Д. КОКС
Д. ХИНКЛИ

Задачи по теоретической статистике с решениями

Издательство
·МИР·

Problems and Solutions in Theoretical statistics

D. R. COX
Department of Mathematics,
Imperial College, London

D. V. HINKLEY
School of Statistics, University of Minnesota

London
Chapman and Hall
A Halsted Press Book: John Wiley & Sons,
New York 1978

Д. КОКС, Д. ХИНКЛИ

Задачи по теоретической статистике с решениями

Перевод с английского
Е. В. ЧЕПУРИНА

под редакцией
Ю. К. БЕЛЯЕВА

Издательство «Мир»
Москва 1981

517.8
К 59

Кокс Дж., Хинкли Дж.

К 59 Задачи по теоретической статистике с решениями. Пер. с англ. Е. В. Чепурина. Под ред. и с предисловием Ю. К. Беляева. — М.: Мир, 1981.
224 с.

Первое на русском языке учебное пособие с упражнениями по курсу математической статистики. Оно принадлежит перу известных английских математиков и дополняет их же монографию «Теоретическая статистика» (М.: Мир, 1978). Позволяет ознакомиться как с конкретными приложениями общих статистических методов, так и с современными достижениями математической статистики. В книге около 150 задач с решениями.

Книга рассчитана на преподавателей, аспирантов, студентов, специализирующихся в области математической статистики и ее приложений.

К $\frac{20203-008}{041(01)-81}$ 8—81, ч. 1. 1702060000 517.8

Редакция литературы по математическим наукам

Д. Кокс, Д. Хинкли

Задачи по теоретической статистике с решениями

Научный редактор А. А. Брянднкская. Младший научный редактор Н. С. Полякова
Художник И. И. Каледин. Художественный редактор В. И. Шаповалов. Технический редактор Л. П. Бирюкова. Корректор Е. К. Монякова

ИБ № 2213.

Сдано в набор 25.09.80. Подписано к печати 20.02.81. Формат 60×90¹/₁₆. Бумага типографская № 2. Гарнитура литературная. Печать высокая. Объем 7 бум. л. Усл. печ. л. 14. Уч.-изд. л. 11,59 Изд. № 1/0914. Тираж 26500 экз. Зак. 2133. Цена 85 коп.

ИЗДАТЕЛЬСТВО „МИР"
Москва, 1-й Рижский пер., 2.

Ордена Октябрьской Революции и ордена Трудового Красного Знамени Первая Образцовая типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54. Валовая, 28. Заказ 2133

От редактора перевода

Предлагаемая вниманию читателя книга Д. Кокса и Д. Хинкли является первым сборником задач по математической статистике, издаваемым на русском языке. Тексты самих задач содержатся в изданной ранее в русском переводе книге тех же авторов „Теоретическая статистика“ (М.: Мир, 1978). Однако здесь эти задачи публикуются вместе с их решениями. Авторы позаботились о том, чтобы чтение книги было возможно без обращения к основному тексту „Теоретической статистики“, снабдив каждую главу сводкой необходимых результатов, хотя знакомство с предыдущей книгой существенно облегчает разбор решений. Как и в предыдущей книге, привлекает внимание неформальный и в то же время достаточно строгий стиль изложения. Следует подчеркнуть, что почти каждая задача содержит какой-либо существенный элемент, освещающий общие положения теории, показывающий их в нестандартном ракурсе, и тем самым облегчает их более глубокое понимание.

Книга безусловно будет полезна тем, кто использует статистические методы и желает понять их содержание и возможности. В круг этих лиц, конечно, включены и те, кто преподает элементы математической статистики во втузах, а также студенты и аспиранты тех специальностей, где предусмотрена расширенная программа обучения математике.

Хотя задачи и снабжены решениями, отнюдь не следует думать, что разбор этих решений будет совсем простым делом. Разбор почти каждой задачи даст чувство удовлетворения благодаря знакомству пусть с небольшим, но содержательным разделом или направлением исследований математической статистики. При этом мы как бы вместе с авторами участвуем в процессе исследования и получении соответствующего результата. В некоторых задачах указаны альтернативные изложенным пути решения задач.

Итак, можно сказать, что это книга об использовании методов математической статистики в решении различных статистических задач.

Ю. К. Беляев

Предисловие

В нашей книге „Теоретическая статистика" в разделах „Дальнейшие результаты и упражнения" мы привели около 150 задач, главным образом для того чтобы проиллюстрировать подлинно важный материал, который невозможно было охватить в основном тексте. Во многих случаях формулировки задач были основаны непосредственно на вышедших недавно научных статьях.

В предлагаемой книге приведены наброски решений и обсуждение этих задач. Чтобы сделать материал книги замкнутым, мы предпослали каждой группе задач краткое изложение необходимых теоретических сведений. Совокупность этих сведений составляет краткий обзор методов математической статистики. Книга содержит значительное количество материала общего характера, не публиковавшегося ранее в книжной форме.

Подробное решение конкретных задач жизненно важно при изучении любой математической дисциплины, и поэтому мы надеемся, что преподаватели и студенты, специализирующиеся в области статистики, извлекут пользу из представленных в книге задач и набросков решений.

Мы также надеемся, что для ученых, работающих в области статистики и интересующихся конкретными задачами, эта книга послужит эффективным обзором некоторых полезных идей теории, а также и соответствующего элементарного математического аппарата.

Хотя нумерация и порядок задач остались теми же, что в „Теоретической статистике", мы переработали заново некоторые задачи, чтобы сделать их замкнутыми и внести некоторые исправления.

Д. Р. Кокс
Лондон
Д. В. Хинкли
Твин Сити
Июнь 1977

1. Введение

Необходимые сведения

Подавляющее большинство теоретических результатов математической статистики касается ситуаций, в которых задана вероятностная модель для вектора данных y, представляющего наблюденное значение случайного вектора Y, имеющего неизвестную плотность вероятностей $f_Y(y)$. Часто эта плотность известна с точностью до значения конечномерного вектора неизвестных параметров θ. В этом случае плотность обозначается $f_Y(y;\ \theta)$. Важнейшим аспектом прикладного статистического исследования является именно выбор подходящей модели.

Теоретическая статистика имеет дело с общими идеями, которые полезны при нахождении ответа на вопросы о неизвестной плотности. Книга основана на полностью эклектической позиции, согласно которой необходимо разнообразие подходов в зависимости от степени детализации задачи. Действительно, чем более детализирована постановка, тем четче очерчиваются пути решения и тем меньше необходимость в привлечении довольно частных критериев.

Во второй главе рассматриваются некоторые общие понятия, такие, как, например, правдоподобие и достаточность. Последующие главы имеют дело с процедурами, которые строятся, вообще говоря, в условиях возрастающего уровня детализации постановки от слабых критериев значимости, в которых задается только нулевая гипотеза в третьей главе до задач теории решений, в которых для неизвестных параметров имеется априорное распределение, а вместе с ним существует и перечень возможных решений, и соответствующая функция полезности в 11 гл.

Насколько это возможно, будут использоваться стандартные обозначения. Для вероятности события A примем обозначение $\mathrm{pr}(A)$, а для ожидания, дисперсии и ковариации — $E(X)$, $\mathrm{var}(X)$ и $\mathrm{cov}(X,\ Y)$ соответственно. Нормальное распределение со средним μ и дисперсией σ^2 будет обозначаться как $N(\mu,\ \sigma^2)$, а p-мерное нормальное распределение с вектором средних μ и ковариационной матрицей Σ — как $MN_p(\mu,\ \Sigma)$. Суммы квадратов и среднеквадратические, связанные с линейной моделью, записываются соответственно как SS и MS. Стандартный нормальный интеграл обозначается как $\Phi(.)$[1], а числовые постоянные будут отмечены звездочкой. В частности, k_{α}^{*} — верхняя α-точка стандартного нормального распределения, определяемая равенством $\Phi(k_{\alpha}^{*}) = 1 - \alpha$.

Число сокращений сведено до минимума. В то же время нам довольно часто приходится использовать такие сокращения, как н.о.р. — независимые и одинаково распределенные, о.п. — отношение правдоподобий, м.п. — максимум правдоподобия, п.р.в. — плотность распределения вероятностей.

---

[1] То есть $\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-u^2/2}\, du$. — *Прим. перев.*

2. О некоторых общих понятиях

Необходимые сведения

Предположим, что вектор наблюдений y является реализацией значений случайной величины Y, имеющей плотность распределения вероятностей (п.р.в.) $f_Y(y;\ \theta)$, которая зависит от параметра θ. Тогда для заданного y правдоподобие $\operatorname{lik}(\theta;\ y)$ полагается равным $f_Y(y;\ \theta)$ и рассматривается как функция параметра θ. Если Y — вектор с независимыми компонентами, логарифм правдоподобия $l(\theta;\ y) = \log \operatorname{lik}(\theta;\ y)$ представляется суммой слагаемых (вкладов), соответствующих компонентам вектора y.

Статистика определяется как функция $T = t(Y)$. В общем случае это вектор. Соответствующее наблюденное значение $t = t(y)$. Статистика S является достаточной для определенного выше семейства распределений, если условная плотность вектора Y при условии $S = s$ не зависит от θ. Иногда достаточность легко установить, вычисляя в явном виде условную плотность. Однако чаще используют для этой цели следующую факторизационную теорему Неймана. Необходимым и достаточным условием того, чтобы статистика S была достаточной для θ, является существование таких функций $m_1(s, \theta)$ и $m_2(y)$, что для всех θ

$$\operatorname{lik}(\theta;\ y) = m_1(s,\ \theta)\, m_2(y). \tag{1}$$

Если только модель считается верной, существуют весьма веские соображения в пользу того, чтобы ожидать, что выводы о параметре θ должны зависеть от данных только через s. Дополнительная информация, которую можно получить из y, соответствует наблюдению с заданным распределением, не зависящим от θ. Для проверки соответствия модели наблюдениям распределение полных наблюдений y можно сравнивать с условным распределением случайной величины Y при заданном $S = s$.

Мы почти всегда используем минимальную достаточную статистику, которая является функцией всех других достаточных статистик и осуществляет для заданной модели наиболее полное сокращение данных. Минимальную достаточную статистику можно получить посредством следующей процедуры: двум выборочным точкам y и z будем приписывать одно и то же значение в том и только том случае, если эти точки порождают пропорциональные функции правдоподобия, т. е. $f_Y(y;\ \theta)/f_Y(z;\ \theta)=h(y,\ z)$ для всех θ.

Достаточная статистика полна в том и только том случае, если из равенства $E\{k(S;\ \theta)\}=0$ для всех θ вытекает, что за исключением множества значений s вероятностной меры нуль функция $k(s)=0$. С математической точки зрения это свойство оказывается важным при исследовании вопроса единственности процедур. Из полноты вытекает минимальная достаточность.

Один важный класс достаточных статистик возникает для моделей, в которых Y состоит из независимых компонент $Y_1,\ \ldots,\ Y_n$, а Y_j имеет п.р.в.

$$\exp\{a(\theta)\,b_j(y_j)+c_j(\theta)+d_j(y_j)\}. \tag{2}$$

Тогда достаточной статистикой будет $\Sigma b_j(Y_j)$. Выделяется случай, когда Y_j одинаково распределены. Для многих стандартных распределений п.р.в. представимы в указанной форме. Семейство распределений, имеющее п.р.в. вида (2), является наиболее простым примером семейства распределений экспоненциального типа. Более общая форма п.р.в. для Y_j, используемая, в частности, когда параметр является вектором, имеет вид

$$\exp\left\{\sum_{k=1}^{m} a_k(\theta)\,b_{jk}(y_j)+c_j(\theta)+d_j(y_j)\right\}. \tag{3}$$

Если $m=q$, где q — размерность вектора θ и носитель меры распределения не зависит от θ, то п.р.в. вида (3) определяет наиболее общий вид семейства распределений, для которого размерность минимальной достаточной статистики совпадает с размерностью параметра.

Когда параметр ϕ и единичное наблюдение выбраны так, что п.р.в. типа (2) принимает вид

$$\exp\{-\phi z+c^{\dagger}(\phi)+d^{\dagger}(z)\}, \tag{4}$$

то говорят, что наблюдение и параметр заданы в естественной форме.

Если $m>q$, то в некоторых случаях часть компонент минимальной достаточной статистики имеет заданные распределения, не зависящие от θ. Такие компоненты статистик называют подчиненными. Имеются веские причины для того, чтобы выводы делались условно относительно наблюденных значений любых подчиненных статистик.

Перечислим основные общие подходы к использованию наблюдений при построении выводов относительно θ:

(i) выборочная теория, при которой интерпретация процедур связана с их вероятностным поведением при воображаемых повторениях наблюдений;

(ii) формальная теория правдоподобия, в которой непосредственно используется лишь $\mathrm{lik}(\theta; y)$;

(iii) различные формы байесовской теории, в которой параметр θ рассматривается как реализованное значение случайной величины Θ, имеющей известную априорную п.р.в. $f_\Theta(\theta)$. В этом случае интерпретация процедур зависит от найденного с помощью теоремы Байеса апостериорного распределения, а именно от условной п.р.в. случайной величины Θ при заданном $Y=y$:

$$f_{\Theta|Y}(\theta/y) \propto f_\Theta(\theta) f_{Y/\Theta}(y|\theta);$$

(iv) теория решений, в которой в дополнение к перечню возможных решений, на одном из которых следует остановиться, можно количественно выразить последствия от заданного способа действия для каждого θ.

Задачи

2.1. Проверьте справедливость факторизации (1) для нормальной линейной модели частного вида, соответствующей линейной регрессии, проходящей через начало координат. Получите отсюда факторизацию для двухфакторной классификации, т. е. для случайных величин Y_{jk} $(j=1, \ldots, m_1;\ k=1, \ldots$ $\ldots, m_2)$, для которых

$$E(Y_{jk}) = \mu + \alpha_j + \beta_k, \quad \Sigma\alpha_j = \Sigma\beta_k = 0.$$

Покажите, что если модель дополнить слагаемым $\gamma\alpha_j\beta_k$, где γ — неизвестно, то при переходе к достаточным статистикам сокращения размерности не происходит.

Решение

Обычно самой хорошей исходной точкой исследования достаточности является логарифм правдоподобия. Если $Y_1, \ldots, Y_n$ независимы и нормально распределены с дисперсией σ^2 и ожиданиями $E(Y_j) = \beta x_j$, то с точностью до постоянной логарифм правдоподобия равен

$$-n \log \sigma - \Sigma (y_j - \beta x_j)^2/(2\sigma^2) =$$
$$= -n \log \sigma - \{\Sigma (y_j - \hat{\beta} x_j)^2 + (\hat{\beta} - \beta)^2 \Sigma x_j^2\}/2\sigma^2,$$

где $\hat{\beta} = \Sigma x_j y_j / \Sigma x_j^2$. Таким образом, логарифм правдоподобия зависит от данных только через $\hat{\beta}$ и $SS_{\text{ост}} = \Sigma (y_j - \hat{\beta} x_j)^2$ — остаточную сумму квадратов. Тем самым получен частный случай факторизационной теоремы Неймана (см. равенство (1) раздела „Необходимые сведения“), т. е. доказана требуемая достаточность. Если σ^2 известна, то происходит дальнейшее сокращение достаточной статистики до β^2.

Логарифм правдоподобия для двухфакторной классификации получается аналогично:

$$-m_1 m_2 \log\sigma - \Sigma\Sigma(y_{jk}-\mu-\alpha_j-\beta_k)^2/(2\sigma^2) =$$
$$= -m_1 m_2 \log\sigma - \Sigma\Sigma\{y_{jk}-\bar{y}_{j.}-\bar{y}_{.k}+\bar{y}_{..}) +$$
$$+(\bar{y}_{j.}-\bar{y}_{..}-\alpha_j)+(\bar{y}_{.k}-\bar{y}_{..}-\beta_k)+(\bar{y}_{..}-\mu)\}^2/(2\sigma^2).$$

Здесь использованы стандартные обозначения. В частности, $\bar{y}_{j.}=\Sigma y_{jk}/m_2$. Возводя в квадрат выражение, стоящее в фигурных скобках, и рассматривая его как сумму четырех компонент, получим, что сумма всех перекрестных произведений обратится в нуль. Таким образом, логарифм правдоподобия является функцией суммы $\Sigma\Sigma(\bar{y}_{jk}-\bar{y}_{j.}-\bar{y}_{.k}+\bar{y}_{..})^2$, строка $\times$ $\times$ вектор взаимодействия сумм квадратов, строки и столбца средних.

Если в модели появляется дополнительно слагаемое $\gamma\alpha_j\beta_k$, то осуществить полученную факторизацию уже не удастся. Таким образом, минимальной достаточной статистикой будет весь массив данных. Чтобы доказать это строго, следует предположить, что для двух различных массивов данных $\{y_{jk}\}$ и $\{z_{jk}\}$ разность логарифмов правдоподобий не зависит от параметров, и в результате приходим к противоречию.

Обсуждение целесообразности введения γ при исследовании взаимодействий проводилось Шеффе (1963, стр. 191) и Мандлом (1971).

[Теоретическая статистика, § 2.1, 2.2 (ii); Рао, § 4; Силвей, стр. 58[1]]

2.2. Предположим, что случайные величины образуют процесс авторегрессии первого порядка

$$Y_r=\mu+\rho(Y_{r-1}-\mu)+\varepsilon_r,$$

где $\varepsilon_1, \ldots, \varepsilon_r$ независимы и одинаково нормально $N(0, \sigma^2)$ распределены, $|\rho|<1$. Выпишите правдоподобие для наблюдений $y_1, \ldots, y_n$, когда начальным значением y_0 является:

(i) заданная постоянная;

(ii) независимая от $\varepsilon_1, \ldots, \varepsilon_n$ случайная величина, имеющая нормальное $N(\mu, \sigma_\varepsilon^2/(1-\rho^2))$ распределение;

(iii) значение y_n.

---

[1] Ссылки на „Теоретическую статистику", а также на работы Лемана, Линдли, Рао и Силвея, которые постоянно встречаются в тексте, детализированы в списке литературы на стр. 220.

В каждом случае найдите минимальную достаточную статистику для параметра $(\mu, \rho, \sigma_\varepsilon^2)$. Покажите, в частности, что в случаях (i) и (ii) статистику $(\Sigma Y_j, \Sigma Y_{j-1}Y_j, \Sigma Y_j^2)$ необходимо дополнить поправками на крайнее значение.

Решение

Правдоподобие для временного ряда и случайного процесса обычно проще всего получить, вычисляя последовательно условные вероятности, т. е. представляя плотность вероятности в следующем виде:

$$f_{Y_0, \ldots, Y_n}(y_0, \ldots, y_n) = f_{Y_0}(y_0) \prod_{j=1}^{n} f_{Y_j \mid Y^{(j-1)}}(y_j / y^{(j-1)}), \qquad (1)$$

где $Y^{(j)} = (Y_0, \ldots, Y_j)$. Для марковских процессов, частным случаем которых является гауссовский процесс авторегрессии первого порядка, имеем

$$f_{Y_j \mid Y^{(j-1)}}(y_j \mid y^{(j-1)}) = f_{Y_j \mid Y_{j-1}}(y_j \mid y_{j-1}).$$

Таким образом, для рассматриваемой частной задачи вклад j-го наблюдения ($j = 1, \ldots, n$) в правдоподобие равен

$$-\log \sigma_\varepsilon - \{(y_j - \mu) - \rho(y_{j-1} - \mu)\}^2/(2\sigma_\varepsilon^2).$$

В случаях (i) и (iii) вкладом в правдоподобие, отвечающим y_0, можно пренебречь, за исключением последнего случая, когда $y_0 = y_n$. В то же время в случае (ii) первый сомножитель в правой части равенства (1) определяется нормальной плотностью для y_0.

Таким образом, в первом случае правдоподобие имеет вид

$$-n \log \sigma_\varepsilon - \left\{ \sum_{j=1}^{n} y_j^2 + \rho^2 \sum_{j=0}^{n-1} y_j^2 - 2\mu \left(\sum_{j=1}^{n} y_j + \rho^2 \sum_{j=0}^{n-1} y_j - \rho \sum_{j=0}^{n-1} y_j - \right.\right.$$
$$\left.\left. - \rho \sum_{j=1}^{n} y_j \right) - 2\rho \sum_{j=1}^{n} y_j y_{j-1} + n\mu^2 (1-\rho)^2 \right\} \Big/ (2\sigma_\varepsilon^2).$$

Следовательно, минимальная достаточная статистика состоит из ΣY_j, ΣY_j^2, $\Sigma Y_j Y_{j-1}$, где суммирование производится по $j = 1, \ldots, n$, дополненных величиной Y_n и фиксированной постоянной y_0. В „круговом" случае (iii) $y_n = y_0$,

$$\sum_{j=1}^{n} y_j^r = \sum_{j=0}^{n-1} y_j^r,$$

и в „краевой поправке" необходимости не возникает.

В случае (ii) логарифм правдоподобия дополняется слагаемым, соответствующим плотности случайной величины $Y_{\cdot}$. Достаточная статистика та же самая, что и в случае (i).

Когда анализируется один длинный ряд, различия между разными начальными условиями несущественны. В то же время, когда информация сводится воедино из нескольких независимых коротких отрезков, использование модели (ii), когда это обоснованно, дает возможность привлечь значительное количество дополнительной информации. Аналогичные соображения относятся и к более общим моделям временных рядов. Круговой случай в приложениях встречается редко.

[Теоретическая статистика, § 2.1, 2.2 (ii); Бартлетт, 1966, § 8.3]

2.3. В простую систему массового обслуживания поступает пуассоновский поток требований с интенсивностью α. Имеется одно обслуживающее устройство, распределение времени обслуживания — показательное с параметром β. Последнее означает, что условная вероятность закончить обслуживание требования в интервале $(t,\ t+\Delta t)$ при условии, что к моменту t система занята, равна $\beta\Delta t+o(\Delta t)$ вне зависимости от предыстории функционирования системы. Докажите, что время между последовательными „событиями“. т. е. поступлениями и окончаниями обслуживания на интервале занятости системы распределено показательно с параметром $(\alpha+\beta)$, а вероятность, что это будет поступление нового требования вне зависимости от соответствующей длины интервала равна $\alpha\,(\alpha+\beta)$. Система наблюдается фиксированное время t, начальное состояние системы — произвольное состояние. Построив правдоподобие исходя из: (a) интервалов между событиями в состоянии занятости системы, (b) типов этих интервалов, (c) интервалов незанятого состояния системы, (d) длительности любого неполного интервала в конце периода наблюдений, покажите, что правдоподобие процесса функционирования системы на $(0,\ t)$ определяется как $\alpha^{n_a}\beta^{n_b}e^{-\alpha t}e^{-\beta(t-t_0)}$, где n_a и n_b — числа прибывших и обслуженных требований за это время, t_0 — время незанятости системы.

Решение

Этим примером иллюстрируется общее соображение относительно того, что часто случайный процесс можно задать различными, хотя и эквивалентными, способами. Соответствующим выбором способа задания процесса можно значительно упростить вычисление правдоподобия.

Пусть A и S — времена от момента осуществления произвольного события до ближайшего следующего поступления требования и ближайшего следующего окончания обслуживания соответственно. И пусть T — время между моментами осуществления последовательности событий, в предположении, что в системе присутствуют еще не обслуженные требования. Тогда $T = \min(A, S)$ и

$$\mathrm{pr}(T \geqslant t) = \mathrm{pr}(A \geqslant t \text{ и } S \geqslant t) = e^{-\alpha t} e^{-\beta t},$$

т. е. T имеет показательное распределение с параметром $\alpha + \beta$. Далее,

$$\mathrm{pr}(\text{следующее событие — поступление требования} \mid t \leqslant T < t+\delta) \sim$$
$$\sim \mathrm{pr}(t \leqslant A < t+\delta,\ S \geqslant t+\delta \mid t \leqslant T < t+\delta) =$$
$$= \mathrm{pr}(t \leqslant A < t+\delta,\ S \geqslant t+\delta) / \mathrm{pr}(t \leqslant T < t+\delta) =$$
$$= \alpha/(\alpha+\beta) + o(1)$$

вне зависимости от значения t.

Вклады в правдоподобие образуются следующим образом:

(i) сомножитель $(\alpha+\beta) d^{-(\alpha+\beta)x}$ отвечает каждому временному интервалу длительности x между осуществлением событий на периоде занятости системы;

(ii) сожножитель $\alpha e^{-\alpha x}$ отвечает временному интервалу длительности x пребывания системы в состоянии, свободном от требований;

(iii) сомножитель $e^{-(\alpha+\beta)x}$ отвечает интервалу длительности x от последнего события до конца наблюдений, если система занята требованиями;

(iv) сомножитель $e^{-\alpha x}$ отвечает интервалу длительности x от последнего события до конца наблюдений, если система не занята.

В дополнение к этому следует учесть сомножитель $\alpha/(\alpha+\beta)$ для временного интервала между событиями, заканчивающегося прибытием требования, и сомножитель $\beta/(\alpha+\beta)$ для каждого такого интервала, заканчивающегося завершением обслуживания требования.

Сразу после перемножения перечисленных сомножителей получим требуемый вид правдоподобия.

[Теоретическая статистика,
§ 2.1; Билингсли, 1961b;
Кокс, 1964a]

2.4. В модели со случайными факторами однофакторного дисперсионного анализа случайные величины Y_{jk} ($j = 1, \ldots, m$; $k = 1, \ldots, r$) имеют следующую структуру:

$$Y_{jk} = \mu + \eta_j + \varepsilon_{jk},$$

где η_j и ε_{jk} независимы в совокупности и нормально распределены с нулевым средним и дисперсиями σ_b^2 и σ_W^2 соответственно. Таким образом, неизвестным параметром является вектор $(\mu, \sigma_b^2, \sigma_W^2)$. Покажите, что минимальная достаточная статистика имеет вид $\{\bar{Y}_{..}, \Sigma(\bar{Y}_{j.}-Y_{..})^2, \Sigma\Sigma(Y_{jk}-\bar{Y}_{j.})^2$, где $\bar{Y}_{j.}=\Sigma Y_{jk}/r$ и $\bar{Y}_{..}=\Sigma\bar{Y}_{j.}/m$. Какова будет минимальная достаточная статистика, если задано, что $\mu=0$? Обобщите результаты на случай, когда каждое Y_{jk} является p-мерным нормальным вектором.

Решение

Один из подходов к построению правдоподобия для модели со случайными факторами состоит в том, чтобы рассмотреть сначала условные распределения при заданных $\eta_1, \ldots, \eta_m$. Более непосредственные соображения связаны с тем, что векторы $Y_j=(Y_{j1}, \ldots, Y_{jr})^T$ независимы и имеют многомерное нормальное распределение со средним $(\mu, \ldots, \mu)^T=\mu 1$ и ковариационной матрицей

$$\Sigma=\sigma_W^2 \mathbf{I}+\sigma_b^2 \mathbf{J}=\sigma_W^2 \mathbf{I}+\sigma_b^2 1 1^T,$$

где $\mathbf{I}$ — единичная матрица размера $r\times r$, а $\mathbf{J}$ — матрица размера $r\times r$, все элементы которой равны единице. Легко показать, что

$$\Sigma^{-1}=\sigma_W^{-2}\{\mathbf{I}-\sigma_b^2\mathbf{J}/(\sigma_W^2+r\sigma_b^2\}.$$

Логарифм правдоподобия принимает вид

$$-\frac{1}{2} m\log|\Sigma|-\frac{1}{2}\sum_{j=1}^{m}(y_j-\mu 1)^T\Sigma^{-1}(y_j-\mu 1)=$$

$$=-\frac{1}{2} m\log|\Sigma|-(2\sigma_W^2)^{-1}\{\Sigma y_j^T y_j-2\mu 1^T\Sigma y_j+m\mu^2 1^T 1-$$

$$-\lambda(\Sigma y_j^T \mathbf{J} y_j-2\mu 1^T\mathbf{J}\Sigma y_j+\mu^2 1^T\mathbf{J}1\},$$

где $\lambda=\sigma_b^2/(\sigma_W^2+r\sigma_b^2)$. Поскольку $\mathbf{J}=11^T$, то выражение в фигурных скобках преобразуется к виду

$$\Sigma\Sigma y_{jk}^2-2\mu rm\bar{y}_{..}+rm\mu^2-\lambda(r\Sigma\bar{y}_{j.}^2-2\mu rm\bar{y}_{..}+r^2\mu^2).$$

Отсюда вытекает достаточность статистики

$$\{\bar{Y}_{..}, \Sigma(Y_{j.}-\bar{Y}_{..})^T, \Sigma\Sigma(Y_{jk}-\bar{Y}_{j.})^T\}.$$

По существу это следует из экспоненциальной формы распределения рассматриваемого семейства. Из полноты вытекает минимальная достаточность.

Если $\mu = 0$, минимальная достаточная статистика принимает вид $\{\Sigma\Sigma Y_{jk}^2, \Sigma\Sigma(Y_{jk} - \bar{Y}_{j.})^2\}$. Если Y_{jk} являются многомерными случайными величинами, введенная выше статистика принимает вид

$$\{\bar{Y}_{..}, \Sigma(\bar{Y}_{j.} - \bar{Y}_{..})(\bar{Y}_{j.} - \bar{Y}_{..})^{\mathrm{T}}, \Sigma\Sigma(Y_{jk} - \bar{Y}_{j.})(Y_{jk} - \bar{Y}_{j.})^{\mathrm{T}}\},$$

т. е. состоит из общего среднего и матриц межгрупповых и внутригрупповых сумм произведений. Это минимальная достаточная статистика для (μ, Σ_b, Σ_W.)

Аналогичные результаты имеют место и для нормальных сбалансированных моделей со случайными эффектами.

[Теоретическая статистика,
§ 2.2; Рао, § 4f, 4j]

2.5. Независимые случайные величины $Y_1, \ldots, Y_n$ таковы, что вероятности единичных значений зависят от контролируемой переменной x, которая принимает значения $x_1, \ldots, x_n$ соответственно. Покажите, что для модели

$$\rho_j = \log\left\{\frac{\operatorname{pr}(Y_j = 1)}{\operatorname{pr}(Y_j = 0)}\right\} = \gamma + \beta x_j$$

минимальной достаточной статистикой является $(\Sigma Y_j, \Sigma x_j Y_j)$. Величину ρ_j называют логическим преобразованием. Обобщите результат на случай, когда h-мерный вектор ρ задается линейной моделью $\rho = \mathbf{x}\beta$, где $\mathbf{x}$ — известная матрица размера $n \times q$ и ранга q, а β — q-мерный вектор неизвестных параметров.

Решение

Вклад в правдоподобие, отвечающий j-му наблюдению, в сжатом виде можно написать как $\exp(\gamma y_j + \beta x_j y_j)/\{1 + \exp(\gamma + \beta x_j)\}$. Перемножая эти величины, получим, что правдоподобие зависит от данных только через $(\Sigma y_j, \Sigma x_j y_j)$. Для общей логистической модели минимальной достаточной статистикой является $\mathbf{x}^{\mathrm{T}} Y$.

[Теоретическая статистика, § 2.2;
Кокс, 1970]

2.6. Плотность распределения вероятностей н. о. р. случайных величин $Y_1, \ldots, Y_n$ на отрезке $0 \leqslant y \leqslant \frac{1}{2}\theta$ равна $2/(3\theta)$, а на интервале $\frac{1}{2}\theta < y \leqslant \theta$ равна $4/(3\theta)$. Докажите, что минимальная достаточная статистика состоит из числа наблюдений, меньших, чем $\frac{1}{2}Y_{(n)}$, и тех порядковых статистик, которые превышают $\frac{1}{2}Y_{(n)}$.

Решение

Обозначим $y_{(1)}\leqslant\ldots\leqslant y_{(n)}$ вариационный ряд выборки, и пусть $\mathrm{hv}(x)=1$ при $x\geqslant 0$ и $\mathrm{hv}(x)=0$ при $x<0$. В этих обозначениях функция правдоподобия записывается в весьма простом виде:

$$\left(\frac{4}{3\theta}\right)^n \prod_{s=1}^{n}\left\{\mathrm{hv}(\theta-y_{(s)})-\frac{1}{2}\mathrm{hv}\left(\frac{1}{2}\theta-y_{(s)}\right)\right\}.$$

Таким образом, она равна нулю при $\theta<y_{(n)}$, а для всякого $y_{(s)}<\frac{1}{2}y_{(n)}$ величина вклада в произведение не зависит от самого значения $y_{(s)}$. Далее, любому $y_{(s)}>\frac{1}{2}y_{(n)}$ в зависимости от того, превосходит ли значение θ величину $2y_{(s)}$ или нет, отвечает сомножитель, равный 1 или 1/2 соответственно. Более формально, отношение правдоподобий для различных выборочных точек y и z не зависит от θ в том и только том случае, если $y_{(n)}=z$ и равны отрезки вариационных рядов, составленных из порядковых статистик, величина которых превосходит половину $y_{(n)}$ и $z_{(n)}$ соответственно. Таким образом, минимальная достаточная статистика образуется из множества порядковых статистик, превосходящих по величине $\frac{1}{2}y_{(n)}$. По нему определяется, в частности, число таких порядковых статистик.

Заметим в связи с этим, что распределение случайной величины R не зависит от θ. Действительно, это немедленно вытекает из инвариантности упомянутого распределения относительно масштабных преобразований. Таким образом, R — подчиненная статистика.

[Теоретическая статистика,
§ 2.2]

2.7. Пусть Y_1 и Y_2 — независимые случайные векторы, распределение которых зависит от одного и того же параметра θ, а S_1 и S_2 — соответствующие минимальные достаточные статистики. Покажите, что для составного вектора $Y=(Y_1, Y_2)$ статистика $S=(S_1, S_2)$ достаточна, хотя и не обязательно минимально достаточна. Для того чтобы показать, что этот результат для зависимых случайных величин вообще-то не справедлив, используйте в качестве ясного примера гауссовский процесс авторегрессии первого порядка. Предложите такое обобщенное определение достаточности, чтобы при объединении совокупности зависимых данных достаточная статистика получалась объединением достаточных статистик отдельных совокупностей. Проиллюстрируйте это определение гауссовским процессом авторегрессии первого порядка.

Решение

Для независимых векторов Y_1 и Y_2 составное правдоподобие имеет вид

$$\begin{aligned}f_{Y_1, Y_2}(y_1, y_2; \theta) &= f_{Y_1}(y_1; \theta) f_{Y_2}(y_2; \theta) = \\ &= f_{Y_1 | S_1}(y_1 | s_1) f_{Y_2 | S_2}(y_2 | s_2) f_{S_1}(s_1; \theta) f_{S_2}(s_2; \theta) = \\ &= m_2(y_1, y_2) m_1(s_1, s_2; \theta),\end{aligned}$$

откуда и вытекает достаточность статистики S. Если распределения величин Y_1 и Y_2 принадлежат одному и тому же семейству распределений экспоненциального типа, то S_1 и S_2 можно выбрать так, чтобы минимальной достаточной статистикой была сумма $S_1 + S_2$.

Практическая интерпретация полученного результата заключается в том, что если отдельные множества данных сокращаются до соответствующих минимальных достаточных статистик, то при объединении множества данных никакой дополнительной информации не требуется.

В случае зависимых величин рассмотрим процесс авторегрессии

$$X_{j+1} = \theta X_j + \varepsilon_{j+1} \quad (j = 1, 2, \ldots),$$

где н. о. р. случайные величины ε_j нормально $N(0, 1)$ распределены. Пусть $Y_1 = (X_1, \ldots, X_n)$, $Y_2 = (X_{n+1}, \ldots, X_{n+r})$, где X_1 задается стационарным распределением процесса, т. е. имеет нормальное $N(0, 1/(1-\theta^2))$ распределение. Тогда, как и в задаче 2.2, получим

$$S_1 = s(Y_1) = \left(X_1^2, \sum_{j=2}^{n-1} X_j^2, \sum_{j=1}^{n-1} X_j X_{j+1}, X_n^2 \right),$$

$$S_2 = s(Y_2) = \left(X_{n+1}^2, \sum_{j=n+2}^{n+r-1} X_j^2, \sum_{j=n+1}^{n+r-1} X_j X_{j+1}, X_{n+r}^2 \right).$$

Однако из вектора (S_1, S_2) не представляется возможным получить

$$s(Y_1, Y_2) = \left(X_1^2, \sum_{j=2}^{n+r-1} X_j^2, \sum_{j=1}^{n+r-1} X_j X_{j+1}, X_{n+r}^2 \right).$$

не зная знака произведения $X_n X_{n+1}$. Для того чтобы в рассматриваемом частном случае получить достаточное сокращение данных, когда Y_1 сопоставляется с любым другим множеством данных из одной и той же реализации процесса, необходимо $S_1 = s(X_1, \ldots, X_n)$ заменить на

$$S_1' = (X_1, \sum X_j^2, \sum X_j X_{j+1}, X_n).$$

В общем случае статистика $S = s(X_1, \ldots, X_n)$ должна быть не только достаточной, но и такой, чтобы при заданном значении S условное распределение любого вектора Y, отличающегося от вектора $(X_1, \ldots, X_n)$, совпадало с условным распределением вектора Y при заданном значении вектора $(X_1, \ldots, X_n)$. Для модели авторегрессии первого порядка это условие как раз выполнено для приведенной выше статистики S', поскольку условные распределения при данном значении вектора $(X_1, \ldots, X_n)$ зависят лишь от X и X_n.

[Теоретическая статистика, § 2.2; Фишер, 1925; Бахадур, 1954; Лауритцен, 1974]

2.9. [1] Популяция состоит из неизвестного числа θ индивидуумов. Для определения объема популяции из нее извлекается случайно индивидуум, помечается и возвращается обратно в популяцию. Второй индивидуум извлекается опять же случайно. Если это помеченный индивидуум, то это фиксируется. В противном случае он помечается и возвращается обратно в популяцию. На m-м шаге описанной процедуры из популяции случайно извлекается индивидуум и фиксируется, имеет ли он метку. Если нет, то индивидуум помечается. После этого он возвращается в популяцию. Вся процедура состоит из n шагов. Наблюдения образуют бинарную последовательность, показывающую, был или не был помечен индивидуум при каждом шаге извлечения. Покажите, что правдоподобие пропорционально $(\theta-1)\ldots(\theta-n+r+1)/\theta^{n-1}$, где r — число извлечений индивидуумов, имеющих метку, а следовательно, R — достаточная статистика. Докажите, что то же самое утверждение справедливо, если:

(a) число n является случайным и определяется последовательным правилом остановки, произвольно зависящим от данных;

(b) для каждого индивидуума с меткой имеются числа конкретных номеров испытаний, при которых он появлялся ранее;

[1] Улучшенный вариант. Задача 2.8 опущена.

(c) на m-м шаге извлекается случайно не один, а k_m индивидуумов.

Составьте перечень предположений, сделанных в процессе анализа, и прокомментируйте их обоснованность при исследовании биологических популяций.

Решение

В результате эксперимента получим последовательность $(Y_1, \ldots, Y_n)$, где $Y_j = 0$, если j-й извлеченный индивидуум уже помечен, и $Y_j = 1$ в противном случае. Ясно, что $Y_1 = 1$. Пусть M_j — общее число индивидуумов с метками после j-го извлечения. Тогда $M_j = Y_1 + \ldots + Y_j$ и

$$\mathrm{pr}(Y_j = 0 \mid Y_1, \ldots, Y_{j-1}; \theta) = \mathrm{pr}(Y_j = 0 \mid M_{j-1}; \theta) = \\ = m_{j-1}/\theta \; (j = 1, \ldots, n),$$

где, конечно, $M_0 = 0$. Отсюда правдоподобие записывается как

$$\mathrm{lik}(\theta; y_1, \ldots, y_n) = \prod_{j=1}^{n} (m_{j-1}/\theta)^{1-y_j} (1 - m_{j-1}/\theta)^{y_j}. \qquad (1)$$

Поскольку r — число извлечений индивидуумов, уже имеющих метку, то $m_n = n - r$ и в последовательности $(Y_1, \ldots, Y_n)$ имеется $n - r$ единиц. Таким образом, последовательность $(m_1, \ldots, m_{n-1})$ пробегает значения $(1, 2, \ldots, n - r - 1)$ и сомножитель $(\theta - k)$ появляется в правдоподобии только один раз, а именно при том извлечении, когда $y = 1$ и m возрастает до $k + 1$. Сомножитель $(\theta - m_n)$ в правдоподобии не появляется, поскольку после того, как $m_j = n - r$, единиц в последовательности $Y_{j+1}, \ldots, Y_n$ наблюдаться не будет. Таким образом,

$$\mathrm{lik}(\theta; y_1, \ldots, y_n) \propto \theta^{-n} \theta (\theta - 1) \ldots (\theta - n + r + 1).$$

В случае выборочной схемы (a) отсылаем к примеру 2.34 „Теоретической статистики“.

В случае выборочной схемы (b) правдоподобие (1) умножается на постоянные $m_1^{-1}, \ldots, m_n^{-1}$. Это следует из того, что условная вероятность извлечь на j-м шаге конкретный индивидуум, уже имеющий метку, равна m_{j-1}^{-1}.

В случае выборочной схемы (c) сомножители появляются не по одному, а группами, однако общий вид правдоподобия не изменится. Отметим, что в предельном случае при $k_m = n$ $R = 0$ и информации о θ в выборке не содержится.

Исходные предположения состоят в том, что популяция замкнута (т. е. новые индивидуумы не рождаются, имеющиеся — не умирают, отсутствует миграция индивидуумов), вероятности быть извлеченными на любом этапе процедуры для всех индивидуумов равны, т. е. равны для различных индивидуумов популяции вне зависимости от того, были или не были извлечены они ранее. И, наконец, предполагается независимость извлечения разных индивидуумов. Для достаточно обширной совокупности данных первое предположение можно ослабить. В то же время второе предположение существенно и в этом случае.

[Теоретическая статистика, § 2.2; Гудман, 1953; Кормак, 1968; Себер, 1973]

2.10. Случайные величины $Y_1, \ldots, Y_n$ подчиняются линейной модели второго порядка. Это означает, что они некоррелированы, имеют постоянные дисперсии и $E(Y) = \mathbf{x}\beta$, где $\mathbf{x}$ — матрица полного ранга размера $n \times q_x$, $q_x < n$. Векторную статистику называют *линейно достаточной* для параметра β, если любая некоррелированная с S статистика имеет ожидание, не зависящее от β. Прокомментируйте связь этого определения с обычной достаточностью. Покажите, что статистика $(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}Y$ линейно достаточна.

Решение

Если T — произвольная статистика, не коррелированная с S, то

$$0 = \operatorname{cov}(T, S) = E(TS) - E(T)E(S) = \\ = E\{E(TS \mid S) - E(T)S\} = \\ = E[S\{E(T \mid S) - E(T)\}].$$

Далее, если статистика S полна и достаточна, то из этого следует, что $E(T) = E(T \mid S)$, а из последнего равенства вытекает, что $E(T)$ не зависит от β. Таким образом, из полноты и достаточности следует линейная достаточность.

Для нормальной линейной модели с известной дисперсией статистика $S = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}Y$ полна и достаточна; тем самым получаем требуемый результат. По смыслу данного выше определения линейной достаточности не требуется нормальности модели. На самом деле при непосредственном доказательстве линейной достаточности статистики S не требуется ни нормальности, ни того, чтобы дисперсия была известна. Действительно, если $\operatorname{cov}(T, S) = 0$, то $T = c^{\mathrm{T}}Y$, где $c^{\mathrm{T}} = d^{\mathrm{T}}\{\mathbf{I} - \mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\}$, а d — некоторый вектор. Тогда

$$E(T) = d^{\mathrm{T}}\{\mathbf{x}\beta - \mathbf{x}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{x}\beta\} = 0.$$

Эти соображения довольно прямо ведут к доказательству теоремы Гаусса, состоящей в том, что оценки наименьших квадратов в линейной модели второго порядка имеют минимальные дисперсии среди всех линейных несмещенных оценок или, что эквивалентно, среди всех линейных оценок с ограниченной среднеквадратической ошибкой. С общими подходами в теории точечного оценивания можно познакомиться в гл. 8.

[Теоретическая статистика, § 2.2.; Барнард, 1963; Рао, § 4a, b]

2.11. Пусть $Y_1, \ldots, Y_n$ — н. о. р. с плотностью распределения вероятностей $f(y; \theta)$, а $S = s(Y)$ — одномерная достаточная статистика для параметра θ при всех n. Исходя из факторизационного соотношения (1), покажите, что для любого θ отношение

$$\frac{\partial}{\partial y_j} \log \left\{ \frac{f(y_j; \theta)}{f(y_j; \theta_1)} \right\} \Big| \frac{\partial}{\partial y_j} \log \left\{ \frac{f(y_j; \theta)_2}{f(y_j; \theta_1)} \right\}$$

не зависит от y_j при любых фиксированных значениях θ_1 и θ_2 параметра θ. Следовательно, это отношение зависит только от θ. Поскольку θ_1 и θ_2 фиксированы, выведите отсюда, что $f(y, \theta)$ имеет форму (2) для экспоненциальных семейств.

Решение

Решение основано на работе Питмена (1936). Пусть

$$\operatorname{lik}(\theta; y) = f_Y(y; \theta) = f_{Y_1, \ldots, Y_n \mid S}(y_1, \ldots, y_n \mid s) f_S(s, \theta).$$

Предполагая дифференцируемость плотности распределения вероятностей, получим

$$\frac{\partial}{\partial y_j} \log \left\{ \frac{\operatorname{lik}(\theta; y)}{\operatorname{lik}(\theta_1; y)} \right\} = \frac{\partial}{\partial y_j} \log \left\{ \frac{f(y_j; \theta)}{f(y_j; \theta_1)} \right\} = \frac{\partial}{\partial y_j} \log \left\{ \frac{f_S(s; \theta)}{f_S(s; \theta_1)} \right\} =$$
$$= \frac{\partial s}{\partial y_j} \frac{d}{ds} \log \left\{ \frac{f_S(s; \theta)}{f_S(s; \theta_1)} \right\}.$$

В формулировке задачи рассматривается отношение производных; поэтому величины $\partial s / \partial y_j$ исчезают и остается лишь функция от s, которая является симметричной относительно $(y_1, \ldots, y_n)$. Но поскольку отношение не зависит от y_k $(k \neq j)$, то оно не может зависеть и от y_j. (Полные варианты рассмотренных обоснований и замечания, относящиеся к модели равномерного распределения на $(0, \theta)$, можно найти в работах Питмена (1936) и Купмена (1936).) Если введенное отношение равно $a(\theta)$, то после интегрирования будем иметь

$$\log\left\{\frac{f(y_j;\ \theta)}{f(y_j;\ \theta_1)}\right\} = a(\theta)\log\left\{\frac{f(y_j;\ \theta_2)}{f(y_j;\ \theta_1)}\right\} + c(\theta).$$

Таким образом, если положить

$$b(y_j) = \log\{f(y_j;\ \theta_2)/f(y_j;\ \theta_1)\},\quad d(y_j) = \log f(y_j;\ \theta_1),$$

то получим плотность, принадлежащую стандартному экспоненциальному семейству.

[Теоретическая статистика,
§ 2.2, разд. (vi)]

2.12. Найдите естественную параметризацию для распределения Пуассона (логарифм среднего), биномиального распределения (логарифм шансов), показательного распределения (обратная величина среднего), гамма распределения с известным параметром формы (обратная величина среднего), отрицательно биномиального распределения с известным параметром формы[1]) (логарифм вероятности успеха), нормального распределения, когда оба параметра неизвестны (среднее, деленное на дисперсию, и обратная величина дисперсии) и многомерного нормального распределения с нулевым средним и неизвестной матрицей ковариаций (обратная матрица ковариаций). Обсудите в каждом случае пределы осмысленного использования естественного параметра для интерпретации данных. Соответствующие соображения основаны на непосредственно физическом смысле и области допустимых значений.

Решение.

Соотношением (4), приведенным в разделе „Необходимые сведения“, определяется наиболее простой вид п. р. в. экспоненциального семейства распределений. В многопараметрическом случае первое слагаемое в показателе экспоненты заменяется суммой. В табл. 1 в общепринятых обозначениях приводятся основные частные случаи семейств экспоненциального типа.

С точки зрения простоты теории при сравнении двух выборок в более общем случае при анализе линейных моделей предпочтительнее, чтобы разность параметров была разностью есте-

---

[1]) При параметризации: (a) п. р. в. гамма-распределения $f_\Gamma(y;\ \rho,\ k) = \rho(\rho y)^{k-1}e^{-\rho y}/\Gamma(k)$ параметром формы называют k; (b) вероятностей отрицательно биномиального распределения $f_N(n;\ r,\ p) = \binom{n-1}{r-1}\rho^r(1-\rho)^{n-r}$ параметром формы называют r.— *Прим. перев.*

Таблица 1

Распределение	Плотность	Естественный параметр
Пуассона	$\exp(y\log\mu-\mu-\log!)$	$\log\mu$
Биномиальное	$\exp\left\{y\log\left(\frac{\theta}{1-\theta}\right)+n\log(1-\theta)+\log\binom{n}{y}\right\}$	$\log\{\theta/(1-\theta)\}$
Показательное	$\exp(y/\mu-\log\mu)$	$1/\mu$
Гамма (показатель формы известен)	$\exp\{-k_0y/\mu+k_0\log(k_0/\mu)+(k_0-1)\log y-$ $-\log\Gamma(k_0)\}$	$1/\mu$
Отрицательное биномиальное (показатель формы известен)	$\exp\left\{y\log\theta+k_0\log(1-\theta)+\log\binom{k_0+y-1}{k_0-1}\right\}$	$\log\theta$
Нормальное	$\exp\left\{-\frac{y^2}{2\sigma^2}+\frac{y\mu}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\frac{1}{2}\log(2\pi)\right\}$	$\frac{1}{\sigma^2}$, $\frac{\mu}{\sigma^2}$
Многомерное размерности p нормальное (с нулевым средним)	$\exp\left\{-\frac{1}{2}y^T\Sigma^{-1}y-\frac{1}{2}\log\lvert\Sigma\rvert-\right.$ $\left.-\frac{1}{2}p\log(2\pi)\right\}$	Элемент матрицы Σ^{-1}

ственных параметров, или чтобы линейная модель определялась в естественной параметризации. В этом методическое преимущество естественной параметризации. В качестве иллюстрации сделанного утверждения можно рассматривать задачу 2.5. С прикладной точки зрения указанная шкала для параметра привлекательна для пуассоновского и биномиального распределений и в гораздо меньшей степени в других случаях. Так, обычно $\log\sigma$ является более подходящей функцией для сравнения дисперсий нормальных распределений, чем $1/\sigma^2$. Для показательного распределения $1/\mu$ необходимо интерпретировать как интенсивность. Однако, для интенсивностей, или средних, часто опять же более подходят мультипликативные модели. Поскольку параметрическое пространство для $1/\mu$ совпадает (в лучшем случае) с положительной полупрямой, то линейные модели на основе значений $1/\mu$ имеют смысл только для ограниченного множества параметрических значений. Последнее обстоятельство может вызывать трудности как при проверке согласия данных с моделью, так и при интерпретации результатов. Как для пуассоновского, так и для биномиального распределения естественный параметр может принимать любое вещественное значение.

Нелдер и Веддербэн (1972) предложили важную идею обобщения линейной модели, при котором некоторые функции естественного параметра подчиняются линейной модели.

[Теоретическая статистика, § 2.2 (vi), Леман, 1979, стр. 62 и далее]

2.13. Докажите, что когда плотность экспоненциального семейства берется в ее естественной форме

$$f_Z(z;\ \phi) = \exp\{-z\phi + c^\dagger(\phi) + d^\dagger(z)\},$$

то производящей функцией кумулянтов случайной величины Z является $c^\dagger(t+\phi) - c^\dagger(\phi)$. Чтобы показать это, используйте условие, что интеграл по всему пространству от плотности равен единице. Исходя из вида производящей функции кумулянтов, выразите кумулянты случайной величины Z для малых значений ϕ через кумулянты случайной величины Z при $\phi=0$. Сравните с невырожденным случаем, для которого известны точные выражения кумулянтов.

Решение

Если

$$f_Z(z;\ \phi) = \exp\{-z\phi + c^\dagger(\phi) + d^\dagger(z)\},$$

то

$$E(e^{tZ};\ \phi) = \exp\{c^\dagger(\phi) - c^\dagger(\phi - t)\} \int \exp\{-(\phi - t)z + c^\dagger(\phi - t) + d^\dagger(z)\}\, dz.$$

Интеграл в правой части равенства равен единице. Аналогичное представление имеет место в дискретном случае. Таким образом, для производящей функции кумулянтов имеем

$$K(t;\ \phi) = \log E(e^{tZ};\ \phi) = c^\dagger(\phi) - c^\dagger(\phi - t) = K(t-\phi;\ 0) - K(-\phi;\ 0).$$

Далее, r-й кумулянт случайной величины Z при значении параметра, равном ϕ, определяется как $\varkappa_r(\phi) = [\partial^r K(t,\ \phi)/\partial t^r]_{t=0}$. Таким образом, из теоремы Тейлора получаем

$$\varkappa_r(\phi) = [\partial^r K(u;\ 0)/\partial u^r]_{u=-\phi} = \sum_{s=0}^{\infty} (-\phi)^s \varkappa_{r+s}/s!,$$

где $\varkappa_t = \varkappa_t(0)$.

В качестве простого примера рассмотрим нормальное $N(\mu,\ \sigma_0^2)$ распределение, где σ_0^2 — известный параметр. Пусть

$Z=-Y$ и $\phi=\mu/\sigma_0^2$. В этом случае

$$\varkappa_1(\phi)=\varkappa_1(0)-\phi\varkappa_2(0)=0-(\mu/\sigma_0^2)\,\sigma_0^2,$$

т. е. точный результат получается из первых двух членов разложения в ряд функции $K(t;\ 0)$.

[Теоретическая статистика, § 2.2 (vi)]

2.14. Случайные величины $Y_1, \ldots, Y_n$ н. о. р. с плотностью $\tau^{-1}h\{(y-\mu)/\tau\}$. Найдите такое преобразование $Z_1, \ldots, Z_n$ вариационного ряда $Y_{(1)}, \ldots, Y_{(n)}$, чтобы $Z_3, \ldots, Z_n$ были инвариантны при изменениях сдвига и масштаба. Исходя из этого, найдите подчиненную компоненту минимальной достаточной статистики.

Решение

Сдвигово-масштабное преобразование исходных данных для некоторых a и b переводит $Y_{(j)}\rightarrow aY_{(j)}+b$. Инвариантная статистика принимает одно и то же значение для всех a и b. Разности $Y_{(j+1)}-Y_{(j)}$ преобразуются в $a(Y_{(j+1)}-Y_j)$, а a исключается переходом к отношениям. Таким образом, отношения

$$\frac{(Y_{(3)}-Y_{(2)})}{(Y_{(2)}-Y_{(1)})},\ \ldots,\ \frac{(Y_{(n)}-Y_{(n-1)})}{(Y_{(2)}-Y_{(1)})}$$

инвариантны и могут быть взяты в качестве $Z_3, \ldots, Z_n$. После этого можно положить $Z_1=Y_{(1)}$, $Z_2=Y_{(2)}-Y_{(1)}$, чем и завершится взаимно однозначное преобразование минимальной достаточной статистики $(Y_{(1)}, \ldots, Y_{(n)})$. Легко проверить, что якобиан при таком переходе невырожден. Ясно, что имеется много эквивалентных форм полученного представления.

Статистика $(Z_3, \ldots, Z_n)$ является подчиненной и может быть получена из общей теоремы Басу (1955). Чтобы убедиться в этом непосредственно, положим $Y_j=\mu+\tau Y'_j$. Величины $(Z_3, \ldots, Z_n)$, полученные из $Y_{(1)}, \ldots, Y_{(n)}$, равны величинам $(Z'_3, \ldots, Z'_n)$. В то же время совместная плотность величин $(Z'_3, \ldots, Z'_n)$ является интегралом от $\Pi h(y'_j)$, а следовательно, не зависит от $(\mu,\ \tau)$.

Имеются теоретически важные применения статистик $(Z_3, \ldots, Z_n)$ при получении выводов о параметрах сдвига и масштаба.

[Теоретическая статистика, § 2.2 (vi); Фишер, 1934]

2.16. [1]) Конечная популяция состоит из m индивидуумов, снабженных номерами $1, \ldots, m$. Популяция такова, что каждому индивидууму соответствует или число θ, или число $1-\theta$.

[1]) Задача 2.15 опущена.

Обозначим $\theta_1, \ldots, \theta_m$ значения чисел, отвечающих индивидуумам. Это неизвестные постоянные, каждая из которых равна нулю или единице. Из конечной популяции извлекается выборка без возвращения объема n. В нее попадают в порядке появления индивидуумы с номерами $i_1, \ldots, i_n$ и соответствующие значения θ_{i_j} наблюдаются. Покажите, что правдоподобие имеет вид

$$\begin{cases} (m-n)!/m! & \text{при } \theta_{i_j}=y_j,\ j=1, \ldots, n, \\ 0 & \text{в противном случае} \end{cases}$$

и, следовательно, не зависит от ненаблюденных значений θ_{i_k}, $k=n+1, \ldots, m$. Покажите далее, что по существу та же самая функция правдоподобия получается для любой выборочной схемы, если только эта схема определяется независимо от θ_j, $j=1, \ldots, m$.

Обсудите применение этого факта для различных подходов к статистическому выводу. Означает ли это, что: (a) ничего нельзя узнать о ненаблюденных значениях θ_{i_k}, $k=n+1, \ldots, m$; (b) интерпретация выборочного обследования не должна зависеть от выборочной схемы; (c) принцип правдоподобия не применим, по крайней мере в данной ситуации.

Решение

Весьма важно, что наблюдаются номера $i_1, \ldots, i_n$. Итак, данные имеют вид $(i_1, \ldots, i_n;\ y_1=\theta_{i_1}, \ldots, y_n=\theta_{i_n})$. При случайном выборе все выборки объема n равновероятны, число различных выборок без учета порядка появления номеров равно $m!/(m-n)!$. Таким образом,

$$\operatorname{pr}(I_1=i_1, \ldots, I_n=i_n, y_1, \ldots, y_n;\ \theta)=$$
$$=\begin{cases} \operatorname{pr}(I_1=i_1, \ldots, I_n=i_n) & \text{при } y_j=\theta_{i_j},\ j=1, \ldots, n, \\ 0 & \text{в противном случае,} \end{cases}$$

откуда

$$\operatorname{lik}(\theta;\ i_1, \ldots, i_n;\ y_1, \ldots, y_n)=$$
$$=\begin{cases} (m-n)!/m! & \text{при } \theta_{i_j}=y_j,\ j=1, \ldots, n, \\ 0 & \text{в противном случае.} \end{cases}$$

Это не то же самое, что $\operatorname{lik}(\theta;\ y_1, \ldots, y_n)$, зависящее от θ только через $\Sigma\theta_j$.

Подобного же рода результат получится и в том случае, когда вероятности $\operatorname{pr}(I_j=i_j;\ j=1, \ldots, n)$ не зависят от θ.

Вследствие бинарности модели ответы на вопросы (a)—(c) связаны с выводом утверждений, подобных тому, что:

(i) $E(\Sigma Y_j/n;\ \theta)=\Sigma\theta_j/m$, $\operatorname{var}(\Sigma Y_j/n;\ \theta)=O(1/n)$,

(ii) $\sum_{j=1}^{m}\theta_j=\sum_{s\notin\{i_j\}}\theta_s+\sum_{j=1}^{n}y_j$;

(iii) $\operatorname{lik}(\theta;\ y_1, \ldots, y_n)=$

$$=\operatorname{pr}\left(Y_1=y_1, \ldots, Y_n=y_n;\ \tau=\sum_{j=1}^{m}\theta_j\right)\propto\binom{\tau}{\Sigma y_j}\binom{m-\tau}{n-\Sigma y_j}.$$

[Теоретическая статистика, § 2.2, 2.4 (viii); Годамб и Томпсон, 1971; Ройялл, 1976]

2.17. Для модели $f_Y(y;\ \theta)$ статистику S называют *байесовски достаточной*, если вне зависимости от априорного распределения параметра θ апостериорное распределение зависит от данных только через значения s. Покажите, что байесовская достаточность эквивалентна „обычной“ достаточности.

Решение

В соответствии с теоремой Байеса, для байесовской достаточности необходимо, чтобы

$$f_{\Theta|Y}(\theta|y)=f_{Y|\Theta}(y|\theta)f_\Theta(\theta)/f_Y(y)=g(s,\ \theta)$$

для $s=s(y)$. Отсюда вытекает, что

$$f_{Y|\Theta}(y|\theta)=\frac{g(s,\ \theta)}{f_\Theta(\theta)}f_Y(y),$$

т. е. имеется требуемая факторизация для обычной достаточности. Обратное доказывается аналогично.

[Теоретическая статистика, § 2.2, 2.4 (IX); стр 408; Линдли, стр. 21; Райфа и Шлайфер, 1961]

3. Слабые критерии значимости

Необходимые сведения

Нулевая гипотеза H_0 является точно сформулированным предположением относительно характеристик вероятностного распределения случайной величины Y. Гипотезу называют простой, если распределение случайной величины Y определяется ею однозначно. В противном случае гипотезу называют сложной. Предполагается, что, как правило, имеется качественная информация относительно того, какого типа отклонения от H_0 требуется выявить. Критерий согласия с H_0, который строят в ситуации, когда информация о типе отклонения от H_0 определена недостаточно четко, называют слабым критерием значимости.

Такой критерий строится на основе статистики критерия $T = t(Y)$, которая обладает следующими свойствами:

(i) большие значения величины t являются более убедительным доказательством наличия того типа отклонения, которое необходимо проверить;

(ii) в случае справедливости H_0 распределение статистики T известно, в крайнем случае известно приближенно.

Для заданного наблюдения y вычисляем $t = t_{набл} = t(y)$. После этого получаем уровень значимости $p_{набл}$, определяемый соотношением $p_{набл} = \text{pr}(T \geqslant t_{набл};\ H_0)$. Эта величина получает теоретическую интерпретацию как вероятность ошибочного отвержения H_0, когда анализируемые данные рассматривались бы как явно свидетельствующие против H_0.

Если данные имеют непрерывное распределение, то при нулевой гипотезе H_0 случайная величина P, соответствующая $p_{набл}$, равномерно распределена на $(0,1)$. При наличии того типа отклонения от H_0, которое измеряется величиной t, уровень значимости будет иметь тенденцию принимать малые значения.

Когда гипотеза H_0 сложная, распределение статистики T должно быть в точности или приближенно одним и тем же для всех простых гипотез, образующих H_0. Часто такие статистики можно найти, рассматривая распределения, условные относительно достаточных статистик для мешающих параметров, определяемых H_0. В противном случае в качестве T иногда можно взять статистику, инвариантную относительно преобразований данных, порождающих различные простые гипотезы, образующие H_0.

Если для проверки одной и той же гипотезы по одним и тем же данным применяются m критериев, а затем берется наиболее значимый из них, то при определении правила принятия гипотезы это необходимо принять в расчет. В качестве статистики критерия используется в действительности величина $Q = \min(P_1, \ldots, P_m)$, малые значения которой считаются значимыми. Истинный уровень значимости, соответствующий $q_{\text{набл}}$, ограничен сверху величиной $mq_{\text{набл}}$.

Слабые критерии значимости, хотя и полезны, однако их ценность для статистических выводов довольно ограниченна. Дело в том, что они не дают непосредственных указаний о величине возможных отклонений от H_0.

Задачи

3.1. Рассмотрите случайные величины $Y_1, \ldots, Y_n$, представляющие направления на круге, т. е. углы, заключенные между 0 и 2π. Нулевая гипотеза состоит в том, что случайные величины $Y_1, \ldots, Y_n$ независимы и равномерно распределены на $(0, 2\pi)$. Покажите, что для выявления сгущения точек вокруг нуля подходящей статистикой критерия является $\Sigma \cos Y_j$ и что ее можно интерпретировать как абсциссу случайного блуждания на плоскости со скачками единичной длины. Предложите статистики критерия в ситуации, когда возможно сгущение точек вокруг нуля и π, а также когда возможно сгущение вблизи некоторого неизвестного направления. В обоих случаях получите нормальную аппроксимацию распределения статистики при нулевой гипотезе, соответствующей равномерности распределения углов по кругу.

Решение

Вектор единичной длины, имеющий угол Y_j с осью абсцисс,— это вектор с компонентами $(\cos Y_j, \sin Y_j)$. Результирующий вектор суммы n таких векторов имеет компоненты $(\Sigma \cos Y_j, \Sigma \sin Y_j)$. Если векторы имеют тенденцию сгущаться около нулевого угла, то компонента $T = \Sigma \cos Y_j$ будет принимать, как правило, большие значения. При нулевой гипотезе распределение статистики T имеет простую нормальную аппроксимацию с учетом моментов $E(\cos Y_j) = 0$ и $\operatorname{var}(\cos Y_j) = \frac{1}{2}$. Таким образом, статистика T приближенно $N\left(0, \frac{1}{2}n\right)$ нормально распределена.

Если векторы сгущаются около $y = 0$ или $y = \pi$, то имеются две возможности:

а) работать с величинами $Y'_j = 2Y_j$, поскольку для Y'_j будет иметь место сгущение около направлений $y=0$ и $y=2\pi$, что приведет к одинаковым распределениям для статистик $\Sigma\cos(2Y_j)$ и T;

(b) преобразовать $\cos Y_j$ в $|\cos Y_j|$ или $\cos^2 Y_j$. В последнем случае вновь придем к статистике, рассмотренной в пункте (а).

Если сгущение происходит около неизвестного направления, то статистика критерия должна быть инвариантной относительно вращения осей. Наличие сгущения ведет к удлинению результирующего вектора. Таким образом, разумной статистикой будет $T' = (\Sigma\cos Y_j)^2 + (\Sigma\sin Y_j)^2$ — квадрат длины результирующего вектора. Статистика T' инвариантна. При нулевой гипотезе статистики $\Sigma\cos Y_j$ и $\Sigma\sin Y_j$ некоррелированы, обе суммы имеют нулевое среднее и дисперсии, равные $\frac{1}{2}n$. Таким образом, в соответствии с двумерной центральной предельной теоремой статистика $2T'/n$ асимптотически распределена как хи-квадрат случайная величина с двумя степенями свободы, т. е. как показательная случайная величина со средним, равным двум.

Имеется обширная литература по только что рассмотренной и более сложным подобным ситуациям, связанным с анализом данных о направлениях (Мардиа, 1972; Пирсон и Хартли, 1972, табл. 56—64).

[Теоретическая статистика,
§ 3.2, 3.3]

3.2. Предложите статистику критерия для проверки нулевой гипотезы о том, что бинарные данные являются реализацией марковской цепи с двумя состояниями и заданной переходной матрицей. Особый интерес представляют отклонения, отвечающие более частым переходам из состояния в состояние по сравнению с тем, как это можно было бы предположить для заданной матрицы.

Решение

По соображениям, связанным главным образом с принципом достаточности, разумно рассмотреть в первую очередь статистики критерия, которые являются функциями числа переходов m_{rs}, т. е. числа одношаговых переходов из r в s $(r, s=0, 1)$. Далее, если заданная матрица переходов имеет вид

$$\begin{bmatrix} \alpha_0 & 1-\alpha_0 \\ 1-\beta_0 & \beta_0 \end{bmatrix},$$

где α_0 и β_0 известны, то о наличии отклонения проверяемого типа будут свидетельствовать неравенства $m_{00}/(m_{00}+m_{01})<\alpha_0$ и $m_{11}/(m_{10}+m_{11})<\beta_0$.

Простые статистики критерия имеют вид взвешенной линейной комбинации расхождений данных с гипотезой, т. е. линейной комбинацией статистик $m_{00}-\alpha_0(m_{00}+m_{01})$ и $m_{11}-\beta_0(m_{10}+m_{11})$. Наиболее простая среди них — сумма расхождений. Точное распределение статистики критерия при нулевой гипотезе можно найти, в принципе, перебором всех различных бинарных последовательностей. Нормальную аппроксимацию можно получить на основе результатов, содержащихся в обзоре Биллингсли (1961a).

[Теоретическая статистика, § 3.2]

3.3. *Абсолютный критерий значимости* простой гипотезы — это слабый критерий значимости, в котором статистика критерия в наблюдаемой точке y является при нулевой гипотезе строго убывающей функцией от п. р. в. в точке y. Это означает, что уровень значимости является вероятностью получения при нулевой гипотезе данных со столь же малой или даже меньшей плотностью вероятностей по сравнению с п. р. в. данных, полученных фактически. Обсудите критически, является ли плодотворной эта идея. Покажите, в частности, что если критерий применяется к данным, подвергнутым нелинейным преобразованиям в случае непрерывных случайных величин, или если для дискретных данных проводится группировка, включающая и точки выборочного пространства, то все это может сильно воздействовать на уровень значимости.

Решение

Насколько нам известно, первым примером абсолютного критерия значимости является хи-квадрат критерий согласия в том виде, как он введен К. Пирсоном (1900). Для того чтобы проверить согласие полиномиальных данных с заданной совокупностью вероятностей, он ввел упорядочение среди всех точек выборочного пространства в соответствии с их вероятностями при нулевой гипотезе. Хи-квадрат статистика вводилась в связи с этим как средство приближенного упорядочения (см. задачу 3.4). Асимптотические примеры возникают, по крайней мере неявно, в связи с таблицами сопряженности.

При анализе абсолютных критериев значимости возникают, по существу, три затруднения:

(i) Для непрерывных случайных величин упорядочение выборочных точек на основе значений соответствующей п. р. в. можно изменить каким угодно образом, выбрав подходящее монотонное преобразование. Таким образом, необходимы критерии, однозначно определяющие форму распределения Y, или необходимо исключить непрерывные распределения.

(ii) Для дискретных случайных величин упорядочение выборочных точек можно менять как „расщеплением", так и группировкой выборочных точек. Здесь весьма существенную роль играет сокращение данных до значений достаточных статистик.

(iii) Очень часто вероятности при H_0 изменяются гладко вместе со значением естественной статистики критерия. Поэтому с точки зрения проверки гипотез абсолютный критерий значимости является разумным критерием. Однако вероятности могут изменяться и нерегулярно. Примером крайней ситуации является ранговый коэффициент корреляции Спирмена (Кендалл, 1962), когда для проверки гипотезы независимости неразумно использовать абсолютный критерий значимости.

Возникает также более философский вопрос: должны ли мы относиться к H_0 просто с подозрением, когда выборочная точка имеет малую вероятность, или мы должны высказать соображения, содержащие некоторое разумное объяснение данных, делающее данные более правдоподобными?

[Теоретическая статистика, § 3.2; Мартин-Лёф, 1974; Свердруп, 1975]

3.4. Предположим, что при каждом испытании возможны m исходов. Данные состоят из результатов n независимых испытаний при одних и тех же условиях. Покажите, что если случайные величины $(Y_1, \ldots, Y_m)$ задают число осуществлений исходов каждого типа, тогда

$$\operatorname{pr}(Y_1 = y_1, \ \ldots, \ Y_m = y_m) = n! \, \Pi \left(p_j^{y_j} \big/ y_j! \right),$$

где $(p_1, \ldots, p_m)$ — вероятности конкретных исходов в одном испытании. Покажите далее, что для больших n логарифм упомянутых вероятностей является линейной функцией от

$$\Sigma \, (y_j - np_j)^2/(np_j)$$

в соответствующем интервале значений $y_j - np_j = O(\sqrt{n})$. Покажите, что абсолютный критерий значимости нулевой гипотезы, которая полностью определяет вероятности p_j, $j = 1, \ldots, m$, эквивалентен критерию, основанному на статистике хи-квадрат критерия согласия.

Решение

В стандартной полиномиальной формуле для $\mathrm{pr}(Y=y)$ положим $y_j = np_j + z_j\sqrt{np_j}$ для $j=1, \ldots, m$. После этого рассмотрим логарифм этой вероятности, устремив $n \to \infty$, считая z_j фиксированными величинами, на которые наложено условие $\Sigma z_j \sqrt{p_j} = 0$. Таким образом, используя разложение Стирлинга, при $n \to \infty$ получим

$$\log \mathrm{pr}(Y=y) = \log n! + \Sigma \left\{np_j + z_j\sqrt{(np_j)}\right\} \log p_j -$$
$$- \Sigma \log \left\{np_j + z_j\sqrt{np_j}\right\}! = \mathrm{const} - \frac{1}{2}\Sigma z_j^2 + o(1).$$

Далее, совместное асимптотическое рапределение случайных величин Z_j, отвечающих z_j, — это условное нормальное сферическое распределение $MN_m(0, 1)$ при условии, что $\Sigma Z_j\sqrt{p_j} = 0$.

Для задачи проверки нулевой гипотезы, состоящей в том, что вероятности отдельных событий равны $p_1, \ldots, p_m$, из сказанного выше вытекают два главных следствия. Во-первых, что упорядочение выборочных точек в соответствии с убыванием $\mathrm{pr}(Y=y)$ совпадает асимптотически с упорядочением на основе убывания Σz_j^2-статистик классического хи-квадрат критерия. Таким образом, хи-квадрат критерий является аппроксимацией для абсолютного критерия значимости, определенного в том виде, как это было сделано в задаче 3.3. Второе следствие состоит в том, что распределение величин Z_j асимптотически совпадает с распределением m н. о. р. нормальных $N(0,1)$ случайных величин, между которыми имеется одна линейная связь $\Sigma Z_j\sqrt{p_j} = 0$. Отсюда вытекает, что ΣZ_j^2 асимптотически имеет хи-квадрат распределение с $(m-1)$ степенями свободы.

[Теоретическая статистика, § 3.2; Рао, § 6b; Силвей, стр. 120]

3.5. Покажите, что для непрерывных распределений статистики Колмогорова и Крамера — фон Мизеса не зависят от конкретного проверяемого распределения $G(.)$. Докажите, что вторую статистику можно записать в виде

$$\frac{1}{12n^2} + \frac{1}{n}\Sigma\left\{G(Y_{(j)}) - \frac{2j-1}{2n}\right\}^2.$$

Решение

Строго монотонные преобразования на носителе меры функции распределения $G(\cdot)$, переводящие различные выборочные точки в различные, оставляют инвариантным любой функционал от эмпирической функции распределения $\bar{F}_n(y)$ и $G(y)$, и в том числе интегралы вида

$$\int a\{G(y),\ \bar{F}_n(y)\}\, dG(y) + \int b\{G(y),\ \bar{F}_n(y)\}\, d\bar{F}_n(y).$$

Как статистка Колмогорова, так и статистика Крамера—фон Мизеса представляются в указанной форме. Таким образом, их выборочные значения, а следовательно, и распределения при нулевой гипотезе инвариантны при монотонных преобразованиях. Преобразуем теперь y в $G(y)$. После этого проверяемое распределение при нулевой гипотезе превратится в равномерное на (0,1). Статистика Крамера—фон Мизеса примет вид

$$\int_0^1 [\bar{F}_n\{G^{-1}(y)\} - y]^2\, dy = \sum_{j=1}^{n+1} \int_{G(y_{(j-1)})}^{G(y_{(j)})} \left(\frac{j-1}{n} - y\right)^2 dy,$$

где $G(y_{(0)}) = 0$, $G(y_{(n+1)}) = 1$, т. е. сведется к требуемой форме.

При исследовании точных распределений таких статистик при нулевой гипотезе удобно, конечно, считать, что $G(y)$ — равномерное распределение.

[Дурбин, 1973; Рао, § 6; Силвей, стр. 140]

3.6. Нулевая гипотеза для бинарных данных заключается в том, что они порождены марковской цепью с двумя состояниями и неизвестной матрицей переходов. Предложите некоторые из типов отклонений, которые было бы интересно проверить, и укажите соответствующие статистики критерия в каждом случае.

Решение

Основными типами отклонений от гипотезы простой марковской цепи можно считать следующие:

(a) переходы могут зависеть от дополнительных наблюдений, которые можно производить для каждого испытания;

(b) переходные вероятности могут быть нестационарными, например может иметь место временной тренд или периодические флуктуации;

(c) может существовать зависимость переходных вероятностей от более чем одного текущего состояния или сама зависимость совсем другая, чем формулируемая в нулевой гипотезе.

Довольно общее семейство статистик критерия можно определить следующим образом. Сопоставим каждому переходу

$Y_j \to Y_{j+1}$ контролируемую переменную z_j. Так, для отклонений типа (b) можно положить $z_j = j$ или $z_j = \cos(\omega_0 j)$, а для отклонений типа (c) можно принять $z_j = Y_{j-1}$. В соответствии с четырьмя типами переходов, а именно $0 \to 0$, $0 \to 1$, $1 \to 0$, $1 \to 1$, можно сформулировать четыре группы z-величин. Пусть $\bar{z}_{rs}$ (r, $s = 0, 1$) — это средние групповых значений z-величин. Среди разумных статистик критерия содержатся, в частности, статистики $(\bar{z}_{01} - \bar{z}_{00})^2 + (\bar{z}_{10} - \bar{z}_{11})^2$ и $n_0(\bar{z}_{01} - \bar{z}_{00})^2 + n_1(\bar{z}_{01} - \bar{z}_{11})^2$, где состояние i случалось n_i раз ($i = 0, 1$).

Хотя и представляется возможным предлагать для каждого случая такие частные статистики критерия, существует класс относительно сложных ситуаций, в которых, вероятно, более плодотворно рассмотреть более формальный подход последующих глав, где альтернативные вероятностные модели формулируются в явном виде. Это было сделано для сформулированного выше отклонения типа (3) (см. Андерсон и Гудмен, 1957).

[Теоретическая статистика, §3.3]

3.7. Используя геометрические соображения, докажите, что если $Y_1, \ldots, Y_n$ — независимые и одинаково нормально распределенные случайные величины, то коэффициенты асимметрии и эксцесса не зависят от оцениваемого среднего и дисперсии, а распределения коэффициентов асимметрии и эксцесса не зависят от среднего и дисперсии исходного распределения.

Решение

Произведем сначала ортогональное преобразование к новым переменным $\bar{Y}\sqrt{n}$ и $Z_1, \ldots, Z_{n-1}$, а затем перейдем к сферической полярной системе координат

$$Z_1 = R\cos\Theta_1,\ Z_2 = R\sin\Theta_1\cos\Theta_2,\ \ldots,\ Z_{n-1} = R\sin\Theta_1 \ldots \sin\Theta_{n-2}.$$

Положим $R^2 = \Sigma(Y_i - \bar{Y})^2$. Как вытекает из сферической нормальности случайных величин $Z_1, \ldots, Z_{n-1}$, статистика R распределена независимо от угловых координат. В то же время из соображений размерности следует, что оценки коэффициентов асимметрии и эксцесса являются функциями лишь угловых координат.

3.8. Перечислите и классифицируйте некоторые статистики критерия для проверки нулевой гипотезы о нормальности, построенные по н.о.р. случайным величинам. Исследуйте каждую статистику критерия с точки зрения возможности обобщения ее для проверки многомерной нормальности.

Прокомментируйте для многомерных критериев следующие моменты:

(a) сколь просто можно найти или аппроксимировать распределение статистики критерия при нулевой гипотезе;

(b) просто ли можно вычислить саму статистику даже при большой размерности наблюдений;

(c) в какой степени статистики чувствительны к отклонениям от многомерной нормальности, при которых маргинальные распределения остаются в точности нормальными или близкими к нормальным;

(d) для каких типов отклонений от нулевой гипотезы каждый из критериев особенно чувствителен.

Решение

Классификация статистик критерия для проверки согласия с гипотезой одномерного нормального распределения с неизвестными средним и дисперсией довольно произвольна. Все эти статистики критериев являются функциями порядковых статистик или, что эквивалентно, эмпирической функцией распределения. Однако удобно сгруппировать их следующим образом:

(a) меры согласия группированных наблюденных частот с теоретическими нормальными частотами (часто), как, например, хи-квадрат статистика:

$$\sum \frac{(\text{набл. част.} - \text{предполагаемые част.})^2}{\text{предполагаемые част.}}$$

(b) меры расстояния между эмпирической функцией распределения и проверяемой теоретической функцией распределения $\Phi\{(y-\tilde{\mu})/\tilde{\sigma})\}$, где $\tilde{\mu}$ и $\tilde{\sigma}$ — некоторые оценки μ и σ, как, например, $\bar{y}_.$ и $\sqrt{MS}$;

(c) меры согласия простых общих характеристик, как, например, третий и четвертый кумулянты, с их теоретическими значениями;

(d) меры линейности графика вариационного ряда, когда значения порядковых статистик наносятся против значений их ожиданий. Примером такой меры может служить квадрат коэффициента корреляции, полученный для такого графика (см. Шапиро и Уилк, 1965).

Распределение этих статистик, в принципе, можно найти, рассчитывая их условные распределения при фиксированных значениях достаточных статистик $\bar{Y}_.$ и MS. Однако в практических приложениях более необходимыми представляются аппроксимации для больших выборок. Шапиро и др. (1968) провели обширные исследования по моделированию чувствительности этих критериев. В своих выводах они склоняются к использованию статистик типа (d).

Одно важное соображение связано с тем, в какой мере результат применения критерия помогает указать тип отклонения от H_0, если таковое имеется. С этой точки зрения предпочтительнее использовать наиболее значимую из нескольких статистик, измеряющих различные типы отклонений. Графический анализ является заведомо существенным дополнением ко всем критериям (см. задачу 3.12). При обсуждении практической значимости задачи нахождения точного распределения статистик критерия должен возникать основной вопрос о связи моделей и реальных ситуаций.

Все перечисленные выше критерии можно приспособить для проверки некоторых характеристик многомерной нормальности. Типов отклонений от многомерной нормальности так много, что было бы разумно попытаться явно сформулировать, какие характеристики многомерного нормального распределения мыслятся первостепенно важными.

Один из общих подходов состоит в том, чтобы проверять маргинальные распределения или в исходной системе координат, или в координатной системе главных компонент. При этом предлагается найти распределение при нулевой гипотезе некоторых подходящих сложных статистик, например таких, как наиболее значимая из всех маргинальных статистик. Основная трудность здесь состоит в нахождении распределения выбранной статистики при нулевой гипотезе.

Второй подход состоит в том, чтобы использовать статистики, которые нельзя сформировать непосредственно из комбинаций одномерных статистик. Упомянутый выше метод (a) применим лишь в том случае, когда размерность весьма мала. Метод (b), по-видимому, не исследовался. Метод (c) рассматривался Мардиа (1971). Наиболее плодотворный подход для графического анализа связан, по-видимому, с методом (d). Простая процедура (Хили, 1968; Кокс, 1968) основана на вычислении для каждого отдельного наблюдения, входящего в данные, величины $d_j = (y_j - \tilde{\mu})^T \tilde{\Sigma}^{-1} (y_j - \tilde{\mu})$, где $\tilde{\mu}$ и $\tilde{\Sigma}$ — оценки вектора средних и ковариационной матрицы, с последующим нанесением упорядоченных по величине значений d_j против среднего значения соответствующей порядковой статистики хи-квадрат распределения с p степенями свободы. Можно определить прямой аналог мер, предложенных для метода (d), однако, по-видимому, ничего не известно об их распределениях при нулевой гипотезе.

[Теоретическая статистика, § 3.3; Эндрюс и др., 1971, 1973; Гнанадесикан, 1977)]

3.9. Для проверки согласия данных с гипотезой о том, что п. р. в. имеет вид $\tau^{-1}g(y/\tau)$, $(y \geqslant 0)$, τ неизвестно, предложите удобный график упорядоченных значений. Как можно было бы интерпретировать прямую линию, не проходящую через начало координат? Предложите статистику критерия для проверки такого эффекта и покажите, как найти ее точное распределение при нулевой гипотезе, когда $G(.)$—показательное распределение.

Решение

Из вида плотности r-й порядковой статистики $Y_{(nr)}$ следует, что $E(Y_{(nr)}) = \tau g_{nr}$, где

$$g_{nr} = \frac{n!}{(r-1)!\,(n-r)!} \int x \{G(x)\}^{r-1} \{1 - G(x)\}^{n-r} g(x)\,dx.$$

Таким образом, если форма g выбрана верно, то график значений $Y_{(nr)}$, нанесенных против g_{nr}, должен образовывать приблизительно прямую линию. Отметим, что наклон графика будет оценивать τ. Если в данных присутствуют подозрительно большие значения, тогда для оценки τ можно использовать наклон графика на начальном участке вариационного ряда.

Если прямая линия графика проходит через точку $(\alpha, 0)$, то отсюда вытекает, что п.р.в. имеет вид $\tau^{-1}g\{(y-\alpha)/\tau\}$ $(y \geqslant \alpha)$, т. е. носитель меры распределения сдвинут на α. Наличие такого сдвига естественно проверять на основе значений величины $Y_{(n1)}$. Рассматривая распределения, условные относительно достаточной статистики $Y_1 + \ldots + Y_n$, в случае показательного распределения можно построить критерий, не зависящий от τ (см, задачу 5.5). Чтобы получить условное распределение при нулевой гипотезе $H_0: \alpha = 0$, используем хорошо известный факт, что для показательно распределенных Y_j

$$Y_{(nr)} = \frac{W_1}{n} + \ldots + \frac{W_r}{n-r+1},\quad r = 1, \ldots, n,$$

где $W_1, \ldots, W_n$ — н. о. р. показательные случайные величины со средним τ. Таким образом, требуемое условное распределение величины $Y_{(n1)}$ при заданном $Y_1 + \ldots + Y_n = s$ совпадает с условным распределением W_1/n при заданном $W_1 + \ldots + W_n = s$. Последнее же распределение оказывается ничем иным, как распределением минимума из $n-1$ независимых случайных величин, равномерно распределенных на $(0, s/n)$. Отсюда получаем

$$\mathrm{pr}\,(Y_{(n1)} \geqslant y_{(n1)} \mid \Sigma Y_j = s;\ H_0) = (1 - ny_{(n1)}/s)^{n-1}.$$

При оценивании значимости необходимо сконцентрировать внимание на верхнем хвосте распределения, поскольку отрицательные значения для $y_{(n)}$ с определенностью опровергают гипотезу H_0.

[Теоретическая статистика, § 3.3;
Секхэм, 1936]

3.10. Пусть $\varkappa_1$, $\varkappa_2$ и $\varkappa_3$ — первые три кумулянта для гамма распределения с неизвестными средним и параметром формы. Покажите, что $\varkappa_3\varkappa_1 = 2\varkappa_2^2$. Исходя из этого, предположите статистику критерия для проверки согласия данных с гипотезой гамма распределения. Подумайте, какие методы могут быть использованы при определении распределения статистики при нулевой гипотезе.

Решение

Для плотности вероятностей $\alpha(\alpha y)^{\beta-1}e^{-\alpha y}/\Gamma(\beta)$ производящая функция кумулянтов $\log E\{\exp(sY)\} = \beta\log(1-s/\alpha)$. Разлагая в ряд по степеням s, получим

$$\varkappa_1 = \beta/\alpha,\ \varkappa_2 = \beta/\alpha^2,\ \varkappa_3 = 2\beta/\alpha^3,$$

откуда и вытекает соотношение $\varkappa_3\varkappa_1 = 2\varkappa_2^2$. Его можно записать также следующим образом:

$$\gamma_1 = \varkappa_3/\varkappa_2^{3/2} = 2\varkappa_2^{1/2}\varkappa_1^{-1} = 2\gamma_0,$$

где γ_0 — коэффициент вариации, т. е. стандартное уклонение, деленное на среднее. Для логнормального распределения $\gamma_1 = 3\gamma_0 + \gamma_0^3$, тогда как для распределения Вейбулла справедливо совершенно другое соотношение. Таким образом, если для нескольких совокупностей данных в плоскости с координатами γ_1 и γ_0 нанести точки $(\bar\gamma_0, \bar\gamma_1)$, где $\bar\gamma_0$ и $\bar\gamma_1$ — оценки γ_0 и γ_1, то близость к соответствующей кривой укажет, какое из основных двухпараметрических семейств распределений на $(0, \infty)$ согласуется с данными, если такое согласие вообще имеется. Рассмотренный прием аналогичен использованию графика величины $\bar\gamma_2$ в зависимости от $\bar\gamma_1$ для семейств, определенных на $(-\infty, \infty)$ (Пирсон и Хартли, 1970, стр. 87).

На пути превращения высказанных выше соображений в критерий значимости возникают определенные затруднения. Первый шаг заключается в том, чтобы исследовать статистики $\bar\gamma_1 - 2\bar\gamma_0$ или $\log\left(\frac{1}{2}\bar\gamma_1/\bar\gamma_0\right)$. При нулевой гипотезе средние этих статистик равны приблизительно нулю. Для эмпирических оценок кумулянтов выборочные моменты известны вплоть до высоких порядков (см. Кендалл и Стьюарт, т. I, гл. 12). Поэтому для рассмотренных выше статистик можно приближенно оценить стандартные ошибки и использовать нормальную аппроксимацию. Для получения более точных распределений при нулевой гипотезе, по-видимому, более полезны методы моделирования.

Совсем другой подход состоит в том, чтобы при нулевой гипотезе перейти к условным распределениям относительно достаточной статистики (ΣY_j, $\Sigma \log Y_j$), исключая тем самым мешающие параметры α и β. Одна возможность при этом состоит в том, чтобы рассмотреть условное распределение статистики ΣY_j^2 при заданных $\Sigma Y_j = s_1$ и $\Sigma \log Y_j = s_2$. Однако оказывается, что трудно работать с точным условным распределением статистики ΣY_j^2. Поэтому вновь необходимо использовать аппроксимации, основанные на большом объеме выборки.

Общее заключительное замечание состоит в том, что содержательные утверждения о форме распределения можно обычно сделать только после анализа нескольких аналогичных совокупностей данных, полученных при довольно различных условиях.

[Теоретическая статистика, § 3.3]

3.11. Покажите, как можно построить для биномиального, геометрического и общего однопараметрического экспоненциального семейства распределений точные критерии типа показателя рассеяния. Объясните, почему при соответствующим образом сформулированных условиях хи-квадрат распределение будет хорошей аппроксимацией распределения статистики при нулевой гипотезе в первом случае, плохой — во втором и вообще не обязательно хорошей — в третьем случае.

Решение

Рассмотрим дискретную случайную величину, имеющую плотность из однопараметрического экспоненциального семейства распределений, заданного в естественной форме:

$$f_Z(z; \varphi) = \exp\{-z\varphi + c^\dagger(\varphi) + d^\dagger(z)\}.$$

Для n н. о. р. наблюдений совместная вероятность при гипотезе H_0 примет вид

$$\exp\{-\varphi \Sigma z_j + nc^\dagger(\varphi)\} \exp\{\Sigma d^\dagger(z_j)\}.$$

Достаточной статистикой для мешающего параметра φ будет $S = \Sigma Z_j$. Условное же распределение при заданном $S = \Sigma Z_j$

будет определяться вероятностями

$$\operatorname{pr}\{Z_j = z_j,\ j = 1, \ldots, n \mid S = s; H_0\} = \frac{\exp\{\Sigma d^{\dagger}(z_j)\}}{\sum\limits_{z=(z_1, \ldots, z_n): \Sigma z_k = s} \exp\{d^{\dagger}(z_k)\}}. \quad (1)$$

Возьмем теперь любую статистику критерия, измеряющую рассеяние, например $\Sigma (Z_j - \bar{Z})^2$. В принципе, точное H_0 распределение этой статистики определяется вероятностями (1).

В задаче 2.13 было показано, что $\operatorname{var}(Z) = -(c^{\dagger}(\varphi))''$ и $E(Z) = (c^{\dagger}(\varphi))'$. Обозначим эти функции $v(\varphi)$ и $b(\varphi)$ соответственно. Таким образом, $v\{b^{-1}(S/n)\}$ является оценкой для $v(\varphi)$, и можно использовать эту вычисленную на основе теоретической модели оценку, чтобы нормировать непосредственную оценку для $\operatorname{var}(Z)$. Тем самым получим статистику типа показателя рассеяния

$$\Sigma (Z_j - \bar{Z})^2 / v\{b^{-1}(S/n)\}. \quad (2)$$

Статистику (2) можно рассматривать приближенно как скорректированную длину квадратов n н. о. р. случайных величин, имеющих при гипотезе H_0 единичную дисперсию. Если случайные величины Z_j распределены приближенно нормально, то можно ожидать, что хи-квадрат распределение с $(n-1)$-й степенью свободы будет хорошей аппроксимацией для распределения статистики (2). Если случайные величины Z_j имеют распределение Пуассона с большим средним, или биномиальное с большим средним, то случайные величины Z_j и на самом деле приближенно нормальны, а следовательно, хи-квадрат аппроксимация будет справедлива. В то же время, если случайные величины Z_j геометрически распределены, то легко видеть, что рассеяние статистики (2) во много раз сильнее, чем рассеяние хи-квадрат случайной величины. Это вытекает по-существу из того, что для геометрически распределенных наблюдений четвертый момент значительно больше квадрата дисперсии. Распределение выборочной дисперсии будет, таким образом, отражать большую разбросанность значений, чем это можно ожидать в случае нормально распределенных величин.

[Теоретическая статистика, § 3.3]

3.12. Для графических методов анализа предполагаются выполненными следующие желаемые свойства:

(а) соответствие с простой моделью должно отображаться прямой линией графика или там, где непосредственный интерес представляет случайная часть изменчивости, полностью случайным рассеянием;

(b) нанесенные на график точки должны быть подвержены независимым ошибкам;

(c) координаты всех точек должны иметь одинаковую точность;

(d) нелинейные преобразования должны быть такими, чтобы они акцентировали внимание на областях значений параметров, представляющих непосредственный интерес.

Обсудите в свете всех этих требований различные графические методы для исследования согласия с гипотезой нормального распределения.

Решение

Графические методы важны как на начальном этапе анализа данных, так и в окончательном представлении выводов. В то же время трудно провести систематическое обсуждение и регламентацию использования графических методов. Два исходных принципа формулируются следующим образом. Во-первых, систематические связи лучше представлять обычно в виде линейной зависимости. Во-вторых, там, где исследуется наличие отклонения от случайной структуры и где нельзя построить график линейной зависимости, необходимо использовать график ожидаемого полностью случайного рассеяния. Отклонения от этих двух графических моделей относительно легко видны глазом. Сказанное характеризует требование (a) графического метода анализа.

Если точки наносятся с ощутимыми коррелированными ошибками, то, исходя из общих соображений, можно ожидать систематические отклонения от „истинной“ модели, и необходимо это учитывать при интерпретации результатов. Аналогично если точки наносятся на график с неодинаковой точностью, то это тоже должно учитываться при интерпретации. Иногда это можно сделать, принимая в расчет величину стандартной ошибки, отвечающую каждой точке. Так, например, иногда варьируют определенным образом размеры нанесенных на график точек. Это все, что можно сказать о требованиях (b) и (c) графического метода анализа. Последнее требование (d) существенно тем, что если имеется практически важная шкала, то она должна быть отражена в графическом методе. К сожалению, часто требования (a)—(d) оказываются до некоторой степени противоречивыми.

Среди многих графических приемов, используемых при проверке нормальности, упомянем следующие:

(i) порядковые статистики выборки наносятся против ожиданий соответствующих им порядковых статистик нормального $N(0, 1)$ распределения (см. задачу 3.8) или против некоторых аппроксимаций для таковых ожиданий, как, например, $\Phi^{-1}\{(r-3/8)/(n+1/4)\}$ (см. Теоретическая статистика, Дополнение 2);

(ii) в качестве гипотетической функции распределения (ф. р.) принимается оценка $\Phi\{(y-\tilde{\mu})/\tilde{\sigma}\}$, где $\tilde{\mu}$ и $\tilde{\sigma}$ — оценки параметров μ и σ, а затем принятая кривая и эмпирическая ф. р. вычерчиваются вместе;

(iii) используется метод (ii), но наносится на график разность между теоретической ф.р. и эмпирической ф.р. для соответствующей последовательности значений аргумента, причем возможно шкалирование разностей на основе оценок их стандартных ошибок;

(iv) чертится логарифм ординат гистограмм, когда в качестве абсциссы берется $(y-\tilde{\mu})^2$;

(v) на одном и том же графике наносится принятая в качестве истинной плотность нормального распределения $\tilde{\sigma}^{-1}\varphi\{(y-\tilde{\mu})/\tilde{\sigma}\}$ и выборочная гистограмма;

(vi) строится график разности плотности распределения $\tilde{\sigma}^{-1}\varphi\{(y-\tilde{\mu})/\tilde{\sigma}\}$ и гистограмма. При этом возможно использование некоторой стандартизованной формы графика. Можно, например, делить разность на квадратный корень из принятой плотности или можно работать с разностями квадратных корней частот.

(vii) выборочные значения посредством преобразования $\Phi^{-1}\{(y_j-\tilde{\mu})/\tilde{\sigma}\}$ превращаются в приближенно равномерно распределенные случайные величины. После этого применяются методы, связанные с равномерным распределением.

Среди рассмотренных графических приемов нельзя раз и навсегда указать один наилучший, хотя метод, приведенный в п. (i), по-видимому, самый лучший для использования в общей ситуации. Сравнения принятых кривых лучше производить, по-видимому, на основе нормированных разностей. Однако какой бы метод ни использовался, для его успешного применения необходимы некоторые представления о влиянии уклонений от нормальности. Чтобы закончить обсуждение графических методов анализа, полезно построить несколько графиков для имеющихся в наличии смоделированных данных.

[Теоретическая статистика, § 3.4; Кокс, 1978]

3.13. В пяти испытаниях первые четыре дают нуль, а последнее единицу. Нулевая гипотеза заключается в том, что это испытание Бернулли с исходами 0 и 1 и с вероятностью появления единицы, равной 1/2. Требуется проверить значимость преобладания нулей. Докажите, что если число испытаний фиксировано, т. е. число единиц имеет биномиальное распределение, то уровень значимости равен 3/32. В то же время если испытания продолжаются до первого появления единицы, т. е. число испытаний имеет геометрическое распределение, то уровень значимости равен 1/16. Обсудите все это с учетом усиленного принципа правдоподобия, в соответствии с которым правдоподобие должно использоваться непосредственно. Заметьте при этом, что для постановки задачи, когда в явном виде формулируется только нулевая гипотеза, предпосылки для возникающего противоречия отсутствуют.

Решение

Если число испытаний фиксировано, то вероятности значений случайной величины Y' — числа нулей в последовательности испытаний выражаются в следующем виде:

$$\mathrm{pr}(Y' = r;\ H_0) = 2^{-5} 5!/\{r!\,(5-r)!\}.$$

Таким образом, уровень значимости при проверке возможности преобладания нулей равен

$$\mathrm{pr}(Y' = 4;\ H_0) + \mathrm{pr}(Y' = 5;\ H_0) = 3/16.$$

Если испытания продолжают до первого успеха, то число испытаний Y'' — случайная величина, для которой $\mathrm{pr}(Y'' = r;\ H_0) = 2^{-r}$. Уровень значимости в этом случае

$$\sum_{r=5}^{\infty} \mathrm{pr}(Y'' = r;\ H_0) = 1/16.$$

Правдоподобие можно использовать лишь тогда, когда точно сформулированная модель содержит более одного распределения. Поскольку в рассматриваемой ситуации имеется лишь одно распределение, то функция правдоподобия не определена. Тем самым нет возможности провести какие-либо сравнения с принципом правдоподобия. Чтобы ввести функцию правдоподобия, необходима еще хотя бы одна вероятностная модель. Одна из возможностей состоит в предположении, что испытания независимы и единица появляется с вероятностью θ, отличной от единицы. Далее, если правило остановки или одно из указанных выше, или на самом деле любое другое, для которого вероятность прекращения испытаний зависит только от результатов предыдущих испытаний, то функция правдоподобия для полученной выборки имеет вид $(1-\theta)^4\theta$. В соответствии с усиленным принципом правдоподобия для двух указанных выше схем проведения испытаний должны быть сделаны одни и те же выводы. Однако, если в дополнение к H_0 рассмотрим ровно одну модель, а именно, что в первом, четвертом испытаниях обязательно появляется нуль, а в пятом обязательно появляется единица, то положение существенно меняется. Таким образом, конкретный выбор модели является решающим.

Принятие усиленного принципа правдоподобия означало бы, что содержание гл. 4—9 этой книги можно было бы в значительной степени не рассматривать. Наше отношение к этому можно выразить следующим образом:

(a) действительно было бы удобно, если бы не было необходимости задавать правило остановки;

(в) доводы, выдвинутые при выводе усиленного принципа правдоподобия из более простых требований, не являются неотразимыми (Теоретическая статистика, стр. 50; Кэлбфлейш, 1975), так же как не являются таковыми и доводы в пользу чисто байесовского подхода (гл. 10);

(c) в усиленном принципе правдоподобия не указывается общих методов использования правдоподобия, и в этом его недостаток.

[Теоретическая статистика, § 2.3; 2.4; 3.4 (ii)]

3.14.[1]) При проверке гипотезы H_0, когда нет четких альтернатив, считается, что решение о выборе наиболее подходящей среди ряда статистик можно сделать, опираясь только на их выборочные распределения, вычисленные в предположении справедливости гипотезы H_0. Общее представление о назначении критериев значимости сводится в этом случае к частному утверждению, что если две статистики имеют одно и то же выборочное распределение при H_0, тогда они одинаково эффектны для проверки H_0. Чтобы быть конкретными, рассмотрим гипотезу H_0, что заданные наблюдения $y_1, \ldots, y_n$ являются реализациями н.о.р. случайных величин $Y_1, \ldots, Y_n$, которые нормально $N(0, \sigma^2)$ распределены. Относительно рассматриваемой ситуации полагают, что t-статистика Стьюдента: $T = \sqrt{n}\,\bar{Y}_. / \{\sum(Y_j - \bar{Y}_.)^2/(n-1)\}^{1/2}$, является подходящей статистикой для проверки H_0. Опровергните сформулированное выше утверждение, рассмотрев статистику

$$T' = \frac{(n-1)^{1/2} \Sigma a_j Y_j}{\{\Sigma Y_j^2 - (\Sigma a_j Y_j)^2\}^{1/2}},$$

где $a_j = y_j / \{\Sigma y_k^2\}^{1/2}$ $(j = 1, \ldots, n)$.

---

[1]) Улучшенный вариант.

Решение

Оно связано с обсуждением абсолютных критериев значимости в задаче 3.3. Там было отмечено, что при применении абсолютных критериев значимости к непрерывным данным необходим некоторый критерий выбора шкалы измерений. Рассматриваемый пример иллюстрирует несколько иную черту статистик критериев. А именно из н.о.р. нормальных $N(0, \sigma^2)$ случайных величин можно построить много статистик, имеющих t-распределение Стьюдента.

Для заданных a_j ортогональным преобразованием можно перейти к новым величинам $U_1, \ldots, U_n$, для которых статистика

$$T' = \frac{U_1}{\left\{\sum_{j=1}^{n} U_j^2/(n-1)\right\}^{1/2}}$$

будет иметь то же самое t-распределение Стьюдента, что и статистика T. Заметим, однако, что если $Y = y$, то $T' = \infty$. Другими словами, в некотором смысле каждая выборка по-своему экстремальна. В то же время априорно выбираемая форма статистики T должна отражать возможные отклонения от нулевой гипотезы, а не предопределяться характеристиками своего распределения при нулевой гипотезе.

[Теоретическая статистика, § 3.4 (ii); Нейман, 1952]

3.15. Предположим, что при выборе наиболее значимой из ряда зависимых статистик известны попарные распределения статистик. Покажите, как можно использовать формулу вероятности объединения событий, чтобы дополнить упомянутую в разд. „Необходимые сведения“ верхнюю границу для уровня значимости нижней границей, учитывающей специфику задачи выбора статистики.

Решение

Отождествим B_j с событием $P_j = q_{\text{набл}}$ и воспользуемся цепочкой неравенств

$$\Sigma \operatorname{pr}(B_j) - \sum_{j>k} \operatorname{pr}(B_j \cap B_k) \leqslant \operatorname{pr}(B_1 \cup \ldots \cup B_m) \leqslant \Sigma \operatorname{pr}(B_j)$$

(см. Феллер, 1968, § IV.5). Правая часть неравенств дает общую верхнюю границу Бонферрони для вероятности объединения событий, а именно $mq_{\text{набл}}$. В то же время для нижней границы такого простого общего выражения установить не удается, поскольку оно зависит от конкретных форм совместных распределений величин (P_j, P_k). Когда статистики независимы, точный уровень значимости равен $1-(1-q_{\text{набл}})^m$, а нижняя граница будет равна $mq_{\text{набл}}-\frac{1}{2}m(m-1)q^2_{\text{набл}}$.

[Теоретическая статистика, § 3.4 (iii)]

3.16. Предположим, что m независимых совокупностей данных объема $n_1, \ldots, n_m$ проверяются на предмет совместимости с гипотезой H_0. Имеется в виду альтернатива конкретного типа. Пусть результаты критериев значимости представляются вероятностями значимости $P_1, \ldots, P_m$. Исследуйте сравнительные достоинства суммарных статистик

$$T_1 = -2\Sigma \log P_j, \quad T_2 = \max(P_j),$$
$$T_3 = \min(P_j), \quad T_4 = \Sigma n_j P_j / \Sigma n_j,$$

имея в виду следующие возможности: (i) ожидается, что H_0 не верна самое большое для одной из совокупностей данных;

(ii) ожидается, что H_0 справедлива или для всех, или ни для каких совокупностей данных. Потенциально могут возникать и любые другие возможности.

Решение

Если справедлива нулевая гипотеза, то при нахождении распределений всех перечисленных выше статистик затруднений, в принципе, не возникает, поскольку $P_1, \ldots, P_m$ независимы и равномерно распределены на (0, 1). Например, $-\log P_j$ — показательно распределенная случайная величина с единичным средним, поэтому T_1 имеет хи-квадрат распределение с $2m$ степенями свободы.

Приведем некоторые соображения, имеющие отношение к обсуждению свойств статистик критериев значимости:

(i) возможна лишь косвенная интерпретация статистик, поскольку они не связаны с задачей оценивания;

(ii) если необходимо выявить несколько относительно малых отклонений от H_0, то предпочтительнее использовать взвешивание статистик, подобно тому как это сделано при построении статистик T_4. Однако определение точных значений весов требует более развернутой постановки задачи;

(iii) статистики T_1 и T_3 целесообразно употреблять в том случае, когда весьма вероятно, что гипотеза H_0 не верна только для одной совокупности данных, причем отклонение является весьма заметным. В то же время статистика T_4 чувствительна к воздействию на все множества данных любого отклонения от H_0.

Все выше перечисленные статистики, а в особенности T_1, имеют долгую историю. Статистика T_1, по-видимому, соответствует частному случаю процедуры, предложенной Фишером. Общая процедура состоит в том, что сначала определяется X_j как верхняя P_j-квантиль хи-квадрат распределения с r_j степенями свободы, а после этого определяется общая значимость как вероятность превышения хи-квадрат случайной величиной с Σr_j степенями свободы значения ΣX_j.

Подробный анализ предложенных статистик потребовал бы привлечения соображений, связанных с понятиями мощности и эффективности критериев, вводимых в гл. 4. Один общий результат относительно асимптотической мощности приведен в работе Литтела и Фолкса (1979).

[Теоретическая статистика, § 3.4 (iii)]

4. Критерии значимости: простые нулевые гипотезы

Необходимые сведения

В гл. 3 предполагалось, что явно формулируется только нулевая гипотеза H_0. В настоящей главе будем считать, что H_0 — простая гипотеза, полностью определяющая плотность $f_0(y)$. В дополнение к этому будем считать, что имеются одна или более альтернативных гипотез H_A, представляющих те направления отклонений от H_0, наличие которых желательно обнаружить. Итак, необходимо дать обоснованный ответ на вопрос, имеются ли убедительные свидетельства в пользу несоответствия данных с гипотезой H_0. При этом особый интерес представляют отклонения в направлении H_A.

Предположим сначала, что H_A — простая гипотеза, в которой плотность определяется как $f_A(y)$. Тогда оптимальный критерий значимости задается следующим образом. Обозначим w_α множество возможных исходов наблюдений, которые следует рассматривать как значимо противоречащие гипотезе H_0 на уровне, не большем α. Вероятностная интерпретация величины α состоит в том, что область w_α имеет размер α, т. е. $\operatorname{pr}(Y \in w_\alpha;\ H_0) = \alpha$. Область w_α называют в этом случае критической областью размера α. Оптимальность достигается максимизацией мощности $\operatorname{pr}(Y \in w_\alpha;\ H_A)$.

Основополагающий результат для непрерывно распределенных случайных величин заключается в том, что оптимальный критерий в качестве критических областей имеет критические области отношения правдоподобия, имеющие вид $\{y: \operatorname{lr}_{A0}(y) = f_A(y)/f_0(y) \geqslant c_\alpha\}$ для соответствующим образом выбранного c_α. Это утверждение и есть лемма Неймана — Пирсона. Сформулированный результат справедлив и в дискретном случае, только уровень α не всегда достигается точно.

Для большинства прикладных задач характерно наличие более одной альтернативы. Так, мы можем иметь однопараметрическое семейство вероятностных моделей, для которых, например, $H_0: \theta = \theta_0$, а $H_A: \theta > \theta_0$ (односторонние альтернативы), или $H_A: \theta \neq \theta_0$ (двусторонние альтернативы). Если одна и та же область w_α оптимальна для всех простых альтернатив, составляющих H_A, то критерий и критическая область носят название равномерно наиболее мощных. Этим свойством обладают некоторые односторонние критерии, связанные с проверкой гипотез о значении параметра однопараметрического семейства экспоненциального типа.

Функция мощности произвольного критерия определяется как $\mathrm{pr}(Y \in w_\alpha;\ H_A)$ и считается при фиксированном α функцией параметра распределений, составляющих H_A.

Из общих соображений, упомянутых в гл. 2, вероятностные вычисления размера критерия можно проводить в принципе на основе обращения к условным распределениям при заданных значениях соответствующей подчиненной статистики.

Двусторонние критерии будут рассматриваться, как правило, в виде некоторой функции двух односторонних критериев. В то же время иногда при исследовании двусторонних критериев требуется, чтобы функция мощности имела минимум при H_0. Критерий, удовлетворяющий такому требованию, называют несмещенным.

Довольно часто равномерно наиболее мощного критерия не существует. В этом случае разумно максимизировать мощность для альтернатив, близких к H_0. Пусть распределение зависит от скалярного параметра θ, а нулевая гипотеза задается как $\theta = \theta_0$. Тогда, разлагая отношение правдоподобия, получим

$$\log f_Y(y;\ \theta_0 + \delta) - \log f_Y(y;\ \theta_0) = u_.(y;\ \theta_0)\,\delta + o(\delta),$$

где $u_.(y;\ \theta_0) = [\partial \log f_Y(y;\ \theta)/\partial\theta]_{\theta=\theta_0}$. Таким образом, при проверке нулевой гипотезы против односторонних локальных альтернатив критические области определяются величиной $u_.(y;\ \theta_0)$, так что ее положительные значения значимы при $\delta > 0$, а отрицательные — при $\delta < 0$.

Статистику $u_.(Y;\ \theta_0) = U_.(\theta_0)$ называют эффективным вкладом. Ее основные свойства таковы:

(a) $E(U_.(\theta_0);\ \theta_0) = 0$;

(b) $\mathrm{var}\{U_.(\theta_0);\ \theta_0\} = i_.(\theta_0) = E\{-\partial^2 \log f_Y(Y;\ \theta_0)/\partial\theta^2;\ \theta_0\}$, называемая информацией по Фишеру;

(c) $E\{U_.(\theta_0 + \delta);\ \theta_0\} = i_.(\theta_0)\,\delta + o(\delta)$;

(d) когда Y образуется из независимых компонент, величины $U_.$ и $i_.$ представляются в виде сумм вкладов этих компонент; тем самым нахождение распределения величины $U_.$ упрощается;

(e) если θ преобразуется в $\varphi = \varphi(\theta)$, то эффективный вклад и информация, вычисленные по новому параметру, примут вид

$$\frac{d\theta_0}{d\varphi_0} U_.(\theta_0), \quad \left(\frac{d\theta_0}{d\varphi_0}\right)^2 i_.(\theta_0).$$

Когда θ — векторный параметр, то эффективный вклад и информация по Фишеру преобразуются в вектор и матрицу соответственно.

При построении критериев для проверки многомерных альтернатив H_A необходимо условиться о некотором компромиссе. Дело в том, что высокая мощность по разным направлениям в параметрическом пространстве может достигаться для различных статистик критерия. Одна из возможностей связана с тем, чтобы в качестве статистики критерия брать некоторую подходящую квадратичную форму от компонент вектора эффективного вклада, например

$$\{U_.(\theta_0)\}^T \{i_.(\theta_0)\}^{-1} \{U_.(\theta_0)\}.$$

Задачи

4.1. Сформулируйте проблему выбора наилучшей критической области размера α при проверке гипотезы $H_0: \theta = \theta_0$ против $H_A: \theta = \theta_A$ как задачу вариационного исчисления при заданных ограничениях. Примените уравнение Эйлера, чтобы получить лемму Неймана — Пирсона.

Решение

Для безусловной максимизации интеграла

$$\int_w \{f_A(y) = kf_0(y)\}\, dy \tag{1}$$

применим метод множителей Лагранжа. Пусть z — координата, определяющая границу между областью w и ее дополнением. Деформируем теперь границу так, чтобы новая граница отстояла от z по нормали к w на расстояние $\varepsilon\varphi(z)$. Тогда интеграл (1) изменится по величине на

$$\varepsilon \int \varphi(z) \{f_A(z) - kf_0(z)\}\, dz + O(\varepsilon^2). \tag{2}$$

Отметим, что интеграл (2) надо рассматривать как интеграл по поверхности. Далее, для нахождения стационарного значения интеграла α (1) необходимо, чтобы интеграл в (2) обращался в нуль. Для произвольных $\varphi(z)$ это будет только в том случае, когда $f_A(z) = kf_0(z)$ на границе области w.

Отметим, что в приведенном доказательстве показано только существование стационарного значения. Обычно приводится более длинное, но в то же время более элементарное доказательство, являющееся в действительности решением частного случая задачи 4.2. При этом показывается, что максимум и на самом деле достигается.

[Теоретическая статистика, § 4.3; Леман, § 3.2; Рао, § 7а, Силвей, стр. 98]

4.2. Обобщенный нерандомизированный вариант леммы Неймана — Пирсона формулируется следующим образом. Пусть $k_1(y), \ldots, k_{m+1}(y)$ — интегрируемые функции на выборочном пространстве, которые не обязательно являются плотностями.

Тогда область w, которая максимизирует

$$\int_w k_{m+1}(y)\,dy$$

при заданных ограничениях

$$\int_w k_j(y)\,dy = a_j \quad (j=1,\ \ldots,\ m),$$

определяется как $\{y : k_{m+1}(y) \geqslant c_1 k_1(y) + \ldots + c_m k_m(y)\}$ при условии, что такие постоянные $c_1,\ \ldots,\ c_m$ существуют. Докажите этот результат и используйте его при нахождении вида критических областей локально наиболее мощных односторонних и двусторонних критериев для одномерного параметра.

Решение

Пусть w' — оптимальная область, а w'' — любая другая область и обе они удовлетворяют равенствам

$$\int_w k_j(y)\,dy = a_j \quad (j=1,\ \ldots,\ m).$$

Обозначив $\overline{w}$ дополнение области w, получим представление для рассматриваемых областей

$$w' = (w' \cap \overline{w''}) \cup (w' \cap w'') \quad \text{и} \quad w'' = (w'' \cap \overline{w'}) \cap (w' \cap w'').$$

Отсюда для $j-1,\ \ldots,\ m$

$$\int\limits_{w' \cap w''} k_j(y)\,dy = \int\limits_{w' \cap w''} k_j(y)\,dy = a_j - \int\limits_{w' \cap w''} k_j(y)\,dy.$$

Предположим далее, что в w'

$$k_{m+1}(y) \geqslant \sum_{j=1}^{m} c_j k_j(y),$$

а для точек $\overline{w'}$ выполняется обратное неравенство. Получим

$$\int\limits_{w' \cap \overline{w''}} k_{m+1}(y)\,dy \geqslant \sum c_j \int\limits_{w' \cap \overline{w''}} k_j(y)\,dy =$$

$$= \sum c_j \int\limits_{\overline{w'} \cap w''} k_j(y)\,dy \geqslant \int\limits_{\overline{w'} \cap w''} k_{m+1}(y)\,dy.$$

Поскольку область w'' произвольна, то, добавляя к обеим сторонам последнего неравенства величину $\int\limits_{w' \cap w''} k_{m+1}(y)\,dy$, получим требуемый результат.

Определение локально наиболее мощного критерия эквивалентно требованию существования такой критической области w_α, чтобы при выполнении равенства

$$\int_{w_\alpha} f(y;\ \theta_0)\,dy=\alpha$$

производная

$$\frac{d}{d\theta_0}\int_{w_\alpha} f(y;\ \theta_0)\,dy$$

была бы максимальна[1]). При выполнении условий регулярности очередность операций дифференцирования и интегрирования в последнем выражении можно поменять местами. Заметив, что $\partial f(y;\ \theta_0)/\partial\theta_0=u_.(y;\ \theta_0)f(y;\ \theta_0)$, получим, что для выбора критерия требуется максимизировать

$$\int_{w_\alpha} u_.(y;\ \theta_0)f(y;\ \theta_0)\,dy.$$

К рассматриваемой задаче применим полученный выше общий результат. Действительно, положив $k_1(y)=f(y,\ \theta_0)$, $a_1=\alpha$, $k_2(y)=u_.(y;\ \theta_0)$, получим, что оптимальная критическая область имеет вид $w_\alpha=\{y:u_.(y;\ \theta_0)\geqslant c_1\}$.

Для двустороннего критерия потребуем, чтобы функция мощности имела в θ_0 локальный минимум и максимальную кривизну. При выполнении условий регулярности эти требования формализуются следующим образом:

$$\int_{w_\alpha} f(y;\ \theta_0)\,dy=\alpha,\quad \int_{w_\alpha}\{\partial f(y;\ \theta_0)/\partial\theta_0\}\,dy=0,$$

$$\int_{w_\alpha}\{\partial^2 f(y;\ \theta_0)/\partial\theta_0^2\}\,dy=\max.$$

Вновь применяя общий результат, получим оптимальную критическую область в следующем виде:

$$\begin{aligned}w_\alpha=\{y:\partial^2 f(y;\ \theta_0)/\partial^2\theta_0\geqslant c_1 f(y;\ \theta_0)+\\+c_2\partial f(y;\ \theta_0)/\partial\theta_0\}=\{y:\partial u_.(y;\ \theta_0)/\partial\theta_0+\\+u_.^2(y;\ \theta_0)\geqslant c_1+c_2u_.(y;\ \theta_0)\}.\end{aligned}$$

[Теоретическая статистика, § 4.3, 4.7; Леман, § 3.6; Нейман и Пирсон, 1933, 1936, 1967; Рао, § 7а]

---

[1]) Речь идет о построении локально наиболее мощных критериев проверки гипотезы $H_0:\theta=\theta_0$ против альтернатив $H_A:\theta>\theta_0$. — *Прим. перев.*

4.3. Чтобы сравнить две простые гипотезы H_0 и H_1, в каждую единицу времени производится по одному наблюдению. После каждого наблюдения для всех полученных к этому моменту наблюдений вычисляется отношение правдоподобий. Это отношение после m наблюдений обозначается $\operatorname{lr}_{10}(y_1, \ldots, y_m) = \operatorname{lr}_{10}(y^m)$. Выбираются положительные постоянные $a < 1 < b$. Далее поступаем следующим образом:

(i) если $a < \operatorname{lr}_{10}(y^{(m)}) < b$, то берется $(m+1)$-е наблюдение;

(ii) если $a \geqslant \operatorname{lr}_{10}(y^{(m)})$, то дальнейших наблюдений не производится и гипотеза H_0 предпочитается гипотезе H_1;

(iii) если $b \leqslant \operatorname{lr}_{10}(y^{(m)})$, то дальнейших наблюдений не производится и гипотеза H_1 предпочитается гипотезе H_0.

Заключение, состоящее в том, что предпочитаемой гипотезой считается H_i, обозначим α_i $(i=0, 1)$.

Записывая $\operatorname{lr}_{10}(y^{(m)}) = f_1(y^{(m)})/f_0(y^{(m)})$ и суммируя по всем точкам окончания наблюдений в бесконечномерном выборочном пространстве, покажите, что

$$a \operatorname{pr}(d_0;\ H_0) \geqslant \operatorname{pr}(d_0;\ H_1), \qquad b \operatorname{pr}(d_1;\ H_0) \leqslant \operatorname{pr}(d_1;\ H_1).$$

Допустим, что процедура выбора закончится с единичной вероятностью. Покажите, что $a \geqslant \beta/(1-\alpha)$ и $b \leqslant (1-\beta)/\alpha$, где α и β — „вероятности ошибок“ [1]).

При каких условиях можно считать, что $a \approx \beta/(1-\alpha)$, а $b \approx (1-\beta)/\alpha$? Докажите, что если наблюдаются н. о. р. случайные величины, то процедура выбора заканчивается с вероятностью единица. Исследуйте, какой вид принимает процедура для задачи с однопараметрическим экспоненциальным распределением. Покажите возникающую здесь связь со случайными блужданиями.

Критически обсудите ограниченность данной постановки с точки зрения ее практической полезности.

Решение

Пусть $w_j^{(m)}$ — область выборочного пространства величин $y^{(m)} = (y_1, \ldots, y_m)$, для которой предпочтительнее принять H_j на m-м шаге. Тогда

$$\operatorname{pr}(d_0;\ H_j) = \sum_m \int_{w_0^m} f_j(y^{(m)})\, dy^{(m)} \quad (j = 1,\ 2).$$

Для точки $y^{(m)}$ из $w_0^{(m)}$ имеем $f_1(y^{(m)}) \leqslant a f_0(y^{(m)})$. Отсюда, срав-

---

[1]) То есть ошибки первого и второго рода соответственно. — *Прим. перев.*

нивая почленно слагаемые двух рядов, получаем

$$a\,\mathrm{pr}(d_0;\ H_0) \geqslant \mathrm{pr}(d_0;\ H_1)$$

и аналогично

$$b\,\mathrm{pr}(d_1;\ H_0) \leqslant \mathrm{pr}(d_1;\ H_1).$$

Если процедура наблюдений заканчивается с вероятностью единица, т. е. $\mathrm{pr}(d_0;\ H_j)+\mathrm{pr}(d_1;\ H_j)=1$, то $\mathrm{pr}(d_0;\ H_0)=1-\alpha$ и $\mathrm{pr}(d_1;\ H_1)=1-\beta$. Таким образом, $a \geqslant \beta/(1-\alpha)$ и $b \leqslant (1-\beta)/\alpha$.

Неравенства обращаются в приближенные равенства, если процедура выбора оканчивается в точках, близких к границам остановки. Разумно предположить, что именно с такой ситуацией имеем дело, когда H_0 и H_1 весьма близки друг к другу. В этом случае число необходимых наблюдений m обычно оказывается большим, а отношение правдоподобий изменяется соответственно на малые доли.

Пусть элементы выборки — н. о. р. случайные величины. Выбор продолжается после m-го наблюдения, если он уже велся до m-го наблюдения, и

$$\log a < \sum_{j=1}^{m} v_j < \log b, \tag{1}$$

где $v_j = \log\{f_1(y_j)/f_0(y_j)\}$. Таким образом, вероятность продолжения выбора после m-й стадии меньше, чем

$$\mathrm{pr}\left(\log a < \sum_{j=1}^{m} V_j < \log b\right).$$

По центральной предельной теореме эта вероятность стремится к нулю при $m \to \infty$. Условие (1) показывает, что процесс выбора можно представить как случайное блуждание с двумя поглощающими экранами, параллельными оси „времени".

Пусть наблюдения образуются из н. о. р. случайных величин, п. р. в. которых принадлежат семейству распределений экспоненциального типа и имеют естественную форму $\exp\{-z\varphi + c^{\dagger}(\varphi)+d^{\dagger}(z)\}$. Если гипотезы H_0 и H_1 определяются как $\varphi=\varphi_0$ и $\varphi=\varphi_1<\varphi_0$ соответственно, то

$$v_j = (\varphi_0-\varphi_1)\,z_j + c^{\dagger}(\varphi_1) - c^{\dagger}(\varphi_0).$$

Используя это представление, выразим неравенства (1) в форме

$$\frac{\log a}{\varphi_0-\varphi_1} - m\,\{c^{\dagger}(\varphi_1)-c^{\dagger}(\varphi_0)\} \leqslant \sum_{j=1}^{m} z_j \leqslant \frac{\log b}{\varphi_0-\varphi_1} - m\,\{c^{\dagger}(\varphi_1)-c^{\dagger}(\varphi_0)\},$$

соответствующей двум параллельным линейным поглощающим экранам, когда $\sum_{j=1}^{m} z_j$ наносится на график как значение ординаты для абсциссы m.

Хотя имеется обширная литература, в которой развиваются высказанные выше соображения, однако все эти работы имеют по крайней мере два существенных недостатка. Первый состоит в том, что рассматриваются только две простые гипотезы, а не непрерывные области потенциально возможных гипотез. Второй недостаток связан с тем, что выборочная схема „открыта“, т. е. верхняя граница необходимого числа наблюдений заранее не определяется. Относительно практически используемых выборочных схем см. монографию Уэзерилла (1975).

[Теоретическая статистика, § 4.3; Леман, § 3.10, 3.11; Рао, § 7с; Силвей, гл. 8; Вальд (1947)]

4.4. При сравнении двух простых гипотез H_0 и H_1 наблюдается отношение правдоподобий $\mathrm{lr}_{\text{набл}}$. Используемая выборочная схема неизвестна. Это может быть, например, выборочная схема, описанная в задаче 4.3, а может быть и выборка фиксированного объема. Применяя к данной ситуации соображения, использованные при анализе задачи 4.3, покажите, что уровень значимости при проверке H_0, т. е. $\mathrm{pr}\{\mathrm{lr}_{10}(Y) \geqslant \mathrm{lr}_{\text{набл}};\ H_0\}$, не превосходит $1/\mathrm{lr}_{\text{набл}}$.

Решение

Пусть C_m — совокупность выборочных точек $y^{(m)} = (y_1, \ldots, y_m)$, для которых процедура выбора прекращается после m наблюдений, а отношение правдоподобия не превосходит $\mathrm{lr}_{\text{набл}}$. Тогда

$$\mathrm{pr}\{\mathrm{lr}_{10}(Y) \geqslant \mathrm{lr}_{\text{набл}};\ H_1\} = \sum \int_{C_m} f_1(y^{(m)})\, dy^{(m)},$$

$$\mathrm{pr}\{\mathrm{lr}_{10}(Y) \geqslant \mathrm{lr}_{\text{набл}};\ H_0\} = \sum \int_{C_m} f_0(y^{(m)})\, dy^{(m)}.$$

Почленное сравнение этих рядов с учетом определения C_m приводит к неравенствам

$$\mathrm{lr}_{\text{набл}}\, \mathrm{pr}\{\mathrm{lr}_{10}(Y) \geqslant \mathrm{lr}_{\text{набл}};\ H_0\} \leqslant \mathrm{pr}\{\mathrm{lr}_0(Y) \geqslant \mathrm{lr}_{\text{набл}};\ H_1\} \leqslant 1,$$

из которых и вытекает требуемый результат.

Как правило, уровень значимости будет много меньше, чем $1/\mathrm{lr}_{\text{набл}}$.

[Теоретическая статистика, § 4.3; Барнард, 1947; Эфрон, 1971]

4.5. Пусть $Y_1, \ldots, Y_n$ — н. о. р. случайные величины, имеющие нормальное распределение с неизвестным μ и известным коэффициентом вариации γ_0, т. е. дисперсия равна $\gamma_0\mu^2$. Найдите статистику отношения правдоподобия для проверки гипотезы $\mu=\mu_0$ против альтернативы $\mu=\mu_A$, $\mu_A>\mu_0$ и вычислите ее распределение при нулевой гипотезе. Обсудите возможность сравнения критерия отношения правдоподобия с критериями, которые основаны на статистике: (a) $\Sigma(Y_j-\mu_0)$ или (b) $\Sigma(Y_j-\mu_0)^2$.

Решение

Критическая область отношения правдоподобия основана на больших значениях случайной величины $T=\Sigma V_j$, где $V_j=(\mu_0^{-1}+\mu_A^{-1})Y_j^2-2Y_j$. Производящую функцию моментов случайной величины T получить довольно просто непосредственно. С другой стороны, справедливо представление $Y_j=\mu_0+Z_j\gamma_0\mu_0$, где Z_j имеет при H_0 нормальное $N(0, 1)$ распределение. Статистика T является линейной функцией от

$$T'=\Sigma(Z_j+K_0)^2,$$

где $K_0=\mu_0/\{\gamma_0(\mu_A+\mu_0)\}$. Таким образом, при H_0 статистика T' имеет нецентральное хи-квадрат распределение с n степенями свободы и параметром нецентральности nK_0^2. При альтернативных гипотезах распределение случайных величин Z остается нормальным, поэтому статистика T' вновь пропорциональна нецентральной хи-квадрат случайной величине. Таким образом, и точные границы значимости, и функцию мощности можно найти из таблиц нецентрального хи-квадрат распределения.

Статистика T' является функцией выборочных среднего и дисперсии, сочетающей информацию в наиболее целесообразной форме. Статистика $\Sigma(Y_j-\mu_0)$ имеет при H_0 нормальное $N(0, n\gamma_0^2\mu_0^2)$ распределение. Для альтернативы $N(\mu, \gamma_0\mu)$ точный уровень α-мощности равен

$$1-\Phi\left\{k_\alpha^*\frac{\mu_0}{\mu}-\frac{\sqrt{n}(\mu-\mu_0)}{\gamma_0\mu}\right\}.$$

Критерий, основанный на статистике $\Sigma(Y_j-\mu_0)^2$, при H_0 использует центральное хи-квадрат распределение, а его мощность выражается через нецентральное хи-квадрат распределение.

На самом деле критерий должен был бы строиться с учетом распределений, условных относительно подчиненной статистики $\Sigma Y_j/(\Sigma Y_j^2)^{1/2}$ (см. Хинкли, 1977).

[Теоретическая статистика, § 4.4]

4.6. Пусть Y_1 — бинарная случайная величина и $E(Y_1) = \mathrm{pr}(Y=1) = \theta$. Рассмотрите гипотезы $H_0: \theta = \varepsilon$ и $H_A: \theta = 1-\varepsilon$, где ε мало. „Очевидная" критическая область состоит из единственной точки $y_1 = 1$. Каков размер и какова мощность критерия?

Далее, пусть и.о.р. случайные величины Y_1 и Y_2 имеют такое же распределение, как определенная выше случайная величина. Процедура „принятия" или „отвержения" гипотезы H_0 определяется следующим образом: для $y_1 = y_2 = 0$ принимается H_0, для $y_1 = y_2 = 1$ принимается H_A, а в противном случае принятие или отвержение H_0 осуществляется случайно с равными вероятностями. Покажите, что размер и мощность этого критерия совпадают с размером и мощностью критерия, построенного по одному наблюдению.

Этот пример используется в качестве довода в пользу того, что при сравнении H_0 и H_A „одно бинарное наблюдение столь же хорошо, как два". Опровергните это утверждение, рассматривая достижимые уровни значимости при проверке H_0.

Решение.

При одном наблюдении для размера и мощности будем иметь соответственно

$$\mathrm{pr}(\text{“принимается” } H_A;\ H_0) = \mathrm{pr}(Y=1;\ H_0) = \varepsilon,$$
$$\mathrm{pr}(\text{“принимается” } H_A;\ H_A) = \mathrm{pr}(Y=1;\ H_A) = 1-\varepsilon.$$

Когда проводятся два наблюдения и используется предложенная выше процедура, то в этом случае

$$\mathrm{pr}(\text{“принимается” } H_A;\ H_0) = \mathrm{pr}(Y_1 = Y_2 = 1;\ H_0) + \frac{1}{2}\,\mathrm{pr}(Y_1 + Y_2 = 1;\ H_0) = \varepsilon,$$

$$\mathrm{pr}(\text{“принимается” } H_A;\ H_A) = \mathrm{pr}(Y_1 + Y_2 = 1;\ H_A) + \frac{1}{2}\,\mathrm{pr}(Y_1 + Y_2 = 1;\ H_A) = 1-\varepsilon.$$

Сделанные выше качественные выводы ошибочны. При одном наблюдении имеются два возможных исхода:

$$\begin{aligned}&\text{“умеренное” свидетельство против } H_0,\\ &\text{“умеренное” свидетельство против } H_A.\end{aligned} \qquad (1)$$

В то же время при двух наблюдениях возможными исходами являются:

$$\begin{aligned}&\text{“сильные” свидетельства против } H_0,\\ &\text{“безразличные” свидельства в выборе между } H_0 \text{ и } H_A,\\ &\text{“сильные” свидетельства против } H_A.\end{aligned} \qquad (2)$$

Множества исходов (1) и (2) с точки зрения их использования для оценивания степени убедительности наличия свидетельств за и против H_0 не являются эквивалентными.

[Теоретическая статистика, § 4.5; Коэн, 1958]

4.7. [1]) Н.о.р. случайные величины $Y_1, \ldots, Y_n$ имеют показательное распределение с плотностью $\rho e^{-\rho y}$ $(y \geqslant 0)$. Получите равномерно наиболее мощный критерий для проверки гипотезы $\rho = \rho_0$ против альтернатив $\rho < \rho_0$ и найдите его функцию мощности.

Чтобы уменьшить зависимость от возможных резко выделяющихся наблюдений, предлагается заменить k наибольших значений выборки на $(k+1)$-е по величине значение. Покажите, что потеря в мощности соответствует переходу от выборки объема n к выборке объема $n-k$.

Решение

Пусть частная альтернативная гипотеза определяется как $\rho_A < \rho_0$. Отношение правдоподобия для наблюденных значений $y_1, \ldots, y_n$ имеет вид

$$\mathrm{lr}_{A0}(y) = \{\rho_A^n \exp(-\rho_A \Sigma y_j)\} / \{\rho_0^n \exp(-\rho_0 \Sigma y_j)\}.$$

Таким образом, критическая область отношения правдоподобия $\mathrm{lr}_{A0}(y) \geqslant c_\alpha$ имеет вид $\Sigma y_i > d_\alpha$ для всех $\rho_A < \rho_0$. Вычисляя производящую функцию моментов, можно убедиться, что статистика $2\rho \Sigma Y_j$ имеет хи-квадрат распределение с $2n$ степенями свободы. Отсюда вытекает, что $(2\rho_0)^{-1} c^*_{2n,\alpha}$ — точка значимости уровня α статистики Σy_i, а мощность задается как $\mathrm{pr}(X_{2n} \geqslant \geqslant \rho_A c^*_{2n,\alpha}/\rho_0)$, где X_{2n} — хи-квадрат случайная величина с $2n$ степенями свободы, а $c^*_{2n,\alpha}$ — верхняя α-точка [2]) хи-квадрат распределения с $2n$ степенями свободы.

Когда k наибольших значений наблюдений заменяются на $Y_{(n,n-k)}$, статистика критерия принимает вид

$$\sum_{j=1}^{n-k} Y_{(n,j)} + kY_{(n,n-k)}.$$

Порядковые статистики $Z_{(n,j)}$ показательного распределения с единичным средним можно представить в следующем виде.

---

[1]) Улучшенный вариант.

[2]) То есть $\mathrm{pr}(X_{2n} > c^*_{2n,\alpha}) = \alpha$. — *Прим. перев.*

(Теоретическая статистика, Приложение 2):

$$Z_{(n,j)} = \frac{W_1}{n} + \ldots + \frac{W_j}{n-j+1} \quad (j = 1, \ldots n), \tag{1}$$

где $W_1, \ldots, W_n$ — н.о.р. показательные случайные величины с единичным средним. Полагая $\rho Y_{(n,j)} = Z_{(n,j)}$ и используя представление (1), получаем

$$2\rho\left(\sum_{j=1}^{n-k} Y_{(n,j)} + kY_{(n,\,n-k)}\right) = 2\sum_{j=1}^{n-k} W_j = X_{2n-2k}.$$

Следовательно, принцип вычисления размера и мощности остается прежним, фактический объем выборки становится равным $n-k$.

[Теоретическая статистика, §4.6]

4.8. Пусть н.о.р. случайные величины $Y_1, \ldots, Y_n$ имеют плотность, пропорциональную $\exp\left(-\frac{1}{2}y^2 - \theta y^4\right)$. Получите равномерно наиболее мощный критерий для проверки гипотезы $H_0: \theta = 0$ против альтернатив $\theta > 0$. Исследуйте распределение статистики критерия при H_0. Для каких практических ситуаций может быть полезен рассматриваемый критерий?

Решение

Пусть конкретная альтернативная гипотеза определяется как $\theta_A > 0$. Отношение правдоподобия имеет вид

$$\operatorname{lr}_{A0}(y) = \exp\left(-\theta_A \Sigma y_j^4\right).$$

Соответствующая критическая область эквивалентна области $\{y: \Sigma y_j^4 \leqslant d_\alpha\}$ для всех $\theta_A > 0$.

Чтобы получить распределение статистики критерия при гипотезе H_0, можно использовать в том или ином варианте центральную предельную теорему (см. Теоретическая статистика, Приложение 1). Если Y имеет нормальное $N(0,1)$ распределение, то

$$E(Y^{4m}) = (4m)!/\{2^{2m}(2m)!\},$$

откуда $E(Y^4) = 3$, $\operatorname{var}(Y^4) = 96$. Таким образом, статистика $T = (\Sigma Y_j^4 - 3n)/\sqrt{96n}$ при гипотезе H_0 распределена приближенно как нормальная $N(0, 1)$ случайная величина. Однако коэффициент асимметрии статистики T, равный $\gamma_{1T} = 99/\sqrt{96n}$, довольно велик и показывает медленную сходимость статистики T к нормальности.

Основное применение рассматриваемого критерия относится, возможно, к ситуациям, в которых Y_j являются статистиками критерия. Однако не существует, по всей видимости, убедительных доводов, чтобы считать этот вид симметрических альтернатив более предпочтительным альтернативам, в которых величины Y_j нормальны с переменными дисперсиями. Таким образом, рассмотренная задача представляется весьма отвлеченной.

[Теоретическая статистика, §3.6]

4.9. Пусть н.о.р. случайные величины $Y_1, \ldots, Y_n$ имеют нормальное $N(0, \sigma^2)$ распределение. Рассмотрите гипотезу $H_0: \sigma^2 = 1$ против альтернативы $H_A: \sigma^2 \neq 1$. Получите наиболее мощный несмещенный критерий и сравните его численно с критерием, для которого большие и малые значения статистики критерия являются одинаково значимыми. Рассмотрите при этом случай $\alpha = 0.05$ и $n = 5$.

Решение

Для произвольной альтернативной гипотезы плотность вектора Y равна

$$(2\pi\sigma^2)^{-\frac{1}{2}n}\exp\left(-\frac{1}{2}\Sigma y_j^2/\sigma^2\right) = q(y, \sigma).$$

Применим поэтому общий результат задачи 4.2, положив $m = 2$, $k_1(y) = q(y, 1)$, $k_2(y) = [\partial q(y, \sigma)/\partial\sigma]_{\sigma=1}$, $k_3(y) = q(y, \sigma)$, $a_1 = \alpha$, $a_2 = 0$. Критическая область определяется в виде квадратического неравенства от $t = \Sigma y_j^2$ и будет иметь вид $\{t \leqslant t_1^*, t \geqslant t_2^*\}$. Поскольку T/σ^2 имеет хи-квадрат распределение с n степенями свободы, п.р.в. которого обозначим $p_n(t)$, то t_1^* и t_2^* можно определить из системы

$$\int_{t_1^*}^{t_2^*} p_n(t)\,dt = 1 - \alpha, \qquad \int_{t_1^*}^{t_2^*} (t - n)\,p_n(t)\,dt = 0.$$

Интегрирование по частям второго уравнения приводит его к явному виду

$$(t_1^*)^{\frac{1}{2}n} e^{-\frac{1}{2}t_1^*} = (t_2^*)^{\frac{1}{2}n} e^{-\frac{1}{2}t_2^*}.$$

Дальнейшее продвижение возможно лишь численными методами. Линдли и др. (1960) табулировали t_1^* и t_2^*. Для $n = 5$, $\alpha = 0.05$ получим $t_1^* = 0.989$, $t_2^* = 14.37$. Границы критической области критерия, для которого одинаково значимыми являются и большие и малые значения статистики критерия, задаются величинами 0.831 и 12.83. Вероятности, отвечающие хвостам распределения для точек t_1^* и t_2^*, приблизительно равны 0.037 и 0.013 соответственно.

[Теоретическая статистика, § 4.7;
Леман, § 5.2]

4.10. Разлагая в ряд отношение правдоподобий, используйте члены высших порядков для нахождения усовершенствованного локально наиболее мощного критерия для проверки гипотезы $H_0: \theta = 0$ против альтернативы $H_A: \theta > 0$, когда н.о.р. случайные величины $Y_1, \ldots, Y_n$ нормально $N(0, 1 + a\theta^2)$ распределены. Здесь a — известная положительная постоянная.

Решение

Для одной случайной величины Y_j логарифм плотности определяется как

$$l_j(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(1 + a\theta)^2 - \frac{(Y_j - \theta)^2}{2(1 + a\theta^2)},$$

а его первая производная — как

$$U_j(\theta) = -\frac{a\theta}{1 + a\theta^2} + \frac{Y_j - \theta}{1 + a\theta^2} + \frac{a\theta(Y_j - \theta)^2}{(1 + a\theta^2)^2}.$$

Таким образом, $U_j(0) = Y_j$. Вторая производная от $l_j(\theta)$ в точке $\theta = 0$ равна

$$U_j'(0) = -a - 1 + aY_j^2,$$

откуда получаем $i_j(0) = 1$.

Аппроксимация логарифма отношения правдоподобий для $\theta = \delta$ против $\theta = 0$, использующая два члена разложения Тейлора

$$\delta \Sigma U_j(0) + \frac{1}{2}\delta^2 \Sigma U_j'(0),$$

приводит к статистике критерия

$$\Sigma Y_j + \frac{1}{2}\delta a \Sigma (Y_j^2 - 1). \qquad (1)$$

Выбор δ довольно произволен, но разумная процедура состоит в том, чтобы максимизировать мощность критерия уровня α вблизи точки, где мощность равна β. Поскольку статистика критерия (1) приближенно нормально $N(\delta n, n)$ распределена, то указанное выше требование приводит к выбору

$$\delta \simeq (k_\alpha^* + k_\beta^*)/\sqrt{n}.$$

По-видимому, выбор $\delta = 2/\sqrt{n}$ столь же хорош, как и любой другой. В этом случае статистика критерия принимает вид

$$T = \Sigma\left(Y_j + \frac{1}{2n}\sqrt{n}\right)^2.$$

Точное распределение статистики T совпадает с распределением случайной величины $(1 + a\theta^2)\, X_n(\phi_n)$, где $X_n(\phi_n)$ распределена как нецентральная хи-квадрат случайная величина с n степенями свободы и параметром нецентральности

$$\phi_n = n\left(2a\theta + \sqrt{n}\right)^2/\{4a^2(1 + a\theta^2)\}.$$

Статистика вклада ΣY_j имеет в точности нормальное $N(n\theta, n + na\theta^2)$ распределение. Приведем пример сравнения мощностей критериев уровня 0.05 для $n = 10$, $a = 1$:

	θ	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$1\frac{1}{4}$	$1\frac{1}{2}$.
Мощность статистики критерия	T	0.11	0.22	0.59	0.87	0.97	0.99	1.00
Мощность статистики критерия	ΣY_j	0.11	0.20	0.48	0.72	0.86	0.93	0.96

Вычисления четко показывают, что мощность критерия, использующего статистику T, больше мощности второго критерия в области средних и высоких значений функции мощности. В то же время вычисления недостаточно точны, чтобы выявить более высокую мощность статистики ΣY_j вблизи нуля.

[Теоретическая статистика, § 4.8; Эфрон, 1975; Рао, § 7а.4]

4.11. Рассмотрим полиномиальное распределение с четырьмя исходами в единичном испытании и вероятностями исходов $\left\{\frac{1}{6}(1-\theta), \frac{1}{6}(1+\theta), \frac{1}{6}(2-\theta), \frac{1}{6}(2+\theta)\right\}$ соответственно. Покажите, что информация для n независимых испытаний имеет вид

$$i_n(\theta) = \frac{(2-\theta^2)\, n}{(1-\theta^2)(4-\theta^2)}.$$

Покажите, что существуют две подчиненные статистики. Используйте дисперсию условной информации для выбора одной из них.

Решение

Пусть $N_j\ (j = 1, \ldots, 4)$ обозначают выборочные частости. Они имеют полиномиальное распределение с заданными вероятностями исходов $p_j(\theta)$. Таким образом, с точностью до постоянной

$$\log f_Y(y;\theta) = n_1 \log(1-\theta) + n_2 \log(1+\theta) + n_3 \log(2-\theta) + n_4 \log(2+\theta),$$

$$U_.(\theta) = -\frac{N_1}{1-\theta} + \frac{N_2}{1+\theta} - \frac{N_3}{2-\theta} + \frac{N_4}{2+\theta},$$

а

$$-U'_.(\theta) = \frac{N_1}{(1-\theta)^2} + \frac{N_2}{(1+\theta)^2} + \frac{N_3}{(2-\theta)^2} + \frac{N_4}{(2+\theta)^2}.$$

Ожидание последней функции представляет собой общую информацию. При использовании равенства $EN_j = np_j(\theta)$ информация вычисляется и оказывается равной

$$i_.(\theta) = n(2-\theta^2)/\{(1-\theta^2)(4-\theta^2)\}.$$

В рассматриваемой модели существуют две возможные подчиненные статистики

$$A = (A_1, A_2),\ \text{где}\ A_1 = N_1 + N_2,\ A_2 = n - A_1 = N_3 + N_4;$$
$$B = (B_1, B_2),\ \text{где}\ B_1 = N_1 + N_4,\ B_2 = n - B_1 = N_2 + N_3.$$

В частности, случайная величина A_1 имеет биномиальное распределение, соответствующее n испытаниям и вероятности успеха, равной $1/3$, т. е. не зависящей от θ. Условная информация при заданном значении статистики A получается как условное ожидание от $-U'_.(\theta)$, т. е.

$$i_.(\theta \mid A = a) = \{3a_1 + n(1-\theta^2)\}/\{(1-\theta^2)(4-\theta^2)\}.$$

Условное распределение статистики N_1 совпадает с биномиальным, соответствующим a_1 испытанию и вероятности успеха, равной $\frac{1}{2}(1-\theta)$. Ожидание величины $i_.(\theta \mid A)$ по мере, порождающей случайную величину A, совпадает, конечно, с $i_.(\theta)$, а ее дисперсия равна

$$2n/\{(1-\theta^2)(4-\theta^2)\}^2.$$

Для подчиненной статистики B ожидание условной информации вновь равно $i_.(\theta)$, а дисперсия имеет вид

$$\theta^2 n/\{(1-\theta^2)(4-\theta^2)\}^2.$$

Итак, поскольку $\theta^2 < 2$, то можно сделать вывод, что подчиненная статистика A более избирательна в сравнении со статистикой B в процессе классификации данных на множества различной информативности. С этой точки зрения статистика A предпочтительнее.

[Теоретическая статистика, § 4.8. 2.2 (viii); Басу, 1964; Кокс, 1971]

4.12. Марковская цепь с конечным числом состояний имеет переходные вероятности $(p_{jk}(\theta))$, зависящие от скалярного параметра θ. Цепь является эргодической, начальное распределение цепи совпадает со стационарным. Исследуя логарифм правдоподобия, найдите общую функцию информации $i_.(\theta)$ для наблюдений в моменты $0, 1, \ldots, n$.

Решение

Обозначим равновесное распределение цепи $\{\pi_j(\theta)\}$. Пусть j_0 — начальное состояние цепи, а m_{jk} — число переходов из состояния j в состояние k в течение n переходов цепи. Тогда полный логарифм правдоподобия запишется как

$$\log \pi_0(\theta) + \Sigma\Sigma m_{jk} \log p_{jk}(\theta).$$

Далее, пусть M_{jk} — случайная величина, соответствующая m_{jk}. Поскольку начальное и все последующие состояния имеют равновесные маргинальные распределения, то

$$E(M_{jk}) = n\pi_j(\theta)\, p_{jk}(\theta).$$

Таким образом, если n велико, то средняя информация, приходящаяся на одно наблюдение, получится при делении второй производной от логарифма правдоподобия (1) на n. При этом необходимо пренебречь первым слагаемым и заменить $m_{jk}(\theta)$ на $n\pi_j(\theta)\, p_{jk}(\theta)$. В результате получим

$$\Sigma\Sigma\pi_j(\theta)\, p_{jk}(\theta) \left\{-\frac{\partial^2 \log p_{jk}(\theta)}{\partial\theta^2}\right\} = \Sigma\pi_j(\theta)\, i(\theta \mid j),$$

где $i(\theta \mid j)$ — информация, вычисленная на основе вероятностного распределения j-й строки переходной матрицы.

[Теоретическая статистика, §4.8]

4.13. Пусть независимые случайные величины $Y_1, \ldots, Y_n$ имеют пуассоновские распределения со средними $\mu, \mu\rho, \ldots, \mu\rho^{n-1}$. Покажите, как можно получить локально оптимальные критерии: (a) для проверки гипотезы $\mu = \mu_0$, когда ρ известно, и (b) для проверки гипотезы $\rho = \rho_0$, когда μ известно.

Решение

Логарифм вероятностей значений случайной величины Y_j равен

$$-\mu\rho^{j-1} + y_j \log(\mu\rho^{j-1}) - \log y_j!$$

При рассмотрении п. (a) задачи произведем частное дифференцирование этого выражения по μ. В результате получим вклад одного наблюдения в эффективный вклад выборки, а именно $U_j(\mu_0) = -\rho^{j-1} + Y_j/\mu_0$, дисперсия которого $i_j(\mu_0) = \rho^{j-1}/\mu_0$ является долей информации выборки, приходящейся на j-е наблюдение. Таким образом, статистика критерия имеет вид

$$U_\cdot(\mu_0) = \frac{1}{\mu_0} \sum_{j=1}^{n} (Y_j - \mu_0 \rho^{j-1}).$$

Как можно проверить непосредственными вычислениями, при нулевой гипотезе эта статистика имеет нулевое среднее и дисперсию $\Sigma\rho^{j-1}/\mu_0$. Статистика асимптотически нормальна при $\rho \geqslant 1$. Нормальная аппроксимация возможна также и в том случае, если ρ не слишком мало по сравнению с единицей, а μ_0 достаточно велико. Отметим, что s, кумулянт статистики $U_\cdot(\mu_0)$, в силу свойства аддитивности кумулянтов равен

$$\Sigma\rho^{j-1}/\mu_0^{s-1}.$$

При рассмотрении пункта (b) задачи, когда роли μ и ρ меняются для эффективного вклада по ρ, удобно ввести другое обозначение, например $U_j^*(\rho_0)$. За исключением частных различий, общие идеи построения статистически критерия те же, что и в пункте (а) задачи. В итоге получим

$$U_j^*(\rho_0) = -\mu(j-1)\rho_0^{j-2} + (j-1)Y_j/\rho_0,$$
$$i_j^*(\rho_0) = \mu\rho_0^{j-3}(j-1)^2.$$

Случай известного ρ соответствует наблюдению прореженных последовательностей. Приведенное выше решение позволяет, в принципе, получать для неизвестного μ доверительные интервалы (см. гл. 7).

[Теоретическая статистика, § 4.8;
Рао; § 7a]

4.14. Независимые случайные величины $Y_1, \ldots, Y_n$ нормально распределены с постоянной дисперсией σ^2 и со средними $E(Y_j) = e^{\beta x_j}$, где $x_1, \ldots, x_n$ — известные постоянные. Получите информационную матрицу для параметра $\theta = (\beta, \sigma^2)$. Найдите локально наиболее мощный критерий для проверки гипотезы $\beta = \beta_0$ против альтернативы $\beta \neq \beta_0$, когда σ^2 известно и равно σ_0^2.

Постоянные $x_1, \ldots, x_n$ можно выбирать из $[0, 1]$. На самом деле предполагается выбрать эти постоянные так, чтобы максимизировать $i(\beta_0)$. Попытайтесь обосновать необходимость этого. Предполагая, что все x_j выбраны равными, найдите это общее значение, максимизирующее $i_\cdot(\beta_0)$.

Решение

Логарифм плотности случайной величины Y_j равен

$$-\frac{1}{2}\log(2\pi)-\log\sigma-\frac{1}{2}(Y_j-e^{\beta x_j})^2/\sigma^2.$$

Отсюда компоненты эффективного вклада в очевидных обозначениях запишутся следующим образом:

$$U_{j\beta}=x_j e^{\beta x_j}(Y_j-e^{\beta x_j})/\sigma^2,\quad U_{j\sigma}=\{(Y_j-e^{\beta x_j})^2-\sigma^2\}/\sigma^3.$$

Взнос j-го наблюдения в информационную матрицу выборки равен ковариационной матрице $U_{j\beta}$ и $U_{j\sigma}$, т. е.

$$i_j=\operatorname{diag}(x_j^2 e^{2\beta x_j}/\sigma^2,\ 2/\sigma^2).$$

Таким образом, статистика эффективного вклада при проверке нулевой гипотезы $\beta=\beta_0$ эквивалентна статистике

$$T=\Sigma x_j e^{\beta_0 x_j}(Y_j-e^{\beta_0 x_j}).$$

При нулевой гипотезе статистика T нормально $N\left(0,\ \sigma^2\Sigma x_j^2 e^{\beta_0 x_{2j}}\right)$ распределена. Отметим, что свойство оптимальности статистики критерия выражено довольно слабо.

Одно из обоснований необходимости максимизации информации $i_.(\beta)$ состоит в том, что при этом максимизируется локальная мощность. Когда $\beta=\beta_0+\delta$, нормированная статистика критерия распределена нормально со средним, равным приближенно $\delta\sqrt{i_.(\beta_0)}$, и дисперсией 1. Аналогично, асимптотические свойства оценки максимального правдоподобия параметра β оптимизируются вблизи $\beta=\beta_0$ (см. гл. 9). Чтобы максимизировать информацию $i_.(\beta_0)$ для плана, в котором все наблюдения проводятся в точке $\tilde{x}$, необходимо максимизировать $\tilde{x}^2\exp\{2\beta_0\tilde{x}\}$ при $0\leqslant\tilde{x}\leqslant 1$. Легко показать, что

$$\tilde{x}=\begin{cases}1 & \text{при } \beta_0\geqslant -1,\\ -1/\beta_0 & \text{при } \beta_0<-1.\end{cases}$$

Качественная форма этого результата представляется весьма правдоподобной и на основе общих соображений. Доказать непосредственным расчетом оптимальность плана, сосредоточенного в одной точке, довольно трудно. Общий изящный подход к решению этой задачи связан с обобщением теорем эквивалентности линейных планов для нелинейных моделей Кифера — Вольфовица (см. Уайт, 1973).

[Теоретическая статистика, § 4.8]

4.15. Обобщите разложение (с) раздела „Необходимые сведения" на случай векторных параметров. Покажите прн этом, что $E\{U_.(\theta_0);\ \theta_0+\delta\} = \mathbf{i}_.(\theta)\,\delta + o(\delta)$, где δ — вектор-столбец. Проверьте это соотношение для линейной нормальной модели с известной дисперсией. Пусть, далее, векторный параметр θ преобразуется в новый параметр ϕ с помощью взаимнооднозначного дифференцируемого отображения. Покажите, что новой информационной матрицей является

$$\left(\frac{\partial\theta}{\partial\phi}\right)^{\mathrm{T}} \mathbf{i}_.(\theta) \left(\frac{\partial\theta}{\partial\phi}\right).$$

Решение

Первая часть задачи следует непосредственно из разложения разности $U_.(\theta) - U_.(\theta+\delta)$ и предположений непрерывности $\mathbf{i}_.(\theta+\delta) = \mathbf{i}_.(\theta) + o(1)$.

Логарифм правдоподобия для линейной нормальной модели с известной дисперсией σ_0^2 с точностью до постоянных равен

$$\frac{(\hat{\beta}-\beta)^{\mathrm{T}}\,(\mathbf{x}^{\mathrm{T}}\mathbf{x})\,(\hat{\beta}-\beta)}{2\sigma_0^2},$$

где $\hat{\beta} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\,\mathbf{y}^{\mathrm{T}}Y$. Для проверяемого значения β_0 частное дифференцирование приводит к следующему выражению:

$$U_.(\beta_0) = (\mathbf{x}^{\mathrm{T}}\mathbf{x})\,(\hat{\beta}-\beta_0)/\sigma_0^2.$$

Поскольку $E(\hat{\beta}) = \beta$ для всех β, то

$$E\{U_.(\beta_0);\ \beta_0+\delta\} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})\,\delta/\sigma_0^2 = \mathbf{i}_.\delta.$$

Это равенство является точным для всех δ. Отметим, что эффективный вклад есть нормированное уклонение оценки наименьших квадратов от значения соответствующего параметра. Обсуждение основных положений асимптотической теории правдоподобия в гл. 9 строится на предположении, что некоторое соотношение выполнено приближенно для всех регулярных оценок максимума правдоподобия.

Соответствующее выражение для информационной матрицы при преобразовании параметров вида $\phi = \phi(\theta)$ можно получить непосредственно из соотношения

$$i_{.rs}(\theta) = E\left\{\frac{\partial \log f_Y(Y;\theta)}{\partial\theta_r}\,\frac{\partial \log f_Y(Y;\theta)}{\partial\theta_s};\ \theta\right\},$$

используя при этом формулу

$$\frac{\partial}{\partial\phi_k} = \sum_l \left(\frac{\partial\theta_l}{\partial\phi_k}\,\frac{\partial}{\partial\theta_l}\right).$$

[Теоретическая статистика. § 4.8(ii); Рао § 5а.3, 4а.4]

4.16. Покажите, что ожидание логарифма отношения правдоподобий

$$\int \log\left\{\frac{f_1(y)}{f_0(y)}\right\} f_0(y)\,dy$$

является отрицательным, за исключением случая, когда рассматриваемые распределения являются идентичными. В последнем случае оно равно нулю. Выведите отсюда, что симметричное положительное расстояние между двумя распределениями можно определить как

$$\int \log\left\{\frac{f_1(y)}{f_0(y)}\right\} \{f_1(y)-f_0(y)\}\,dy.$$

Докажите, что если распределения принадлежат одному и тому же регулярному семейству, т. е. $f_0(y)=f(y;\,\theta)$, а $f_1(y)=f(y;\,\theta+\delta)$, то при $\delta\to 0$ расстояние измеряется величиной $i(\theta)\,\delta^2+o(\delta^2)$.

Решение

Первая часть задачи вытекает из неравенства Йенсена.

На определенное выше расстояние не влияет перестановка местами f_0 и f_1. Таким образом, это симметричная мера расстояния. Поскольку знаки сомножителей под интегралом всегда совпадают, то она строго положительна, за исключением случая индентичных f_0 и f_1.

Далее, при $f_1(y)=f(y;\,\theta+\delta)$ и $f_0(y)=f(y;\,\theta)$ справедливо разложение

$$f_1(y)-f_0(y)=\delta\,(\partial f/\partial\theta)+o(\delta),$$
$$\log\{f_1(y)/f_0(y)\}=\delta\,\frac{1}{f}\left(\frac{\partial f}{\partial\theta}\right)+o(\delta).$$

Отсюда мера расстояния определяется как

$$\delta^2\int\frac{1}{f}\left(\frac{\partial f}{\partial\theta}\right)^2 dy+o(\delta^2),$$

что и требовалось доказать.

Термин „расстояние“ не совсем подходит в данной ситуации, поскольку неравенство треугольника может и не выполняться.

[Теоретическая статистика, § 4.8 (ii); Кендалл, 1973; Кульбак, 1968; Рао, § 5а.4]

5. Критерии значимости: сложные нулевые гипотезы

Необходимые сведения

В гл. 4 предполагалось, что нулевая гипотеза H_0 простая, т. е. она полностью определяла распределение случайной величины Y. В приложениях, однако, нулевая гипотеза, как правило, сложная. Распределение случайной величины Y определяется ею не полностью, а ограничивается лишь некоторым классом распределений, каждый элемент которого зависит от одного или нескольких неизвестных параметров. Обсудим теперь основные методы построения критериев проверки сложных нулевых гипотез, предполагая вновь, что имеющиеся альтернативы представляют те отклонения от H_0, которые представляют особый интерес.

Предположим, что возможные распределения для случайной величины Y параметризованы, параметр $\theta=(\psi, \lambda)$, $\theta \in \Omega=\Omega_\psi \times \Omega_\lambda$. Пусть нулевая гипотеза принимает вид H_0: $\psi=\psi_0$, а λ — неизвестный мешающий параметр. Критическую область w_α размера α называют подобной, если

$$\operatorname{pr}(Y \in w_\alpha;\ \psi_0;\ \lambda)=\alpha$$

для всех λ. Критерий, определенный системой таких критических областей, предполагаемых вложенными друг в друга, называют подобным критерием.

Если при нулевой гипотезе H_0 существует ограниченно полная минимальная достаточная статистика $S_\lambda(\psi_0)$, тогда любая критическая область размера α должна иметь размер α, рассчитанный для условного относительно $S_\lambda(\psi_0)=s(\psi_0)$ распределения почти для всех $s(\psi_0)$. О таких критических областях говорят, что они имеют неймановскую структуру. Оптимальным критерием против полностью определенной альтернативы является критерий отношения правдоподобия, который будет зависеть, вообще говоря, от конкретного выбранного альтернативного значения параметров (ψ_1, λ_1). Если критические области отношения правдоподобия не зависят ни от ψ_1, ни от λ_1, то получающийся в результате критерий называют равномерно наиболее мощным подобным критерием. Если критические области отношения правдоподобия не зависят от λ_1, но зависят от ψ_1, то существуют другие подходы, в частности, и обсуждавшиеся уже в гл. 4. Так, например, если ψ — одномерный параметр, то можно попытаться максимизировать мощность для альтернатив, близких к $\psi = \psi_0$. Если критическая область о. п. зависит от λ_1, то в этом случае общей теории построения хороших подобных критериев не существует. Один из методов построения критерия для таких моделей связан с вычислением отношения правдоподобия

$$f_{Y \mid S_\lambda(\psi_1)}(y \mid s(\psi_1);\ \psi_1)/f_{Y \mid S_\lambda(\psi_0)}(y \mid s(\psi_0);\ \psi_0).$$

Для одного общего класса задач удается с успехом применить теорию подобных критериев. Речь идет о задачах, в которых нулевая гипотеза задается в виде линейных ограничений на естественные параметры экспоненциального семейства. Приведем два типа задач, где подход с использованием подобных областей не проходит:

(i) там, где S_λ не является ограниченно полной;

(ii) там, где S_λ является минимально достаточной для всех θ. Последнее затруднение возникает тогда, когда гипотеза H_0 сформулирована так, что соответствующее ей множество значений параметра не образует подпространства в Ω. Для задач типа (i) подобные критерии могут существовать, но их придется искать другими методами.

Второй общий подход к построению критериев проверки сложных гипотез связан с соображениями инвариантности, олицетворяющему требования, что в некоторых соответствующим образом сформулированных задачах выводы не должны зависеть от принятой шкалы измерений. Предположим вновь, что распределения случайных величин Y параметризованы, θ — соответствующий параметр. Нулевая гипотеза H_0 задается как $\theta \in \Omega_0$, а альтернативная гипотеза H_A задается как $\theta \in \Omega_A$. Задача проверки гипотез инвариантна относительно группы преобразований $\mathcal{G}$, действующей в выборочном пространстве, если для любого преобразования $g \in \mathcal{G}$ и любого множества B из выборочного пространства

$$\mathrm{pr}\,(gY \in B;\ \theta) = \mathrm{pr}\,(Y \in B;\ g^*\theta),$$

где g^* принадлежит группе преобразований, действующей в Ω, причем группа такова, что $g^*\Omega_0 = \Omega_0$ и $g^*\Omega_A = \Omega_A$. Отсюда следует, в частности, что H_0 верна для gY при всех g, если она верна для Y. Критерий с критической областью w_α размера α является инвариантным, если из $Y \in w_\alpha$ вытекает, что $gY \in w_\alpha$ для любого $g \in \mathcal{G}$. Статистика $t\,(Y)$, определяющая инвариантную критическую область, является максимальным инвариантом, если из $t(y) = t\,(y')$ следует, что $y' = gy$ для некоторого $g \in \mathcal{G}$. Таким образом, чтобы построить соответствующий инвариантный критерий, необходимо в качестве статистики критерия выбрать функцию наблюдений y, которая является постоянной на орбите точек gy, когда g пробегает $\mathcal{G}$, и которая изменяет свое значение при переходе от орбиты к орбите. В общем случае для упрощения задачи перед применением соображений, основанных на инвариантности, проводится достаточное сокращение наблюдений y. Сокращение наблюдений до максимально инвариантной статистики порождает аналогичное сокращение параметра θ. Высказанные ранее доводы, связанные с оптимальностью критериев, применяются к сокращенной задаче проверки гипотез, основанной на использовании максимально инвариантной статистики $t(\cdot)$.

В тех случаях, когда метод перехода к условным распределениям или к доводам, связанным с инвариантностью, не дает возможность однозначно выделить критерий, применяют общий метод построения критериев, связанный с образованием статистики максимума отношения правдоподобия:

$$\sup_{\Omega_A} \operatorname{lik}(\theta;\ y)/\sup_{\Omega_0} \operatorname{lik}(\theta;\ y).$$

Показывается (см. гл. 9), что эта статистика в случае выборок большого объема обладает свойством оптимальности. Другой подход, особенно полезный для некоторых многомерных задач, состоит в том, чтобы работать со стохастически наибольшей (наиболее значимой) статистикой критерия, полученной на основе линейных преобразований многомерной величины.

Задачи

5.1. Н. о. р. случайные величины $Y_1, \ldots, Y_{n_1}$ имеют геометрическое распределение с параметром θ_1 и независимы от н. о. р. случайных величин $Y_{n_1+1}, \ldots, Y_{n_1+n_2}$, которые имеют геометрическое распределение с параметром θ_2. Постройте равномерно наиболее мощный подобный критерий для проверки нулевой гипотезы $\theta_1=\theta_2$ против альтернативы $\theta_1>\theta_2$. Какую более общую нулевую гипотезу можно проверить аналогичным образом?

Решение

Имеем $f(y;\ \theta)=\theta^y(1-\theta)$, где $y=0, 1, \ldots, \theta=\theta_1$ для $Y_1, \ldots, Y_n$ и $\theta=\theta_2$ для $Y_{n_1+1}, \ldots, Y_{n_1+n_2}$. Достаточной статистикой для параметра (θ_1, θ_2) является пара (S_1, S_2), где $S_1=Y_1+\ldots+Y_{n_1}$ и $S_2+Y_{n_1+1}+\ldots+Y_{n_1+n_2}$. При $\theta_1=\theta_2$ достаточная статистика примет вид $S=S_1+S_2$.

Пусть альтернатива H_A такова, что $\theta_1>\theta_2$. Тогда $\log\theta_1>\log\theta_2$, т. е. имеем неравенство в естественном параметрическом пространстве. Можно показать, что существует равномерно наиболее мощный подобный критерий. Он получается, если будем строить критерий для любой конкретной альтернативы. Подобные критерии имеют структуру Неймана, поскольку статистика S полна, что вытекает из свойств рассматриваемого семейства экспоненциального типа. Отношение правдоподобий для конкретных (θ_1, θ_2) против (θ_1, θ_1) имеет вид

$$\frac{f_{S_1\mid S}(s_1\mid s;\ \theta_1,\ \theta_2)}{f_{S_1\mid S}(s_1\mid s;\ \theta_1,\ \theta_1)}=\frac{\left(\frac{\theta_1}{\theta_2}\right)^{s_1}\sum\binom{r-1}{n_1-1}\binom{s-r-1}{n_2-s}}{\sum\binom{r-1}{n_1-1}\binom{s-r-1}{n_2-1}\left(\frac{\theta_1}{\theta_2}\right)^{r}}$$

и является при $\theta_1>\theta_2$ возрастающей функцией от s_1. Отсюда вытекает, что критическая область размера α имеет вид $w_\alpha=\{(s_1,\ s_2)\colon\ s_1>c_\alpha(s_2)\}$. Односторонний уровень значимости для $S_{1,\,\text{набл}}$ получим в виде суммы

$$\sum_{r=s_{1,\,\text{набл}}}^{s}\binom{s-1}{n_1-1}\binom{s-r-1}{n_2-1}\Big/\binom{s-1}{n_1+n_2-1}$$

Для рассматриваемой модели естественными параметрами являются $\log\theta_1$ и $\log\theta_2$. Отсюда следует, что при проверке гипотезы H_0: $\theta_1=\gamma_0\theta_2$ против альтернативы H_A: $\theta_1>\gamma_0\theta_2$ подобный критерий будет иметь ту же форму, что и рассмотренный выше. Можно также проверить гипотезу об отношении естественных параметров, т. е. проверить нулевую гипотезу вида H_0: $\theta_1=\theta_2^{\gamma_0}$. Правда, маловероятно, что проверка такой гипотезы может иметь реальный интерес. Соответствующие критические области имеют некоторые нехарактерные особенности.

Относительно критериев проверки геометричности распределения см. задачу 3.11.

[Теоретическая статистика, § 5.2;
Леман, § 4.3, 4.4]

5.2. Пусть события осуществляются в соответствии с неоднородным пуассоновским процессом, имеющим интенсивность $\rho e^{\beta t}$, где ρ и β неизвестны. В интервале $(0,\ t_0)$ наблюдаются моменты осуществления событий. Постройте равномерно наиболее мощный подобный критерий для проверки $\beta=\beta_0$ против альтернативы $\beta>\beta_0$. Исследуйте более подробно вид критерия для частного случая $\beta_0=0$.

Решение

Правдоподобие для общего неоднородного пуассоновского процесса с интенсивностью $\rho(t)$, события которого произошли в моменты $y_1,\ \ldots,\ y_n$, получают следующим образом. Каждому небольшому временному интервалу $(t,\ t+\Delta)$ соответствует в правдоподобии сомножитель $1-\rho(t)\,\Delta t+o(\Delta t)$, если в интервале событие не произошло, и сомножитель $\rho(t)\,\Delta t$, если интервал содержит момент осуществления события. Таким образом, перемножая все подходящие сомножители и устремляя $\Delta t \rightarrow 0$, получим, что правдоподобие имеет вид

$$\left\{\prod_{j=1}^{n}\rho(y_j)\right\}\exp\left\{-\int_0^{t_0}\rho(t)\,dt\right\}.$$

В рассматриваемом частном случае получаем

$$\operatorname{lik}(\rho,\ \beta;\ y_1,\ \ldots,\ x_n)=\exp\{n\log\rho+\beta\Sigma y_j-\rho\beta^{-1}(e^{\beta t_0}-1)\}.$$

Отсюда следует, что $(N,\ \Sigma Y_j)$ — достаточная статистика. При нулевой гипотезе H_0: $\beta=\beta_0$ достаточной статистикой будет уже N. Подобные критерии получаются из условного распределения случайной величины ΣY_j при заданном $N=n$. Маргинальным распределением числа событий N будет пуассоновское распределение со средним $\rho(e^{\beta t_0}-1)/\beta$. Следовательно, $f_{Y_1,\ldots,Y_N|N}(y_1,\ \ldots,\ y_n|n;\ \rho,\ \beta)\propto e^{\beta\Sigma y_j}$. Это условное распределение в качестве естественного параметра имеет β, а соответствующее отношение правдоподобий в H_A монотонно возрастает. Отсюда вытекает, что большие значения статистики $U=\Sigma Y_j$ являются значимыми. Распределение статистики U совпадает с распределением суммы n н. о. р. случайных величин с усеченной до $(0,\ t_0)$ показательной плотностью.

В частном случае $\beta_0=0$ совместное распределение величин $Y_1,\ \ldots,\ Y_n$ совпадает с распределением вариационного ряда равномерно распределенных на $(0,\ t_0)$ n н. о. р. случайных величин. Аппроксимирующее нормальное распределение для статистики U имеет при нулевой гипотезе среднее $\frac{1}{2}nt_0$ и дисперсию $\frac{1}{12}nt_0^2$.

[Теоретическая статистика, § 5.2;
Кокс и Льюис, 1966, гл. 3;
Леман, § 4.3, 4.4]

5.3. Имеется m независимых оценок дисперсии нормального распределения. Каждая оценка имеет d степеней свободы. Пусть $\sigma_1^2,\ \ldots,\ \sigma_m^2$ — соответствующие параметры, $\sigma_j^2=1/(\lambda+\psi j)$, $(j=1,\ \ldots,\ m)$. Получите равномерно наиболее мощный подобный критерий для проверки гипотезы $\psi=0$ против альтернатив $\psi>0$. Предложите более простую, хотя, быть может, и менее эффективную процедуру, основанную на логарифмическом преобразовании дисперсий.

Решение

Пусть оценки дисперсий равны $v_1, \ldots, v_m$. Положим $Y_j = V_j d$, откуда $Y_j = \sigma_j^2 X_{d,j}^2$ $(j = 1, \ldots, m)$. Из соотношения

$$\operatorname{lik}(\psi, \lambda; y_1, \ldots, y_m) \propto \Pi (\lambda + \psi j)^{-d} \exp\left\{-\frac{1}{2}\Sigma(\lambda + \psi j) y_j\right\}$$

следует минимальная достаточность статистики $(\Sigma Y_j, \Sigma_j Y_j)$. При нулевой гипотезе H_0: $\psi = 0$ минимальная достаточная статистика сводится к ΣY_j. Отсюда получаем, что подобные критерии строятся на основе условного распределения статистики $\Sigma_j Y_j$ при заданном значении $\Sigma Y_j = \Sigma y_j$. Указанное распределение имеет монотонное отношение правдоподобий. Для альтернатив $\psi > 0$ значимыми оказываются большие значения статистики $\Sigma_j Y_j$, уровень значимости рассчитывается по соответствующему условному распределению.

Если m достаточно велико, то при нулевой гипотезе можно использовать нормальную аппроксимацию. Используя свойство полноты, легко получить следующие соотношения:

$$E(\Sigma_j Y_j \mid \Sigma Y_j = t) = \frac{1}{2}(m+1)t,$$

$$\operatorname{var}(\Sigma_j Y_j \mid \Sigma Y_j = t) = \frac{1}{6} m(m^2 - 1) t^2 d/(2md + m^2 d^2) \sim \frac{1}{6} m t^2/d.$$

Более простой критерий связан с преобразованием $X_j = \log Y_j$. В этом случае дисперсия $\operatorname{var}(X_j)$ не зависит от σ_j^2 и равна приближенно $2/d$, в то время как $E(X_j) \simeq \log \sigma_j^2 - 1/d$. Для больших d и малых ψ верно приближение

$$X_j = \lambda' + \psi' j + \varepsilon_j, \quad \psi' \propto \psi,$$

где ε_j — н. о. р. случайные величины, имеющие нормальное $N(0, 2/d)$ распределение. Приближенный критерий проверки гипотезы $\psi = 0$ сводится к проверке гипотезы $\psi' = 0$ в нормальной линейной модели. Обсуждение эффективности предложенного подхода проводится в „Теоретической статистике", пример 8.5.

[Теоретическая статистика,
§ 5.2; Леман, § 4.3, 4.4]

5.4. Пусть в моменты $0, \ldots, n$ наблюдается реализация марковской цепи с двумя состояниями, начальное состояние i_0 считается фиксированным. Одношаговые переходные вероятности из состояния u в состояние v обозначим θ_{uv} $(u, v = 0, 1)$, m_{uv} — общее число переходов из состояния u в состояние v за время наблюдений. Докажите, что если $\log(\theta_{01}/\theta_{00}) = \lambda$, а

$\log(\theta_{11}/\theta_{10}) = \lambda + \psi$, то правдоподобие имеет вид

$$\frac{\exp(\lambda m_{\cdot 1} + \psi m_{11})}{(1+e^{\lambda})^{m_{0\cdot}}\,(1+e^{\lambda+\psi})^{m_{1\cdot}}}.$$

По форме оно несколько отличается от правдоподобий для семейств экспоненциального типа, приводящих к равномерно наиболее мощному подобному критерию для нулевой гипотезы $\psi = \psi_0$. Покажите далее, что условная плотность распределения величины M_{11} при заданных r — числе попаданий в состояние один, i_0 и i_n — начальном и конечном состоянии определяется как

$$\frac{c_{i_0 i_n}(r, m_{11})\, e^{\psi m_{11}}}{\Sigma c_{i_0 i_n}(r, t)\, e^{\psi t}},$$

где комбинаторный коэффициент в числителе равен числу различных бинарных последовательностей с заданными начальными и конечными состояниями, с r единицами и $n+1-r$ нулями, а также с требуемым значением величины m_{11}.

Решение

Привлекая соображения, приведенные при обсуждении задачи 2.2, получим

$$\operatorname{lik}(\theta;\, y_1, \ldots, y_n \mid i_0) = \prod_{u,\,v} \theta_{uv}^{m_{uv}}.$$

Далее,

$$m_{0\cdot} = n - r - \delta_{0\, i_n}, \quad m_{1\cdot} = r - \delta_{1\, i_n},$$
$$m_{\cdot 0} = n - r - \delta_{0\, i_0}, \quad m_{\cdot 1} = r - \delta_{1\, i_0},$$

где $\delta_{ab} = 1$ при $a = b$ и $\delta_{ab} = 0$ при $a \neq b$, а r — число единиц.

Из определения параметров λ и ψ имеем

$$\theta_{01} = e^{\lambda}/(1+e^{\lambda}),\ \theta_{11} = e^{\lambda+\psi}/(1+e^{\lambda+\psi}).$$

Правдоподобие, таким образом, принимает вид

$$\frac{e^{\lambda m_{01}}\, e^{(\lambda+\psi) m_{11}}}{(1+e^{\lambda})^{m_{0\cdot}}\,(1+e^{\lambda+\psi})^{m_{1\cdot}}} = \frac{e^{\lambda m_{\cdot 1} + \psi m_{11}}}{(1+e^{\lambda})^{m_{0\cdot}}\,(1+e^{\lambda+\psi})^{m_{1\cdot}}},$$

где $m_{\cdot 1} = m_{1\cdot} + \delta_{1 i_n} - \delta_{1 i_0}$.

Совместная условная вероятность равенства $(R, M_{11}) = (r, m_{11})$ при заданных (i_0, i_n) равна, по определению,

$$\frac{c_{i_0 i_n}(r, m_{11})\, e^{\lambda m_{\cdot 1} + \psi m_{11}}}{(1+e^{\lambda})^{m_{0\cdot}}\,(1+e^{\lambda+\psi})^{m_{1\cdot}}}.$$

Отсюда следует, что вероятность события $R=r$ равна

$$\sum_t c_{i_0 i_n}(r,\, t)\, e^{\lambda m_{\cdot 1}+\psi t}/\{(1+e^{\lambda})^{m_{0\cdot}}(1+e^{\lambda+\psi})^{m_{1\cdot}}\}.$$

Таким образом,

$$\operatorname{pr}(M_{11}=m_{11}\,|\,R=r,\ i_0,\ i_n)=\frac{c_{i_0 i_n}(r,\, m_{11})\, e^{\psi m_{11}}}{\sum_t c_{i_0 i_n}(r,\, t)\, e^{\psi t}}.$$

Заметим, что параметр λ исключен, поскольку статистика $(R,\ I_0,\ I_n)$ достаточна для λ при всех фиксированных ψ. Дальнейший анализ проводится так же, как и при сравнении двух биномиальных вероятностей.

[Теоретическая статистика, § 5.2; Кокс, 1970, § 5.7; Биллингсли, 1961a]

5.5. Пусть н. о. р. случайные величины $Y_1,\ \ldots,\ Y_n$ имеют показательное распределение с параметром положения θ_1 и средним $\theta_1+\theta_2$. Постройте соответствующий подобный или инвариантный критерий для проверки гипотез:

(a) H_0: $\theta_1=0$ против H_A: $\theta_1>0$;

(b) H_0: $\theta_2=1$ против H_A: $\theta_2\neq 1$.

Решение

Плотность одного наблюдения имеет вид

$$\theta_2^{-1}\exp\{-(y-\theta_1)/\theta_2\}\,\operatorname{hv}(y-\theta_1).$$

Следовательно,

$$\operatorname{lik}(\theta;\ y_1,\ \ldots,\ y_n)=\begin{cases}\theta_2^{-n}\exp\{-(\Sigma y_j-n\theta_1)/\theta_2\} & \text{при } y_{(1)}\geqslant\theta_1,\\ 0 \text{ в противном случае,} & \end{cases}$$

откуда вытекает, что $(\Sigma Y_j,\ Y_{(1)})$ — минимальная достаточная статистика.

(a) H_0: $\theta_1=0$ против H_A: $\theta_1>0$. Эта задача инвариантна относительно преобразований вида $y\to ay$, для которых максимальным инвариантом будет $Y_{(1)}/\Sigma Y_j$ или (более удобная статистика)

$$T=nY_{(1)}/(\Sigma Y_j-nY_{(1)}).$$

Из свойств порядковых статистик показательного распределения (задача 3.9) известно, что $X_1=n(Y_{(1)}-\theta_1)$ и $X_2=\Sigma Y_j-nY_{(1)}$ независимы и распределены как $\frac{1}{2}\theta_2\chi_2^2$ и $\frac{1}{2}\theta_2\chi_{2(n-1)}^2$ случайные величины. Наиболее мощный инвариантный критерий, построенный на основе статистики T, отвергает H_0 при больших значениях статистики T. При выполнении гипотезы H_0 статистика $(n-1)T$ имеет стандартное F-распределение с $(2, 2n-2)$ степенями свободы.

Тот же самый критерий получается в том случае, если рассматриваются условные относительно ΣY_j распределения. При выполнении гипотезы H_0 статистика ΣY_j достаточна.

(b) H_0: $\theta_2 = 1$ против H_A: $\theta_2 \neq 1$. Эта задача инвариантна относительно преобразований $y \to y + b$. Соответствующим максимальным инвариантом является $\Sigma Y_j - nY_{(1)} = \frac{1}{2}\theta_2\chi^2_{2(n-1)}$. Следовательно, инвариантный критерий совпадает с дисперсионным критерием в нормальной модели (см. задачу 4.9).

Тот же самый критерий получается при рассмотрении условных относительно $Y_{(1)}$ распределений. При выполнении гипотезы H_0 статистика $Y_{(1)}$ достаточна.

[Теоретическая статистика,
§ 5.2, 5.3; Леман, § 4.3, 4.4,
гл. 6]

5.6. Пусть независимые случайные величины $Y_1, \ldots, Y_n$ таковы, что некоторая, но неизвестно какая именно, перестановка этих величин приводит к гамма-распределениям, с плотностями

$$\rho_j (\rho_j y)^{\beta-1} e^{-\rho_j y}/\Gamma(\beta),$$

зависящими от двух неизвестных параметров ψ и λ через соотношение регрессии $\rho_j = \exp\{\lambda + \psi z_j\}$, где величины z_j известны. Убедитесь, что задача проверки гипотезы H_0: $\psi = 0$ против альтернативы $\psi \neq 0$ инвариантна относительно перестановок наблюдений и изменения масштаба, а максимальным инвариантом является $(Y_{(1)}/Y_{(n)}, \ldots, Y_{(n-1)}/Y_{(n)})$. Покажите, что локально наиболее мощный инвариантный критерий имеет области отвержения гипотезы, образованные большими значениями отношения $\Sigma(Y_j^2)/(\Sigma Y_j)^2$.

Решение

Совместная п. р. в. вектора $Y = (Y_1, \ldots, Y_n)$ определяется как

$$\frac{1}{n!}\Sigma^\bullet \prod_{j=1}^{n} \rho_{i_j} (\rho_{i_j} y_j)^{\beta-1} \exp(-\rho_{i_j} y_j)/\Gamma(\beta)$$

с $\rho_j = \exp(\lambda + \psi z_j)$, где суммирование в $\Sigma^\bullet$ проводится по всем перестановкам $(i_1, \ldots, i_n)$ ряда $(1, \ldots, n)$. Выводы являются автоматически инвариантными относительно перестановок, поскольку зависят только от $Y_{(j)}$. Преобразования $y \to ay$, $z \to z+b$ порождают преобразования $(\psi, \lambda, \beta) \to (\psi, \lambda+\log a + \psi b, \beta)$, оставляющие H_0 и H_A инвариантными. Таким образом,

$$T = (Y_{(1)}/Y_{(n)}, \ldots, Y_{(n-1)}/Y_{(n)})$$

— максимальный инвариант. Без потери общности можно положить $\bar{z}. = 0$.

Пусть $V = Y_{(n)}$. Тогда якобиан перехода $Y_{(\cdot)} \to (T, V)$ равен $V^{-(n-1)}$ и совместная п. р. в. вектора T определяется как

$$f_T(t;\ \psi, \beta) = \{\Gamma(\beta)\}^{-n} \Sigma^* \int_0^\infty v^{n-1} \prod_{j=1}^{n} \rho_{i_j} (\rho_{i_j} v t_j)^{\beta-1} \exp(-\rho_{i_j} v t_j)\, dv,$$

где $t_n = 1$. Таким образом,

$$f_T(t;\ \psi, \beta) = \frac{\Gamma(n\beta)}{\{\Gamma(\beta)\}^n} \prod_{j=1}^{n} t_j^{\beta-1} \Sigma^* \left\{ \sum_{j=1}^{n} t_j \exp(\psi z_{i_j}) \right\}^{-n\beta},$$

$$\left[\frac{\partial f_T}{\partial \psi}\right]_{\psi=0} \propto n!\beta (\Sigma t_j)(\Sigma z_j) - n\beta \Sigma^* (\Sigma z_{i_j} t_j) = 0,$$

откуда вытекает, что $U.(0, \beta) = 0$. Следовательно, главный член в разложении разности $\log f_T(t;\ \psi, \beta) - \log f_T(t;\ 0, \beta)$ равен $\psi^2 u'(0, \beta)$. Отсюда вытекает, что локально наиболее мощная статистика критерия имеет вид

$$\left[\frac{\partial^2 \log f_T}{\partial \psi^2}\right]_{\psi=0} \propto \frac{\Sigma^* \{(n\beta+1)(\Sigma z_{i_j} t_j)^2 - (\Sigma t_k)(\Sigma z_{i_l}^2 t_l)\}}{(\Sigma t_k)^2} =$$

$$= \text{const} + \frac{(n-1)\Sigma z_j^2 \Sigma t_k^2}{(\Sigma t_j)^2} \propto \text{const} + \frac{\Sigma y_j^2}{(\Sigma y_j)^2}.$$

Таким образом, при использовании локально наиболее мощного критерия большие значения статистики $\Sigma y_j^2/(\Sigma y_j)^2$ являются значимыми.

[Теоретическая статистика, § 5.3; Фергюсон, 1961]

5.7. Пусть Y_{jk} $(j = 1, \ldots, m;\ k = 1, \ldots, r)$ подчиняются нормальной модели дисперсионного анализа со случайными факторами, т. е. $Y_{jk} = \mu + \eta_j + \varepsilon_{jk}$, где η_j и ε_{jk} независимо нормально распределены с нулевыми средними и дисперсиями σ_b^2 и σ_w^2 соответственно. Покажите, что минимальной достаточной статистикой является $(\bar{Y}.., SS_b, SS_w)$, где $SS_b = r\Sigma(\bar{Y}_{j.} - \bar{Y}..)^2$, а $SS_w = \Sigma\Sigma(Y_{jk} - \bar{Y}_{j.})^2$. Имеются две независимые совокупности данных описанной выше структуры с одними и теми же значениями m и r. Необходимо проверить нулевую гипотезу, что отношение σ_b^2/σ_w^2 одинаково для обеих совокупностей, тогда как все другие параметры произвольны. Найдите соответствующую группу преобразований для применения теории инвариантности. Сокращая данные до минимальной достаточной статистики с последующим вычислением максимального инварианта, покажите, что соответствующей статистикой критерия является отношение двух значений величины SS_b/SS_w. Для распределения статистики критерия при нулевой гипотезе предложите простую аппроксимацию, использующую логарифмическое преобразование. Какого рода практическим ситуациям может отвечать эта задача?

Решение

На основе соображений, высказанных при обсуждении задачи 2.4, получаем, что для одной совокупности данных минимальной достаточной статистикой является $(\bar{Y}_{..}, SS_b, SS_w)$.

Далее, для H_0: $\sigma_{b1}^2/\sigma_{w1}^2 = \sigma_{b2}^2/\sigma_{w2}^2$ группа преобразований $\mathscr{L}$: $gY = aY + c$ оставляет задачу проверки гипотезы инвариантной при различных (a, c) для каждого множества. Действительно,

$$(\bar{Y}_{..1}, SS_{b1}, SS_{w1}) \to (a_1\bar{Y}_{..1} + c_1, a_1^2 SS_{b1}, a_1^2 SS_{w1}),$$
$$(\bar{Y}_{..2}, SS_{b2}, SS_{w2}) \to (a_2\bar{Y}_{..2} + c_2, a_2^2 SS_{b2}, a_2^2 SS_{w2}).$$

Отсюда вытекает, что максимальный инвариант представляется в виде $(SS_{b1}/SS_{w1}, SS_{b2}/SS_{w2}) = (T_1, T_2)$. Однако группа преобразований $\mathscr{L}$ не сокращает нулевую гипотезу до простой гипотезы. Инвариантная параметрическая функция имеет вид

$$(\sigma_{b1}^2/\sigma_{w1}^2, \sigma_{b2}^2/\sigma_{w2}^2).$$

Далее, $T_i = (1 + r\sigma_{bi}^2/\sigma_{wi}^2) F_i$ $(i = 1, 2)$, где F_1 и F_2 — независимые случайные величины, представимые в виде отношений нормированных оценок дисперсий. Положим $\lambda_i = 1 + r\sigma_{bi}^2/\sigma_{wi}^2$. Получим

$$f_{T_1, T_2}(t_1, t_2) = \lambda_1^{-1} g(t_1/\lambda_1)\, \lambda_2^{-1} g(t_2/\lambda_r), \qquad (1)$$

где $g(\cdot)$ — п. р. в. случайной величины $F_{m-1, mr-m}$. Далее, нулевая гипотеза H_0: $\sigma_{b1}^2/\sigma_{w1}^2 = \sigma_{b2}^2/\sigma_{w2}^2$ эквивалентна гипотезе H_0': $\lambda_1 = \lambda_2$ для статистической модели (1). Последняя задача проверки гипотез инвариантна относительно преобразований $T_i \to aT_i$, $(i = 1, r)$, $a > 0$. Соответствующим максимальным инвариантом является T_1/T_2. Это справедливо лишь в том случае, если значения r и m являются одинаковыми для обеих выборок.

Тот же самый конечный результат можно было бы получить, применяя к совокупности случайных величин $\{Y_{jk};\ k=1,$

..., r} сначала ортогональные преобразования, переводящие их в новые независимые нормальные случайные величины с новыми параметрами σ^2_{wi}, λ_i $(i=1, 2)$, а затем рассматривая общие сдвигово-масштабные преобразования.

При достаточно больших m для $\log SS_{bi}$ и $\log SS_{wi}$ можно использовать нормальную аппроксимацию. В этом случае, как и в задаче 5.3,

$$\log (T_1/T_2) = \log SS_{b1} - \log SS_{b2} + \log SS_{w2} - \log SS_{w1} \simeq \log (\lambda_1/\lambda_2) + \varepsilon,$$

где ε — нормально $N\{0, 4(m-1)^{-1} + 4m^{-1}(r-1)^{-1}\}$ распределенная случайная величина.

Рассмотренный выше анализ можно применить к сравнению чувствительности двух методик измерений.

[Теоретическая статистика, § 5.3; Дар, 1962]

5.8. Предположим, что н. о. р. случайные величины $Y_1, \ldots, Y_n$ распределены или нормально $N(\mu, \sigma^2)$, или равномерно на интервале $\left(\nu - \frac{1}{2}\tau, \ \nu + \frac{1}{2}\tau\right)$, параметры неизвестны. Покажите, что наиболее мощный критерий для проверки гипотез об этих двух распределениях, являющейся инвариантным относительно сдвиго-масштабных преобразований, в качестве статистики критерия имеет отношение $(Y_{(n)} - Y_{(1)})/\{\Sigma(Y_j - \bar{Y}_.)^2\}^{1/2}$.

Решение

Для любого распределения, п. р. в. которого $g(\cdot)$ зависит лишь от параметров сдвига и масштаба, максимальный инвариант представляется в виде

$$T = \{(Y_3 - Y_2)/(Y_2 - Y_1), \ \ldots, \ (Y_n - Y_2)/(Y_2 - Y_1)\}.$$

Чтобы найти плотность распределения максимально инвариантной статистики, положим $y_2 = u$, $y_2 - y_1 = v$. Получим

$$f_T(t; H_0) = \iint v^{n-2} g(u) \, g(u - v) \prod_{j=3}^{n} g(vt_j + u) \, du \, dv.$$

Критерий отношения правдоподобий, построенный на основе значений статистики T, представляется в виде отношения соответствующих форм полученных п. р. в. или, что эквивалентно, в виде отношений статистик

$$h(y; g) = \int_0^{\infty} \int_{-\infty}^{\infty} a^{n-2} \prod_{j=1}^{n} g(ay_j - b) \, da \, db.$$

Простые вычисления показывают, что для нормального распределения

$$h(y;\ \text{норм. распр.}) \propto \{\Sigma(y_j - \bar{y}.)^2\}^{-\frac{1}{2}(n-1)},$$

а для равномерного распределения

$$h(y;\ \text{равном. распр.}) = \int\limits_0^{r_n^{-1}} \int\limits_{ay_{(n)}^{-1}}^{ay_{(1)}} a^{n-2}\,da\,db \propto r_n^{-(n-1)},$$

где $r_n = y_{(n)} - y_{(1)}$. Таким образом, отношение правдоподобий представляется в виде $(n-1)$-й степени дроби $(y_{(n)} - y_{(1)})/\{\Sigma \times \times (y_j - \bar{y}.)^2\}^{1/2}$, которую часто называют стьюдентизированной ранговой статистикой (Пирсон и Хартли, 1970, табл. 29)

[Теоретическая статистика, § 5.3; Уткофф, 1970]

5.9. Производятся независимые наблюдения из многомерного нормального распределения с неизвестными средним и коварнационной матрицей. Нулевой гипотезой H_0 утверждается, что первые q координат не зависят от оставшихся $p-q$ координат. Альтернативная гипотеза является общей гипотезой о нормальности распределения наблюдений. Найдите такую подходящую группу преобразований, чтобы максимально-инвариантной статистикой была совокупность канонических корреляций. Это означает, что если выборочная матрица попарных произведений SS разбита на блоки SS_{11}, SS_{12}, SS_{21} и SS_{22}, отвечающие первым q и последним $p-q$ координатам, то максимальный инвариант есть совокупность решений уравнения $|SS_{11} - l SS_{12} SS_{22}^{-1} SS_{21}| = 0$. Какая функция от максимального инварианта используется в критерии максимума отношения правдоподобий?

Решение

Минимальной достаточной статистикой является пара $(\bar{Y}, SS)$. Пусть

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^{\mathrm{T}} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

где $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(Y_i, Y_j)$, а Y_1 — это первые q координат вектора Y. Тогда нулевая гипотеза H_0: $\boldsymbol{\Sigma}_{12} = 0$, параметры μ, $\boldsymbol{\Sigma}_{11}$ и $\boldsymbol{\Sigma}_{12}$ — произвольны и неизвестны.

Рассмотрим общие невырожденные линейные преобразования

$$g\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{bmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

порождающие преобразования

$$g^*\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{bmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

в пространстве параметров. При вычислении ковариационной матрицы преобразованных векторов увидим, что $g^*\mathbf{\Sigma}_{12} = \mathbf{0}$ при заданном $\mathbf{\Sigma}_{12} = \mathbf{0}$ в том и только том случае, если $\mathbf{a}_{21} = \mathbf{0}$, $\mathbf{a}_{12} = \mathbf{0}$. В остальном преобразование произвольно. Итак, задача проверки гипотез инвариантна относительно преобразований

$$g\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{22} \end{bmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Эти преобразования отображают $(\overline{Y}, \mathrm{SS})$ в

$$\begin{pmatrix} \mathbf{a}_{11}\overline{Y}_1 + b_1 \\ \mathbf{a}_{22}\overline{Y}_2 + b_2 \end{pmatrix}, \quad \begin{bmatrix} \mathbf{a}_{11}\mathbf{SS}_{11}\mathbf{a}_{11}^{\mathrm{T}} & \mathbf{a}_{11}\mathbf{SS}_{12}\mathbf{a}_{22}^{\mathrm{T}} \\ \mathbf{a}_{22}\mathbf{SS}_{12}^{\mathrm{T}}\mathbf{a}_{11}^{\mathrm{T}} & \mathbf{a}_{22}\mathbf{SS}_{22}\mathbf{a}_{22}^{\mathrm{T}} \end{bmatrix}.$$

Среди этих двух матриц только из второй матрицы можно получить инвариантную статистику.

Рассмотрим конкретное преобразование, определяемое $\mathbf{a}_{11} = \mathbf{SS}_{11}^{-1/2}$ и $\mathbf{a}_{22} = \mathbf{SS}_{22}^{-1/2}$. Оно переводит матрицу $\mathbf{SS}$ в следующую:

$$\begin{bmatrix} \mathbf{I} & \mathrm{SS}_{11}^{-1/2}\mathrm{SS}_{12}\mathrm{SS}_{22}^{-1/2} \\ \mathrm{SS}_{21}^{-1/2}\mathrm{SS}_{12}^{\mathrm{T}}\mathrm{SS}_{11}^{-1/2} & \mathbf{I} \end{bmatrix}.$$

Предположим, что матрица $\mathbf{G} = \mathrm{SS}_{11}^{-1}\mathrm{SS}_{12}\mathrm{SS}_{22}^{-1}\mathrm{SS}_{12}^{\mathrm{T}}$ для двух множеств данных имеет одинаковые наборы ненулевых собственных значений общим числом $\min(q, p-q)$. Тогда существует такая невырожденная матрица $\mathbf{G}$, что $\mathbf{G} = \mathbf{c}^{-1}\tilde{\mathbf{G}}\mathbf{c}$[1]). Аналогичные соображения применяются к матрице $\mathrm{SS}_{22}^{-1}\mathrm{SS}_{12}^{\mathrm{T}}\mathrm{SS}_{11}^{-1}\mathrm{SS}_{12}$, роль матрицы C будет играть некоторая невырожденная матрица $\mathbf{d}$. Из этих двух утверждений вытекает, что

$$\widetilde{\mathrm{SS}} = \begin{bmatrix} \mathbf{c}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{d}^{\mathrm{T}} \end{bmatrix} \mathrm{SS} \begin{bmatrix} \mathbf{c} & \mathbf{d} \\ \mathbf{0} & \mathbf{d} \end{bmatrix}.$$

Таким образом, множества данных связаны преобразованием,

---

[1]) Статистики, снабженные знаком „тильда", вычислены для второго множества данных.— *Прим. перев.*

форма которого определена выше. Можно проверить, что максимальным инвариантом являются собственные значения.

Функция правдоподобия при заданных $y_1, \ldots, y_n$ определяется как $\operatorname{lik}(\mu, \Sigma; y) \propto |\Sigma|^{-\frac{1}{2}n} \exp\left\{-\frac{1}{2}\Sigma(y_i-\mu)^T\Sigma^{-1}(y_j-\mu)\right\}$ и является функцией параметров (μ, Σ). По μ она максимизируется для $\mu = \bar{y}.$ при любой Σ. Далее, в $\Omega = \Omega_A \cup \Omega_0$ оценкой максимума правдоподобия ковариационной матрицы будет $\hat{\Sigma} = n^{-1}SS$, а в Ω_0

$$\hat{\Sigma} = n^{-1}\begin{bmatrix} SS_{11} & 0 \\ 0 & SS_{22} \end{bmatrix}.$$

Отсюда совсем просто получить, что статистика максимума правдоподобия является функцией величины

$$V = |SS|/\{|SS_{11}||SS_{22}|\}.$$

Далее,

$$\begin{aligned}|SS| &= |SS_{11} - SS_{12}SS_{22}^{-1}SS_{12}^T||SS_{22}| = \\ &= |SS_{22} - SS_{12}^TSS_{11}^{-1}SS_{12}||SS_{11}|,\end{aligned}$$

откуда

$$V = |I - SS_{11}^{-1}SS_{12}SS_{22}^{-1}SS_{12}^T| = |I - SS_{22}^{-1}SS_{12}^TSS_{11}^{-1}SS_{12}| = \Pi(1 - l_i),$$

где l_i — собственные значения определенной выше матрицы **G**. Могут использоваться, конечно, и другие функции собственных значений.

[Теоретическая статистика, § 5.3, 5.4; Андерсон (1958), гл. 12; Рао, § 8c, 8f]

5.10. Выборочное пространство двумерной случайной величины Y состоит из двух концентрических окружностей и их центра. При нулевой гипотезе H_0 центральная точка имеет известную вероятность p_0, а внутренняя и внешняя окружности имеют полные вероятности $1-2p_0$ и p_0 соответственно. Распределения на каждой окружности являются равномерными. При сложной альтернативной гипотезе центр окружностей имеет известную вероятность $p_A > p_0$, внутренняя окружность имеет нулевую вероятность, на внешней окружности определено произвольное ограниченное распределение. Таким образом, задача является инвариантной относительно поворотов плоскости вокруг центральной точки. Найдите максимально инвариантную статистику и опишите тем самым равномерно наиболее мощный инвариантный критерий для проверки H. Сравните этот критерий с критерием, основанным на максимуме отношения правдоподобий.

Решение

Если $Y=(Y_1, Y_2)^{\mathrm{T}}$, тогда подмножества параметрического пространства Ω_0 и Ω_A остаются инвариантными относительно группы преобразований вида

$$gY=\begin{bmatrix}\sin a & \cos a\\ \cos a & -\sin a\end{bmatrix}Y,$$

где за общий центр окружностей взято начало координат. Максимальным инвариантом является статистика $T=Y_1^2+Y_2^2$, которая имеет три возможных значения: 0, t_1 и $t_2>t_1$. Имеем

$$\begin{gathered}\operatorname{pr}(T=0;\ H_0)=p_0,\ \operatorname{pr}(T=0;\ H_A)=p_A>p_0,\\ \operatorname{pr}(T=t_1;\ H_0)=1-2p_0,\ \operatorname{pr}(T=t_1;\ H_A)=0,\\ \operatorname{pr}(T=t_2;\ H_0)=p_0,\ \operatorname{pr}(T=t_2;\ H_A)=1-p_A.\end{gathered}$$

Вид критерия отношения правдоподобия зависит от выполнения или невыполнения неравенства $1-p_A\geqslant p_A$. Если $p_A\leqslant\frac{1}{2}$, то

$$\operatorname{lr}_{A0}(t_2)\geqslant\operatorname{lr}_{A0}(0)>\operatorname{lr}_{A0}(t_1)=0.$$

Поскольку для произвольной непрерывной плотности на внешней окружности наблюденному значению y может соответствовать сколь угодно большое значение плотности вероятности, то максимум отношения правдоподобия неограничен, если $T=t_2$. Таким образом, если $p_A>\frac{1}{2}$, то два рассматриваемых критерия отличаются, хотя уровень значимости, приписываемый значению $T=t_1$, один и тот же в обоих случаях.

[Теоретическая статистика,
§ 5.3, 5.4; Леман, § 6.12]

5.11. Пусть н. о. р. случайные векторы $Y_1, \ldots, Y_{n_1}$ нормально $MN_p(\mu, \Sigma)$ распределены, а н. о. р. случайные векторы $Y_1', \ldots, Y_{n_2}'$ нормально $MN(\mu', \Sigma')$ распределены. Выведите инвариантный критерий для проверки гипотезы H_0: $\Sigma=\Sigma'$ против общей альтернативы. Для этого найдите наиболее значимую из статистик критерия $F(a)$, используемых при проверке соответствующих гипотез о дисперсиях скалярных комбинаций $a^{\mathrm{T}}Y$ и $a^{\mathrm{T}}Y'$.

Решение

Итак, случайная величина a^TY_j нормально $N(a^T\mu, a^T\Sigma a)$ распределена, а случайная величина $a^TY'_j$ нормально $N(a^T\mu', a^T\Sigma' a)$ распределена.

Пусть $a^T\Sigma a=\sigma_a^2$, $a^T\Sigma' a=\tau_a^2$. Критерии проверки гипотез H_0: $\sigma_a^2=\tau_a^2$ инвариантны относительно сдвигово-масштабных преобразований случайных величин a^TY_j и $a^TY'_j$. Это приводит к использованию критериев, основанных на статистиках

$$MS_a/MS'_a=(a^TMSa)/(a^TMS'a)U_a.$$

Как большие, так и малые значения статистики U_a свидетельствуют против H_0. Таким образом, наиболее значимыми являются наибольшее и наименьшее значения статистики U_a, когда a пробегает все возможные значения.

В силу инвариантности относительно шкалы измерений, без потери общности можно предположить, что $a^TMS'a=1$. Максимизация статистики U_a эквивалентна максимизации величины

$$a^TMSa-\lambda a^TMS'a.$$

Максимум написанной выше разности достигается при $(MS-\lambda MS')a=0$, т. е. в том случае, когда λ — собственное значение матрицы $MS(MS')^{-1}$, а a — собственный вектор. Наибольшее значение статистики U_a совпадает с наибольшим собственным значением l_p. Аналогично, наименьшее собственное значение l_1 матрицы $MS(MS')^{-1}$ совпадает с наименьшим значением статистики U_a. Таким образом, построение критерия проверки гипотезы H_0: $\Sigma=\Sigma'$ включает в себя вычисление двух уровней значимости

$$p^+=\mathrm{pr}(L_p\geqslant l_p;\ \Sigma=\Sigma'),\quad p^-=\mathrm{pr}(L_1\leqslant l_p;\ \Sigma=\Sigma').$$

[Теоретическая статистика, § 5.4]

6. Критерии, свободные от распределения, и критерии рандомизации

Необходимые сведения

Термины непараметрический и свободный от распределения используются, заменяя друг друга, для обозначения критериев, в которых в нулевые гипотезы в явном или неявном виде включаются произвольные неизвестные распределения. Наиболее просто формулируемая среди таких гипотез состоит в том, что $Y_1, \ldots, Y_n$ — н. о. р. с одной и той же неизвестной плотностью распределения. В качестве возможных отклонений от нее укажем следующие: (a) наличие тренда, зависящего от порядка последовательности; (b) зависимость от контролируемой переменной; (c) систематическое различие между наблюдениями с номерами $1, \ldots, n_1$ и наблюдениями с номерами $n_1+1, \ldots, n$, т. е. задача о двух выборках. Критерий называют свободным от распределения, если распределение статистики критерия является одним и тем же для некоторого богатого класса плотностей, например в общем случае для всех плотностей или для всех непрерывных плотностей. Как правило, такие критерии получают при обращении к распределениям, условным относительно вариационного ряда выборки. При нулевой гипотезе все перестановки порядковых статистик равновероятны. Критерии, построенные таким образом, называются перестановочными критериями.

В соответствии с рекомендациями гл. 3, наиболее простой подход состоит в том, чтобы выбрать неформальным образом статистику критерия, например, по аналогу с некоторой параметрической задачей. Перестановочное распределение рассматриваемой статистики можно найти методом перебора или на основе приближения, использующего несколько первых моментов статистики. Часто статистики критерия можно значительно упростить, поскольку вариационный ряд считается фиксированным. Например, для задачи двух выборок t-статистика Стьюдента эквивалентна статистике выборочных средних.

Двувыборочный критерий Вилкоксона получается на основе ранговых чисел, а не из первоначальных наблюдений и использования в качестве статистики критерия суммы (или среднего) рангов в одной из выборок. Это важный частный случай критериев, в которых от наблюдений используется только их ранговый порядок.

Один из формальных подходов к выбору ранговых статистик критерия связан с нахождением ранговой статистики, имеющей максимальную локальную мощность для некоторых параметрических вариантов задачи. Это означает, что параметрическая постановка используется для нахождения статистики критерия. Распределение при нулевой гипотезе получается на основе метода перестановок, а тем самым сохраняется свойство свободы от распределения. Важным классом ранговых статистик являются линейные ранговые статистики. Они образуются заменой линейных статистик $\Sigma c_j Y_j$ на статистики $\Sigma c_j w(R_j)$, где R_j — ранговый номер наблюдения Y_j. Например, для задачи двух выборок конкретный выбор "весовой функции" $w(k)=k$ приводит к критерию Вилкоксона. Имеются общие результаты об асимптотической нормальности линейных ранговых статистик.

Совершенно другой подход к обоснованию правомерности использования критериев перестановок применяется при рассмотрении экспериментов, в которых воздействия размещаются по экспериментальным единицам случайно. При этом вероятностные вычисления основываются не на предположениях независимости и одинаковой распределенности случайных величин, а на использовании специального рандомизирующего устройства. Критерии, при построении которых используется указанный подход, назовем критериями рандомизации.

Другой путь построения критериев состоит в том, чтобы для рассматриваемой задачи определить общую меру расстояния, например в задаче двух выборок определить общую меру расстояния между функциями распределения двух популяций. После этого, используя эмпирические функции распределения, указанные расстояния можно оценить и тем самым получить статистику критерия. Свойство свободы от распределения может как выполняться, так и не выполняться. Другой метод заключается в том, чтобы получать для указанного расстояния несмещенную оценку с равномерно минимальной дисперсией на основе теоремы Рао — Блэкуэлла (см. гл. 8).

Задачи

6.1. Предположим, что $Y_1, \ldots, Y_n$ — н. о. р. с плотностью $f(y-\theta)$, где $f(\cdot)$ — неизвестная непрерывная плотность, симметричная относительно нуля. Рассмотрим задачу проверки нулевой гипотезы H_0: $\theta=0$ против альтернативы $\theta>0$. Покажите, что H_0 эквивалентна гипотезе сравнения положительных и отрицательных значений выборки. Постройте критерий рангового типа для проверки H_0.

Решение

Удобно записать $Y_j=(I_j, Z_j)$, где $Z_j=|Y_j|$, а $I_j=1$ или 0 в соответствии с тем, что $Y_j=+Z_j$ или $Y_j=-Z_j$. При нулевой гипотезе $\theta=0$ случайные величины I_j и Z_j независимы и $\operatorname{pr}(I_j=1)=\frac{1}{2}$. Если $\theta>0$, то при $I_j=1$ случайные величины Z_j стохастически больше и $\operatorname{pr}(I_j=1)>\frac{1}{2}$. Таким образом, если сгруппировать наблюдения в соответствии с их знаками и определить $R_j=\operatorname{rank}(Z_j)$, то можно ожидать, что при $\theta>0$: (a) ранги, отвечающие положительным наблюдениям, будут иметь тенденцию принимать бо́льшие значения в сравнении с рангами отрицательных наблюдений, и (b) положительных наблюдений будет больше отрицательных.

Один из прямых аналогов двувыборочной статистики Вилкоксона имеет вид $T=\Sigma I_j R_j$. Конечно, ΣI_j сама является естественной статистикой. Она определяет критерий знаков. При нулевой гипотезе условное распределение статистики T при заданном значении $\Sigma I_j=m$ имеет такое же распределение, как и двувыборочная статистика Вилкоксона (см. задачу 6.2). Таким образом, усредняя его по биномиальному распределению статистики ΣI_j, получим распределение статистики T при нулевой гипотезе. В частности, используя факт, что при H_0

$$E(T)=\frac{1}{4}n(n+1),\ \operatorname{var}(T)=\frac{1}{24}n(n+1)(2n+1),$$

можно получить для распределения статистики T нормальную аппроксимацию.

Для заданной плотности $f(\cdot)$ можно развить общую теорию оптимальных ранговых критериев. Совместная плотность пар (I_j, R_j) для $j=1, \ldots, n$ записывается в следующем виде:

$$2^{-n}E\left[\prod_{j=1}^{n}\left\{f(Z^{(r_j)}+\theta)\right\}^{i_j}\left\{f(Z^{(r_j)}-\theta)\right\}^{1-i_j}\Big/ f(Z^{(j)});\ H_0\right],$$

где $Z^{(1)}\leqslant\ldots\leqslant Z^{(n)}$ обозначает вариационный ряд, составленный из величин Z_j. Используя эту плотность, получим статистику локально наиболее мощного рангового критерия

$$-E\left\{\Sigma i_j\frac{f'(Z^{(r_j)})}{f(Z^{(r_j)})};\ H_0\right\},$$

распределение которой при нулевой гипотезе определяется перестановочным критерием рангов R_j и бернуллиевскими вероятностями знаков I_j.

[Теоретическая статистика,
§ 6.2; Рао, § 7е; Силвей, § 9.3,
9.4; Леман, § 6.9]

6.2. Для получения первых четырех моментов выборочного среднего в случайных выборках объема n_1, извлеченных без возвращения из конечной совокупности $\{y_1, \ldots, y_n\}$, используйте соображения, связанные с симметрией, или какие-либо другие соображения. Исследуя эксцесс и асимметрию, покажите, что скорость приближения к нормальному распределению при $n \to \infty$, $n_1/n \to a$, $0 < a < 1$, для выборки без возвращения быстрее, чем для выборки с возвращением. Приведите некоторые явные формулы для выборки из конечной популяции $\{1, \ldots, n\}$. Продемонстрируйте применение этих результатов к двувыборочному критерию Вилкоксона.

Решение

Существуют различные пути вычисления моментов статистик при случайном выборе без возвращения из конечной популяции. С математической точки зрения наиболее привлекательным представляется использование соображений, связанных с симметрией. Обозначим $m_r = \Sigma (y_j - m_1)^r/n$ моменты относительно среднего конечной популяции, а $\bar{Y}$ — выборочное среднее. Из соображений, связанных с симметрией, можно показать сначала, что $E\bar{Y} = a_1 m_1$, где a_1 — постоянная, зависящая от n и n_1, но не зависящая от значений $y_1, \ldots, y_n$. Рассматривая частный случай популяции с $y_1 = \ldots = y_n = 1$, получим $a_1 = 1$. Из аналогичных соображений вытекает, что

$$\operatorname{var}(\bar{Y}) = a_2 m_2, \quad \mu_3(\bar{Y}) = a_3 m_3, \quad \mu_4(\bar{Y}) = a_4 m_4 + b_4 m_2^2,$$

где постоянные a_2, a_3, a_4, b_4 выбираются соответствующим образом. Рассматривая популяции с конкретными значениями $y_1, \ldots$ $\ldots, y_n$, получим, что

$$a_2 = \frac{n_1 (n - n_1)}{(n-1)}, \quad a_3 = \frac{n_1 (n - n_1)(n - 2n_1)}{(n-1)(n-2)},$$

$$a_4 = \frac{(n^2 - 6nn_1 + n + 6n_1^2)\, n_1 (n - n_1)}{(n-1)(n-2)(n-3)},$$

$$b_4 = \frac{3(n_1 - 1)\, n\, (n - n_1 - 1)\, n_1 (n - n_1)}{(n-1)(n-2)(n-3)}.$$

Теперь можно вычислить стандартные меры ассимметрии и эксцесса для $\bar{Y}$. Если положить $n_1/n = f$ и устремить $n \to \infty$, асимптотически получим

$$\frac{\mu_3(\bar{Y})}{\{\operatorname{var}(\bar{Y})\}^{3/2}} \sim \frac{(1-2f)\, m_3/m_2^{3/2}}{(1-f)\sqrt{n_1}},$$

$$\frac{\mu_4(\bar{Y})}{\{\operatorname{var}(\bar{Y})\}^{2}} - 3 \sim \frac{-6}{n} + \left\{\frac{1-6f(1-f)}{(1-f)}\right\} \frac{(m_4/m_2^2-3)}{n_1}.$$

При выборке с возвращением, порождающей н. о. р. случайные величины, должно иметь место, например, следующее соотношение:

$$\frac{\mu_3(\bar{Y})}{\{\operatorname{var}(\bar{Y})\}^{3/2}} \sim \frac{m_3/m_2^{3/2}}{\sqrt{n_1}}.$$

Таким образом, полученное соотношение показывает, что более близко к нормальному распределение статистики $\bar{Y}$ в случае выборки без возвращения. Заметим, в частности, что если $f = \frac{1}{2}$, то распределение симметрично. Это с очевидностью вытекает из соотношения $\frac{1}{2}\bar{Y} + \frac{1}{2}\bar{Y}' = m_1 = \text{const}$, где $\bar{Y}'$ — среднее доли y_j-х, не попавших в выборку.

Применить рассмотренную теорию к двувыборочному критерию Вилкоксона можно следующим образом. Рассмотрим наблюдения двух независимых выборок объемов n_1 и n_2 как одно множество $n = n_1 + n_2$ значений и упорядочим их. Пусть статистикой критерия является среднее значение рангов в первой выборке. Тогда если выполняется нулевая гипотеза о тождественности исходных распределений, то статистика критерия представляет собой среднее выборки объема n_1, извлеченной случайно без возвращения из "конечной популяции" рангов. В частности, если нет совпадающих значений в исходных выборках, то конечная популяция представляется рядом $\{1, 2, \ldots, n\}$ и необходимые значения моментов $m_1, \ldots, m_4$ легко получаются. При нулевой гипотезе сходимость распределения статистики Вилкоксона к нормальному довольно быстрая.

[Теоретическая статистика, § 6.2; Рао, § 7е; Силвей, § 9.3]

6.3. Предположим, что для сопоставленных парных экспериментов наблюденные разности откликов между способом воздействия A и способом воздействия B равны соответственно 8, 5, 4, —2, —2, 0 и 1. Используя метод перебора, покажите, что точный двусторонний уровень значимости для проверки нулевой гипотезы об эквивалентности способов воздействия равен 1/4. Сравните это с результатами (a) применения "обычного" одновыборочного t-критерия Стьюдента; (b) использования нормальной аппроксимации с поправкой на непрерывность для перестановочного распределения; (c) аппроксимации перестановочного распределения, использующей первые четыре момента.

Решение

Разности откликов $(x_1, \ldots, x_7) = (8, 5, 4, -2, -2, 0, 1)$ возникают при случайном распределении способов воздействия внутри каждой пары индивидуумов. Таким образом, множество исходов для разностей откликов $(Y_1, \ldots, Y_7)$ образуется совокупностью из $2^7 = 128$ векторов $(\pm x_1, \ldots, \pm x_7)$. Все эти векторы при нулевой гипотезе наблюдаются с равными вероятностями. Таблица, приведенная ниже, показывает, какие выборки $(y_1, \ldots, y_7)$ имеют значение суммы Σy_j по крайней мере столь же большое, как и наблюденное значение 14.

x_j	-2	-2	0	1	4	5	8	Σy_j	Частота
	$-$	$-$	$\pm$	$+$	$+$	$+$	$+$	22	2
	$-$	$-$	$\pm$	$-$	$+$	$+$	$+$	20	2
	$+$	$-$	$\pm$	$+$	$+$	$+$	$+$	18	2
	$-$	$+$	$\pm$	$+$	$+$	$+$	$+$	18	2
	$-$	$+$	$\pm$	$-$	$+$	$+$	$+$	16	2
	$+$	$-$	$\pm$	$-$	$+$	$+$	$+$	16	2
	$-$	$-$	$\pm$	$+$	$-$	$+$	$+$	14	2
	$+$	$+$	$\pm$	$+$	$+$	$+$	$+$	14	2

Итак, при нулевой гипотезе

$$\operatorname{pr}\left(\sum Y_j \geqslant 14\right) = \frac{16}{128} = 0.125,$$

что является односторонним уровнем значимости.

Для t-критерия Стьюдента получим $t = 1.40$. Соответствующая статистика имеет 6 степеней свободы, а значит односторонний уровень значимости равен 0.1.

Отметим, что при нормальной аппроксимации с поправками на дискретность поправка для верхнего хвоста распределения равна -1, поскольку значения суммы $\sum y_j$ отличаются на 2. Итак, с поправкой на непрерывность нормированным значением будет

$$z = \left(\sum y_j - 1\right) / \left(\sum x_j^2\right)^{1/2} = 1.22;$$

односторонний уровень значимости этой величины равен 0.111.

Чтобы получить улучшенную нормальную аппроксимацию на основе разложения Эджворта, необходимо вычислить первые четыре рандомизированных момента для $\sum Y_j$. Они равны

$$0,\ \sum x_j^2 = 114,\ 0,\ 3(\sum x_j^2)^2 - 2(\sum x_j^4) = 28\,968.$$

Коэффициент асимметрии $\gamma_1 = 0$, а коэффициент эксцесса $\gamma_2 = -0.77$.

Разложение Эджворта для верхних хвостов распределения, соответствующее (подправленной) нормированной переменной z при $\gamma_1 = 0$, имеет вид

$$1 - \Phi(z) + \frac{\gamma_2}{24}(z^3 - 3z)\Phi(z).$$

При $z = 1.22$ это выражение принимает значение 0.123, что является превосходной аппроксимацией.

[Теоретическая статистика, § 6.2]

6.4. Пусть $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ — н.о.р. двумерные случайные величины с произвольным непрерывным распределением. Покажите, что: (a) ранговая статистика

$$T = \frac{1}{n(n-1)} \sum_{j>k} \operatorname{sgn}\{(Y_j - Y_k)(Z_j - Z_k)\} + \frac{1}{2}$$

эквивалентна τ-статистике Кендалла; (b) она имеет ожидание, равное вероятности того, что знаки $Y - Y'$ и $Z - Z'$ совпадают. Здесь (Y, Z) и (Y', Z') — независимые случайные векторы, имеющие то же самое распределение, что и (Y_j, Z_j), $j = 1, \ldots, n$. Для нулевой гипотезы независимости Y-ов и Z-ов покажите, что перестановочное распределение статистики T имеет среднее $1/2$ и дисперсию $(2n+5)/\{18n(n-1)\}$.

Покажите, что если двумерные пары имеют двумерное нормальное распределение с коэффициентом корреляции ρ, то ожидание статистики T имеет вид $\frac{1}{2} + \pi^{-1}\arcsin\rho$.

Решение

Обозначим ξ вероятность того, что $Y - Y'$ и $Z - Z'$ имеют одинаковые знаки. Пусть $I_{jk} = \operatorname{sgn}\{(Y_j - Y_k)(Z_j - Z_k)\}$. Тогда

$$E(I_{jk}) = \xi - (1 - \xi) = 2\xi - 1.$$

Далее, сумма, определяющая статистику T, содержит $\frac{1}{2}n(n-1)$ слагаемых, каждое из которых имеет одно и то же ожидание. Таким образом, $E(T) = \frac{1}{2}(2\xi - 1) + \frac{1}{2} = \xi$. Ранговый коэффициент корреляции Кендалла равен $2T-1$ и принимает значение из $[-1, 1]$.

Если Y и Z независимы, $\xi=\frac{1}{2}$, то $E(T)=\frac{1}{2}$. Далее,

$$\operatorname{var}(I_{jk})=E(I_{jk}^2)-\{E(I_{jk})\}^2=1,$$
$$\operatorname{cov}(I_{jk}, I_{lm})=0 \quad (j\neq k\neq l\neq m),$$

и для $j\neq k\neq l$

$$\operatorname{cov}(I_{jk}, I_{jl})=E[\operatorname{sgn}\{(Y_j-Y_k)(Z_j-Z_k)\}\operatorname{sgn}\{(Y_j-Y_l)(Z_j-Z_l)\}]=$$
$$=E[\operatorname{sgn}\{(Y_j-Y_k)(Y_j-Y_l)\}]\,E[\operatorname{sgn}\{(Z_j-Z_k)(Z_j-Z_l)\}].$$

Поскольку

$$E[\operatorname{sgn}\{(Y_j-Y_k)(Y_j-Y_l)\}]=\operatorname{pr}(Y_j>Y_k, Y_j>Y_l)+\operatorname{pr}\{Y_j<Y_k, Y_j<Y_l\}-\operatorname{pr}(Y_k>Y_j>Y_l)-\operatorname{pr}(Y_l>Y_j>Y_k)=\frac{2}{3}-\frac{2}{6}=\frac{1}{3},$$

то $\operatorname{cov}(I_{jk}, I_{jl})=\frac{1}{9}$.

И, наконец,

$$\operatorname{var}(T)=\frac{1}{n^2(n-1)^2}\operatorname{var}\left(\sum_{j>k} I_{jk}\right)=\frac{1}{n^2(n-1)^2}\left\{\frac{1}{2}n(n-1)+\right.$$
$$\left.+\frac{1}{9}n(n-1)(n-2)\right\}=(2n+5)/\{18n(n-1)\}.$$

В случае общего двумерного нормального распределения $\operatorname{corr}(Y-Y', Z-Z')=\operatorname{corr}(Y, Z)$[1]. Таким образом, без потери общности можно считать, что $Y-Y'=U$ и $Z-Z'=V$ имеют двумерное нормальное распределение с нулевыми средними, единичными дисперсиями и коэффициентом корреляции ρ. Искомая вероятность равна, таким образом, $2\operatorname{pr}(U>0, V>0)$. Далее, вычисления, связанные с двумерными нормальным распределением, часто более удобно проводить, обращаясь к условным распределениям. С этой целью положим $V=\rho U+\varkappa W$, где $\varkappa=(1-\rho^2)^{1/2}$, а U и W независимы и нормально $N(0, 1)$ распределены. Искомая вероятность равна, таким образом,

$$\xi=2\int_0^\infty \Phi(-\rho u/\varkappa)\,\phi(u)\,du,$$

где первый сомножитель под знаком интеграла — это $\operatorname{pr}(V>0\mid U=u)$. Интеграл легко вычисляется. Проще всего положить

[1] Здесь и далее $\operatorname{corr}(Y, Z)=\frac{\operatorname{cov}(Y, Z)}{\{\operatorname{var}(Y)\operatorname{var}(Z)\}^{1/2}}$. — *Прим. перев.*

$\rho/\varkappa = \gamma$ и заметить, что

$$\frac{d\xi}{d\gamma} = -2\int_0^\infty \phi(-\gamma u)\,\phi(u)\,du = -\{2\pi(1+\gamma^2)\}^{-1/2}.$$

Интегрирование по γ и использование начального значения $\xi = \frac{1}{2}$ при $\gamma = 0$ приводит к желаемому результату, т. е. получаем так называемую формулу Шеппарда $\xi = \frac{1}{2} + \pi^{-1}\arcsin\rho$.

[Теоретическая статистика,
§ 6.3; Кендалл, 1962]

6.5. Пусть $Y_1, \ldots, Y_{n_1}$ — н.о.р. случайные величины с плотностью $h(y)$, а $Y_{n_1+1}, \ldots, Y_n$ — н.о.р. случайные величины с плотностью $\tau^{-1}h\{(y-\mu)/\tau\}$. Обе совокупности Y независимы. Постройте теорию ранговых критериев проверки нулевой гипотезы H_0: $\tau = 1$, $\mu = 0$ против альтернатив: (a) H_A: $\tau > 1$, $\mu > 0$ и (b) H'_A: $\tau > 1$, $\mu < 0$.

Решение

Первый шаг связан с вычислением плотности вектора рангов R. Она имеет вид

$$f_R(r;\mu,\tau) = \int\limits_{\{y_j = y_{(r_j)}\}} \prod_{k=1}^{n_1} h(y_k) \prod_{l=n_1+1}^{n} \tau^{-1}h\{(y_l-\mu)/\tau\}\,dy.$$

Заметив, что при $\mu = 0$, $\tau = 1$ совместная плотность величин $Y_{(r_j)}$ имеет вид $n!\prod_{j=1}^{n} h(y_{(r_j)})$, получим

$$f_R(r;\mu,\tau) = \frac{1}{\tau^{n_2}n!}E\left[\prod_{j=n_1+1}^{n}\frac{h\{(Y_{(r_j)}-\mu)/\tau\}}{h(Y_{(r_j)})};H_0\right],$$

где $n_2 = n - n_1$, а ожидание берется по полученной выше плотности с $\mu = 0$, $\tau = 1$.

Чтобы построить локальный критерий проверки гипотезы H_0, вычислим

$$[\partial\log f_R(r;\mu,\tau)/\partial\mu]_{\mu=0,\,\tau=1} \text{ и } [\partial\log f_R(r;\mu,\tau)/\partial\tau]_{\mu=0,\,\tau=1}.$$

Отсюда получим статистики критерия

$$S=s(R)=\sum_{j=n_1+1}^{n} v_n(R_j),\ T=t(R)=\sum_{j=n_1+1}^{n} w_n(R_j),$$

где веса $v_n(\cdot)$ и $w_n(\cdot)$ определяются плотностью $h(\cdot)$.

Из статистик критерия S и T необходимо составить одну статистику критерия. Против альтернативы совершенно общего типа в качестве статистики критерия естественно взять квадратичную форму от T и S с коэффициентами, определяемыми матрицей, обратной к ковариационной матрице. В то же время для конкретных альтернатив разумно применить соответствующие односторонние критерии, т. е. взять наиболее значимую из двух компонент. Чтобы вычислить окончательный уровень значимости, необходимо иметь совместное распределение статистик S и T при нулевой гипотезе. Точные значения средних равны $n_1\bar{v}$ и $n_2\bar{w}$, а

$$\operatorname{var}(S;\ H_0)=\frac{n_1 n_2}{(n-1)}\sum\{v_n(j)-\bar{v}\}^2.$$

Для $\operatorname{cov}(S, T)$ и $\operatorname{var}(T)$ можно получить аналогичные формулы. Часто целесообразно использовать двумерную нормальную аппроксимацию. Например, если альтернативной гипотезой является H_A, а в качестве статистики критерия взят $\max\{S/\sqrt{\operatorname{var}(S)}, T/\sqrt{\operatorname{var}(T)}\}$, то вероятность превышения этого наблюденного значения при H_0 можно приближенно вычислить непосредственно на основе таблиц двумерного нормального распределения.

[Теоретическая статистика, § 6.3;
Леман, § 6.8]

6.6. Исследуйте распределение при нулевой гипотезе оптимального двувыборочного рангового критерия для показательных распределений.

Предположите, что при нулевой гипотезе н.о.р. случайные величины $Y_1, \ldots, Y_n$ экспоненциально распределены, тогда как при альтернативной гипотезе Y_j имеет среднее $1/(\gamma+\beta z_j)$, где $z_1, \ldots, z_n$ известны. Постройте оптимальные параметрические и ранговые критерии. Для последнего случая предложите аппроксимацию для распределения статистики критерия при нулевой гипотезе.

Решение

Оптимальный двувыборочный ранговый критерий для проверки показательности исходных распределений по форме аналогичен критерию Вилкоксона, рассмотренному в общих чертах в задаче 6.2. Различие состоит в том, что в рассматриваемом здесь критерии i-му по величине наблюдению приписывается вес $e_{ni} = 1/n + 1/(n-1) + \ldots + 1/(n-i+1)$, являющийся ожидаемым значением i-го по величине наблюдения в выборке объема n из единичного показательного распределения. Вместо того чтобы быть средним рангом, статистика критерия будет здесь средним вкладом элементов первой выборки.

Можно получить точное распределение статистики при нулевой гипотезе. На этот раз необходимо рассматривать конечную популяцию $\{e_{n_1}, \ldots, e_{nn}\}$, которая, как можно показать, имеет единичное среднее. Аппроксимации можно получить следующими методами:

(i) используя нормальную аппроксимацию для „выборочного среднего" $\bar{e}_1$;

(ii) „улучшая" нормальную аппроксимацию, находя асимметрию и эксцесс на основе результатов задачи 6.2;

(iii) используя аналогию с теорией показательных распределений, аппроксимировать распределение статистики $\bar{e}_1/(1-\bar{e}_1)$ посредством F распределения с $(2n_1, 2n_2)$ степенями свободы, т. е. аппроксимировать распределение статистики $\bar{e}_1$ некоторым бета распределением;

(iv) „улучшить" аппроксимацию пункта (iii), используя степени свободы $(2n_1 l, 2n_2 l)$ и выбирая l так, чтобы в точности подогнать среднее и дисперсию статистики $\bar{e}_1$. Как показывают подробные вычисления (см. Кокс, 1964 b)

$$l \sim 1 + (\gamma - 2 + \log n)/n,$$

где γ — постоянная Эйлера. Все методы дают достаточно хорошую аппроксимацию даже для довольно малых выборок. В то же время метод (iv) является наиболее точным.

[Теоретическая статистика,
§ 6.3, 5,2; Леман, § 6.8]

6.7.[1]) В простом рандомизированном блочном плане, используемом для сравнения m способов воздействия, испытываемые единицы расположены в b блоков по m единиц в каждом. Каждый из способов воздействия осуществляется по одному разу в каждом блоке, а в остальном распределение способов воздействия проводится случайно. Пусть $x_{jk,p}$ — отклик, который был бы получен при применении p-го воздействия в k-й испытываемой единице j-го блока. Предполагается, что $x_{jk,p}$ не зависит от распределения способов воздействий по другим испытываемым единицам. В соответствии с предполо-

[1]) Улучшенный вариант.

жением выполняется свойство аддитивности типа „воздействие—испытываемая единица“, т. е. разность $x_{jk,p}-x_{jk,q}$ принимает одно и то же значение для всех испытываемых единиц. Отсюда следует, например, что $x_{jk,p}=\tau_p+x_{jk}$, где x_{jk} можно было бы рассматривать с точностью до постоянной как наблюдение, которое можно было бы получить при однородном испытании, т. е. в отсутствие каких-либо воздействий, а τ_p — параметр способа воздействия. Пусть x_{jp} — наблюдение, полученное при p-м способе воздействия в j-м блоке. Постройте линейную модель, основанную только на рандомизации. Используйте некоторые характеристики стандартного дисперсионного анализа.

Решение

При использовании подхода на основе теории обычной нормальной линейной модели для отклика необходимо иметь представление

$$Y_{jp}=\mu+\tau_p+\beta_j+\varepsilon'_{jp}, \tag{1}$$

где ε'_{jp} — н.о.р. случайные величины, нормально $N(0, \sigma^2)$ распределенные. Такая модель отсылает к тому, что произошло бы в некотором множестве воображаемых повторов опыта, как правило, плохо определенных. При этом делается предположение о частной форме возникающей случайной изменчивости. В противоположность этому в модели рандомизации имеют дело лишь с конечным множеством $x_{jk,p}$. Вероятностные соображения связаны только с рандомизацией, используемой при распределении способов воздействий.

Действительно, Y_{jp} равен τ_p плюс одна из величин $x_{j1}, \ldots, x_{jm}$, выбранная случайно. Это можно записать так:

$$Y_{jp}=\tau_p+\sum_{k=1}^{m}\Delta_{jk,p}\,x_{jk},$$

где $\Delta_{jk,p}=1$, если p-й способ воздействия применялся к k-й экспериментальной единице j-го блока, и $\Delta_{jk,p}=0$ в противном случае. Чтобы получить „ошибку“ с нулевым средним, перепишем это так:

$$Y_{jp}=\tau_p+\bar{x}_{j.}+\varepsilon_{jp}, \tag{2}$$

где $\varepsilon_{jp}=\sum\Delta_{jk,p}(x_{jk}-\bar{x}_{j.})$. Формальная аналогия с моделью (1) очевидна. Индикаторные случайные величины $\Delta_{jk,p}$ характеризуются весьма просто. Например,

$$\operatorname{pr}(\Delta_{jk,p}=1)=1/m,\ \operatorname{pr}(\Delta_{jk,p}=1,\ \Delta_{jl,q}=1)=1/\{m(m-1)\},$$
$$(k\neq l,\ p\neq q). \tag{3}$$

Это означает, что, по крайней мере с первого взгляда, ε_{jp} в модели (2) по своей структуре сильно отличаются от ε'_{jp} в модели (1).

Перечислим три важных результата, вытекающих из теории нормальной модели (1):

(a) контраст $\sum l_p\tau_p$ (при $\sum l_p=0$) оценивается как $\sum l_p\overline{Y}_{\cdot p}$, дисперсия же оценки контраста оценивается как $\sum l_p^2\,\mathrm{MS}_e/b$, где

$$\mathrm{MS}_e=\sum(Y_{jp}-\overline{Y}_{j\cdot}-\overline{Y}_{\cdot p}+\overline{\overline{Y}}_{\cdot\cdot})^2/\{(m-1)(b-1)\};$$

(b) если $\mathrm{MS}_t=b\sum(\overline{Y}_{\cdot p}-\overline{Y}_{\cdot\cdot})^2/(m-1)$, тогда $E(\mathrm{MS}_t)\geqslant E(\mathrm{MS}_e)$, причем равенство достигается в том и только том случае, если $\tau_1=\ldots=\tau_m$;

(c) для проверки гипотез о величине характеристик, рассмотренных в пунктах (a) и (b), применяются t и F критерии.

Чтобы исследовать рандомизированные аналоги процедур, рассмотренных в пунктах (a)—(c), проще всего выразить соответствующие величины через $\Delta_{jk,\,p}$, а затем использовать свойства распределения $\Delta_{jk,\,p}$, т. е. равенства (3). При построении рандомизированных процедур можно избежать довольно длинных выкладок, если применять соображения, аналогичные использованным при решении задачи 6.2. Пусть $E_R(\cdot)$ — ожидание относительно распределения величин $\Delta_{jk,\,p}$.

В частности, линейная форма $\sum l_p\overline{Y}_{\cdot p}$ имеет вид $\sum l_p\tau_p+$ линейная функция от x. Из соображений симметрии следует, что

$$E_R(\Sigma l_p\overline{Y}_{\cdot p})=\Sigma l_p\tau_p+a\sum_{j,\,k}x_{jk}.$$

Выбирая популяцию специальным образом, с $x_{jk}=1$, получаем, что $a=0$. Аналогично, выбирая для простоты контраст $\overline{Y}_{\cdot 1}-\overline{Y}_{\cdot 2}$, найдем его дисперсию по распределению, порождаемому рандомизацией. Она представляется в виде квадратичной формы от x. Соображения, связанные со свойством симметричности рандомизации, показывают, что

$$\mathrm{var}_R(\overline{Y}_{\cdot 1}-\overline{Y}_{\cdot 2})=c'\Sigma x_{jk}^2+d'\sum_{k\neq l}x_{jk}x_{jl}+e'\sum_{j\neq l}x_{jk}x_{lm}=$$

$$=c''\overline{x}^2_{\cdot\cdot}+d''\Sigma(\overline{x}_{j\cdot}-\overline{x}_{\cdot\cdot})^2+e''\Sigma(x_{jk}-\overline{x}_{j\cdot})^2,\quad (4)$$

где c'', d'' и e'' — постоянные, не зависящие от x. Если внутри каждого блока x_{jk} постоянны, то дисперсия равна нулю и $c''=d''=0$.

Аналогичные соображения показывают, что

$$E_R(\mathrm{MS}_e)=e'''\Sigma(x_{jk}-\overline{x}_{j\cdot})^2.\quad (5)$$

Постоянные e'' и e''' можно вычислить, обращаясь к рассмотрению специальных популяций. С другой стороны, опираясь на доводы исключительно математического характера, представим себе, что x_{jk} являются некоррелированными случайными величинами с нулевыми средними и дисперсией τ^2. Тогда, обозначая $\mathscr{E}$ ожидание по распределению этих нормальных величин, получаем

$$\mathscr{E}\{\mathrm{var}_R(\overline{Y}_{.1}-\overline{Y}_{.2})\}=2\tau^2 b,$$
$$\mathscr{E}\{E_R(\mathrm{MS}_e)\}=\tau^2,$$
$$\mathscr{E}\{\Sigma(x_{jk}-\overline{x}_{j.})^2\}=b(m-1)\tau^2.$$

Отсюда вытекает, что

$$\mathrm{var}_R(\overline{Y}_{.1}-Y_{.2})=\frac{2\Sigma(x_{jk}-\overline{x}_{j.})^2}{b^2(m-1)}=\frac{2}{b}E_R(\mathrm{MS}_l).$$

Аналогично,

$$E_R(\mathrm{MS}_t)=\frac{b\Sigma(\tau_p-\overline{\tau})^2}{m-1}+E_R(\mathrm{MS}_e).$$

Полученные формулы являются рандомизированными аналогами характеристик линейной модели, рассмотренных в пунктах (a) и (b). Наиболее важный общий аспект рассмотренных доводов состоит в том, что аналогичные результаты для других планов вытекают из предположения о выполнении свойства аддитивности типа „воздействие — экспериментальная единица" в сочетании с соответствующей рандомизацией. Таким образом, при обращении к латинским квадратам или к неполным сбалансированным блочным планам каких-либо существенно отличных вероятностных предположений по сравнению с рандомизированными блоками или на самом деле с полностью рандомизированными планами не делается.

Точные распределения статистик критериев значимости, основанных на рандомизации, можно получить методом перебора. Возможны приближения на основе вычисления моментов статистик. При нулевой гипотезе об эквивалентности всех способов воздействия можно вычислить точно все наблюдения, которые могли бы быть получены при различных распределениях способов воздействий. Далее, $\mathrm{SS}_t+\mathrm{SS}_e=$ $=\mathrm{const}$. Отсюда вытекает, что вместо $\mathrm{MS}_t/\mathrm{MS}_e$ можно рассмотреть SS_t или MS_t. Обобщая рассмотренные выше соображения, можно вычислить дисперсии и более высокие моменты этих величин. Если маргинальное распределение величин $Y_{jp}-\overline{Y}_j$ не слишком отличается от ожидаемого распределения для модели (1), то рандомизированное распределение статистики критерия окажется близким распределению, отвечающему модели нормальности ошибок. На самом деле это очевидно и из общих соображений. Для „бесконечных" моделей, подобных (1), распределение любой статистики критерия является усреднением отдельных распределений рандомизации. Таким образом, теория „бесконечной" модели и теория рандомизации согласуются в смысле некоторого усреднения.

[Теоретическая статистика § 6.4; Рао, § 7е.4; Силвей § 9.6; Кемпторн, 1952, 1966; Гранди и Хили, 1950]

6.8. Пусть нулевая гипотеза состоит в том, что наблюдаются n н.о.р. случайных величин с заданной непрерывной функцией распределения $F(y)$. Рассмотрим эмпирическую функцию распределения $\tilde{F}_n(y)$, определенную как долю из n наблюдений, меньших или равных y. Колмогоровские статистики расстояния определяются следующим образом:

$$D_n^+ = \sup_y \{\tilde{F}_n(y) - F(y)\},$$

$$D_n^- = \sup_y \{F(y) - \tilde{F}_n(y)\},$$

$$D_n = \max(D_n^+, D_n^-).$$

Докажите, что распределение статистики расстояния Колмогорова при нулевой гипотезе не зависит от произвольных монотонных преобразований данных. Получите отсюда, что без потери общности в качестве проверяемой нулевой гипотезы можно брать равномерное распределение на (0,1).

Докажите идентичность свойств моментов первых двух порядков для $\sqrt{n}\,[\tilde{F}_n(y) - y]$ и броуновского движения с закрепленными концами (иногда его называют броуновским мостиком), т. е. броуновского движения, исходящего из точки (0, 0), на которое наложено условие, что оно проходит через точку (1, 0). Используя этот факт, попытайтесь интерпретировать распределение статистики критерия при нулевой гипотезе с помощью броуновского движения с закрепленными концами и поглощающими барьерами. Покажите, что при $n \to \infty$

$$\lim (D_n^+ \sqrt{n} < x) = 1 - e^{-2x^2}$$

и

$$\lim (D_n \sqrt{n} < x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2}.$$

Решение

Искомая инвариантность видна, если нарисовать на одном графике $\bar{F}_n(y)$ и теоретическую функцию распределения $F(y)$. При произвольных монотонных преобразованиях данных ось абцисс деформируется соответствующим образом, а ось ординат не меняется. Таким образом, совокупность значений разности $\bar{F}_n(y) - F(y)$ не меняется при преобразованиях данных. Положив $F(y) = y$, исходное распределение можно трансформировать в равномерное.

Далее, $\bar{F}_n(y)$ — доля „успешных" среди n независимых испытаний с вероятностью успеха, равной y. Привлекая сюда свойства биномиального распределения, получим

$$E\{\bar{F}_n(y)\} = y, \ \operatorname{var}\{\bar{F}_n(y)\} = y(1-y)/n.$$

Пусть $y_1 < y_2$. Можно считать, что наблюдения принадлежат одному из трех интервалов: $[0, y_1)$, $[y_1, y_2)$, $[y_2, 1]$. К этой схеме применимо полиномиальное распределение с тремя исходами. Отсюда вытекает, что $\operatorname{cov}\{\bar{F}_n(y_1), \bar{F}_n(y_2)\} = y_1(1-y_2)/n$.

Броуновское движение — это случайный процесс, который эффективно представим в виде накопленных сумм независимых нормально распределенных случайных величин. Вторые моменты броуновского движения с закрепленными концами можно вычислить на основе простых соображений. Введем нормальные случайные величины Z_1, Z_2, Z_3 с дисперсиями y_1, $y_2 - y_1$ и $1 - y_2$. Пусть до введения граничных условий значения $\bar{F}_n(y_1)$ и $\bar{F}_n(y_2)$ представляются величинами Z_1 и $Z_1 + Z_2$, а $\bar{F}_n(1) = Z_1 + Z_2 + Z_3$. Тогда требуемая ковариационная матрица совпадает с ковариационной матрицей Z_1 и $Z_1 + Z_2$ при условии, что $Z_1 + Z_2 + Z_3 = 1$. На основе известных свойств трехмерного нормального распределения последняя матрица легко вычисляется.

Из эвристических соображений, столь ценимых в прикладной математике, необходимо считать „очевидным" тот факт, что биномиально распределенный процесс $\{\bar{F}_n(y)\}$ аппроксимируется нормально распределенным процессом с той же самой ковариационной структурой. Строгое доказательство, которое имеет значительный математический интерес, лучше всего вести на основе использования идей слабой сходимости (см. Биллингсли, 1977). Задача нахождения распределения статистик D_n и D_n^+ приближенно будет соответствовать тогда задаче с двумя поглощающими экранами. Соответствующие методы решения таких задач приводятся в книгах по случайным процессам.

[Теоретическая статистика, § 6.5, 3.2; Рао, § 6; Силвей, § 9.1, Бартлетт, 1966, § 4.1]

6.9. Проверка независимости случайных величин Y и Z эквивалентна проверке того, представляется ли их совместная функция распределения в виде $F_{Y,Z}(y, z) = G_Y(y) K_Z(z)$, где $G_Y(\cdot)$ и $K_Z(\cdot)$ — маргинальные функции распределения. Постройте свободный от распределения перестановочный критерий проверки монтонной зависимости, основанный на мере зависимости

$$\Delta = \int\int \{F_{Y,Z}(y, z) - G_Y(y) K_Z(z)\} \, dG_Y(y) \, dK_Z(z).$$

Получите выборочное распределение статистики критерия при нулевой гипотезе о независимости Y и Z.

Решение

Один из простых методов оценки меры зависимости Δ основан на использовании эмпирических функций распределения. Рассмотрим метод U-статистик как альтернативный метод построения несмещенных оценок с минимальной дисперсией для Δ. Чтобы построить оценку, заметим, что

$$\Delta = \mathrm{pr}(Y < Y', \ Z < Z'') - \frac{1}{4},$$

где (Y, Z), Y' и Z' независимы. Отсюда следует, что несмещенная оценка величины Δ строится по трем независимым парам (Y_1, Z_1), (Y_2, Z_2) и (Y_3, Z_3) в следующем виде:

$$t\{(Y_1, Z_1), (Y_2, Z_2), (Y_3, Z_3)\} =$$
$$= \frac{1}{2}\{\mathrm{hv}(Y_2 - Y_1)\,\mathrm{hv}(Z_3 - Z_1) + \mathrm{hv}(Y_3 - Y_1)\,\mathrm{hv}(Z_2 - Z_1)\} - \frac{1}{4}.$$

Соответствующей симметрической оценкой будет

$$k\{(Y_1, Z_1), (Y_2, Z_2), (Y_3, Z_3)\} =$$
$$= \frac{1}{3}[t\{(Y_1, Z_1), (Y_2, Z_2), (Y_3, Z_3)\} + t\{(Y_2, Z_2), (Y_3, Z_3), (Y_1, Z_1)\} +$$
$$+ t\{(Y_3, Z_3), (Y_1, Z_1), (Y_2, Z_2)\}.$$

Для заданных n пар $(Y_1, Z_1), \ldots, (Y_n, Z_n)$, $n \geqslant 3$, полной минимальной достаточной статистикой для F будет совокупность упорядоченных значений Y_j и сопоставленных им Z_j. Следовательно, из теоремы Рао — Блекуэлла, приведенной в гл. 8, вытекает, что для Δ несмещенная оценка с минимальной дисперсией получается из предшествующей оценки усреднением по минимальной достаточной статистике. Оценка принимает вид

$$T = \frac{1}{\binom{n}{3}} \sum k\{(Y_{j_1}, Z_{j_1}), (Y_{j_2}, Z_{j_2}), (Y_{j_3}, Z_{j_3})\},$$

где суммирование ведется по всем $1 \leqslant j_1 \leqslant j_2 \leqslant j_3 \leqslant n$.

При нулевой гипотезе независимости, в принципе, можно найти точное выборочное распределение статистики T. Воспользуемся тем фактом, что при фиксированных порядковых статистиках $y_{(1)} \leqslant \ldots \leqslant y_{(n)}$ и $z_{(1)} \leqslant \ldots \leqslant z_{(n)}$ все перестановки последовательности $(y_{(1)}, \ldots, y_{(n)})$ являются равновероятными значениями величин $(Y_1, \ldots, Y_n)$, а все перестановки последовательности $(z_{(1)}, \ldots, z_{(n)})$ являются равновероятными значениями величин $(Z_1, \ldots, Z_n)$ при заданном значении $(Y_1, \ldots, Y_n) = (y_{(1)}, \ldots, y_{(n)})$.

В приложениях нормальная аппроксимация распределения статистики T при нулевой гипотезе будет, по-видимому, хорошо действовать для $n > 10$. Среднее статистики T равно нулю, а приближенное выражение для дисперсии в больших выборках равно (см. Хёфдинг, 1948)

$$9n^{-1} \operatorname{var}(E[k\{(Y_1, Z_1), (Y_2, Z_2), (Y_3, Z_3)\} \mid (Y_1, Z_1)]).$$

Как показывают простые вычисления, последнее выражение равно $(16n)^{-1}$.

[Теоретическая статистика, §6.5]

7. Интервальное оценивание

Необходимые сведения

Центральная проблема статистических выводов во многом состоит в том, чтобы представить в суммарном виде все, что можно узнать из данных о значении неизвестного параметра. При этом все точки параметрического пространства априорно рассматриваются как равноправные. После анализа данных на разном вероятностном уровне выделяется та совокупность значений параметров, которые согласуются с данными. При подходе на основе выборочной теории такие совокупности значений параметров задаются доверительными границами, интервалами и областями.

Предположим сначала, что имеется одномерный неизвестный параметр θ, однозначно определяющий распределение случайной величины Y. Пусть для $\alpha > 0$ статистика $T^{\alpha} = t^{\alpha}(Y)$ такова, что для всех $\theta \in \Omega$

$$\operatorname{pr}(T^{\alpha} \geqslant \theta;\ \theta) = 1 - \alpha, \tag{1}$$

а при $\alpha_1 > \alpha_2$ $T^{\alpha_1} \leqslant T^{\alpha_2}$. В этом случае назовем T^{α} верхней $(1-\alpha)$-доверительной границей для θ. Если $g(\ )$ строго возрастает, то $g(T^{\alpha})$ будет верхней $(1-\alpha)$-доверительной границей для $g(\theta)$. Нижняя доверительная граница определяется аналогично. Если знак равенства в соотношении (1) заменяется на $\geqslant$, то доверительные границы называют консервативными. Соотношение (1) дает теоретическую интерпретацию доверительным границам, которая, в принципе, проверяема экспериментально.

В приложениях не останавливаются на выборе одного определенного значения α, а явно или неявно используют доверительные границы для некоторого набора значений α.

Оптимальность доверительных границ определяется требованием, чтобы вероятность $\operatorname{pr}(T^{\alpha} \geqslant \theta';\ \theta)$ была малой для $\theta' > \theta$ и, если возможно, минимальной для всех пар (θ, θ'), т. е. доверительная граница должна быть близкой к θ. Это приводит к общей процедуре построения хороших или наилучших доверительных границ. Чтобы получить „хорошую" верхнюю $(1-\alpha)$-доверительную границу, берут все те значения параметра, которые не отвергаются „на уровне α" „хорошим" критерием значимости против левых альтернатив.

Дискретным распределениям отвечает лишь некоторое дискретное множество значений уровней α. При этом, как правило, используют консервативные доверительные границы.

Приведенное выше определение обобщается на доверительные интервалы и области. Таким образом, $[T_{\cdot}, T^{\cdot}]$ определяет $(1-\alpha)$-доверительный интервал, если для всех θ

$$\mathrm{pr}\,(T_{\cdot} \leqslant \theta \leqslant T^{\cdot};\ \theta) = 1-\alpha$$

и в дополнение к этому выполняется свойство вложенности. В случае непрерывных распределений сочетание верхней и нижней доверительных границ вида $[T_{\alpha_1}, T^{\alpha_2}]$ с $\alpha_1+\alpha_2=\alpha$ дает такой интервал. Для большинства случаев разумно условиться положить $\alpha_1=\alpha_2=\frac{1}{2}\alpha$, поскольку при этом одновременно задаются нижние и верхние доверительные границы известных уровней

В ситуации отсутствия естественного свойства монотонности доверительные области по своей форме могут отличаться от интервалов. Вложенная система областей $R_\alpha(Y)$ является $(1-\alpha)$-доверительной областью, если для θ

$$\mathrm{pr}\,\{\theta \in R_\alpha(Y);\ \theta\} = 1-\alpha. \tag{2}$$

При построении доверительных областей полезно договориться образовывать их так, если это возможно, чтобы значение правдоподобий параметров, лежащих внутри R_α, было выше, чем для параметров, лежащих вне R_α. Такие доверительные области называют основанными на правдоподобии. Общая процедура построения хороших интервалов и областей связана с соответствующими двусторонними критериями. Интервал или область составляется из всех значений параметра, которые не отвергаются на рассматриваемом уровне.

Когда параметр θ многомерен, доверительные области для θ определяются на основе простого обобщения соотношения (2). Более часто требуется найти доверительные границы, интервалы или области для компоненты ψ параметра θ, считая λ — оставшуюся компоненту параметра θ — мешающим параметром. Идеи подобия и инвариантности, развитые в гл. 5 и примененные там непосредственно для построения критериев, используются и при построении доверительных областей. Не исключена возможность, что конкретная доверительная область совпадает, в частности, со всем параметрическим пространством. Это будет интерпретироваться так, что на данном уровне все допустимые значения параметра согласуются с данными.

Рассмотрим последнее, близко связанное с доверительными областями понятие. Речь идет об областях предсказания уровня $(1-\alpha)$ для некоторой ненаблюдаемой случайной величины $Y^{\dagger}$, распределение которой зависит от того же самого параметра θ, что и распределение наблюдаемой величины Y. Эту задачу также можно формально обратить в задачу теории проверки гипотез. Для этого предполагают, что Y и $Y^\dagger$ имеют распределения, соответствующие значениям параметра θ и $\theta^\dagger$, и после этого находят w_α — наилучшую критическую область размера α для проверки сложной нулевой гипотезы $\theta=\theta^\dagger$. Для заданного y область предсказания $P_\alpha(y)$ определим как совокупность таких значений $y^\dagger$, что пара $(y, y^\dagger)$ не принадлежит ω_α. Таким образом, y и $y^\dagger$ на уровне α согласуются с общим значением параметра. Тогда область $P_\alpha(y)$ для всех θ обладает следующим свойством:

$$\mathrm{pr}\{Y^\dagger \in P_\alpha(Y);\ \theta\} = 1-\alpha,$$

являющимся разумным требованием, предъявляемым к областям предсказания.

Задачи

7.1. Случайная величина Y имеет биномиальное распределение с параметром 1/2 и с неизвестным числом испытаний θ. Постройте для θ доверительные границы.

Решение

Чтобы построить верхнюю доверительную границу, рассмотрим нулевую гипотезу $\theta=\theta_0$ при альтернативе $\theta<\theta_0$. Консервативная критическая область размера α будет иметь вид $y<y^\alpha(\theta_0)$, где

$$\sum_{r=0}^{y^\alpha(\theta_0)} \binom{\theta_0}{r} 2^{-\theta_0} \leqslant \alpha.$$

Пусть для фиксированного наблюдения y величина $\bar\theta^\alpha$ — наибольшее значение параметра θ_0, которое еще не отвергается на уровне α, т. е.

$$\sum_{r=0}^{y} \binom{\theta_0}{r} 2^{-\theta_0} \geqslant \alpha.$$

Тогда $\bar\theta^\alpha$ — консервативная верхняя доверительная граница с консервативным коэффициентом доверия $1-\alpha$. Соответствующие нижние доверительные границы $\underline{\theta}^\alpha$ определяются на основе критических областей вида $y>y_\alpha(\theta_0)$.

Верхнюю доверительную границу можно получить приближенно на основе нормальной аппроксимации левой части неравенства (1). С учетом поправки на непрерывность получим

$$\Phi\left(\frac{y+\frac{1}{2}-\frac{1}{2}\tilde{\theta}^{\alpha}}{\frac{1}{2}\sqrt{\tilde{\theta}^{\alpha}}}\right)=\Phi(-k_{\alpha}^{*}).$$

Это равенство ведет к квадратному уравнению относительно $\tilde{\theta}^{\alpha}$. Его решением будет

$$2y+1+\frac{1}{2}k_{\alpha}^{*2}\pm k_{\alpha}^{*}\left(2y+1+\frac{1}{4}k_{\alpha}^{*2}\right)^{1/2}.$$

[Теоретическая статистика, § 7.2, 7b.2]

7.2. Пусть $Y_1, \ldots, Y_n$ — н. о. р. случайные величины с п. р. в. $h(y-\theta)$, где $h(\cdot)$ — неизвестная плотность, симметричная относительно начала координат. Объясните, как можно проверить нулевую гипотезу $\theta=0$, используя статистику ΣY_j. Чтобы получить доверительную область для θ, этот критерий примените к $\Sigma(Y_j-\theta_0)$. Покажите, что обращение этого критерия для любого фиксированного уровня значимости приводит к интервалу для θ.

Решение

При нулевой гипотезе $\theta=0$ значения y и $-y$ для y_j имеют одинаковую вероятность. Таким образом, если в качестве статистики критерия взята $T=\Sigma Y_j$, а наблюденные значения $y_1, \ldots, y_n$, то распределения статистики T при нулевой гипотезе можно найти, приписывая равные вероятности 2^{-n} всем 2^n точкам $\{\pm y_1, \ldots, \pm y_n\}$. При более формальном подходе заметим, что $\{|Y_1|, \ldots, |Y_n|\}$ — минимальная достаточная статистика для неизвестной мешающей функции $h(\cdot)$. Отсюда вытекает, что при использовании условных распределений в качестве случайных можно считать только знаки наблюдений Y_j. При этом остается, конечно, значительная свобода в выборе статистики критерия.

Распределение статистики T при нулевой гипотезе можно получить методом перебора. С другой стороны, можно использовать нормальную аппроксимацию. При этом моменты статистики T найти легко, если заметить, что условно T представляются как сумма n независимых бинарных случайных величин. Таким образом,

$$E(T;\ \theta=0)=0,\ \operatorname{var}(T;\ \theta=0)=\Sigma y_j^2.$$

Чтобы получить доверительные интервалы, рассмотрим произвольные нулевые гипотезы $\theta = \theta_0$ и возьмем в качестве статистики критерия $T(\theta_0) = \Sigma (Y_j - \theta_0)$. Условное выборочное пространство в этом случае имеет вид $\{\theta_0 \pm (y_1 - \theta_0), \ldots, \theta_0 \pm \pm (y_n - \theta_0)\}$. Тогда

$$E\{T(\theta_0);\ \theta = \theta_0\} = 0, \ \operatorname{var}\{T(\theta_0);\ \theta = \theta_0\} = \Sigma (y_j - \theta_0)^2.$$

Чтобы построить консервативную верхнюю $(1-\alpha)$-доверительную границу, необходимо проверять гипотезу $\theta = \theta_0$ против альтернатив $\theta < \theta_0$. В этом случае значимыми будут малые значения статистики $T(\theta_0)$. Далее, пусть $T_\alpha^*(\theta_0)$ — консервативная нижняя α-точка распределения статистики $T(\theta_0)$, т. е. такое наибольшее значение t, для которого

$$\operatorname{pr}\{T(\theta_0) \leqslant t;\ \theta = \theta_0\} \leqslant \alpha.$$

Тогда консервативная верхняя доверительная область для θ образуется из всех таких значений θ_0, для которых $T(\theta_0) > > T_\alpha^*(\theta_0)$. Это вытекает из того, что все указанные значения согласуются с гипотезой $\theta = \theta_0$ на консервативном уровне α. Если использовать нормальную аппроксимацию на основе точных значений для среднего и дисперсии статистики $T(\theta_0)$, то доверительная область аппроксимируется полуинтервалом $\theta = \tilde{\theta}^\alpha$, где

$$\tilde{\theta}^\alpha = \bar{y}_. + \frac{k_\alpha^* \{\Sigma (y_j - \bar{y}_.)^2\}^{1/2}}{\{n(n - k_\alpha^{*2})\}^{1/2}}.$$

Более обстоятельное исследование, связанное с вычислением точного распределения статистики $T(\theta)$, показывает, что доверительная область имеет вид $\theta \leqslant \bar{y}_. + c_\alpha(y)$. Соответствующий двусторонний критерий дает возможность построить доверительный интервал, и точка $\bar{y}_.$ будет лежать внутри него. Более подробно об этом см. работу Компторна и Дерфлера (1969).

[Теоретическая статистика, § 3.2;
Рао, § 7b, 7e.4]

7.3. Пусть $Y_1, \ldots, Y_n$ — н. о. р. случайные величины с неизвестной непрерывной функцией распределения $F_Y(y)$. Для $0 < p < 1$ определим ξ_p, p-квантиль упомянутого распределения, как решение уравнения $p = F_Y(\xi_p)$. Поскольку это уравнение не обязательно имеет одно решение, то необходимо принять какое-либо разумное соглашение о единственном выборе ξ_p. Покажите, как проверить нулевую гипотезу $\xi_p = \xi_{p0}$ против: (a) односторонних, (b) двусторонних альтернатив, если использовать лишь число наблюдений, превысивших ξ_{p0}. Покажите, что для больших n и односторонних альтернатив $\xi_p >$

$> \xi_{p0}$ данные попадают в критическую область размера α в том и только том случае, если r-я порядковая статистика превышает ξ_{p0}, где r — целая область от

$$np - k_\alpha^* \{np(1-p)\}^{1/2} + \frac{1}{2}.$$

Исходя из этого, постройте доверительные границы для ξ_p.

Решение

Обозначим R число наблюдений, меньших, чем ξ_{p0}. Если нулевая гипотеза $\xi_p = \xi_{p0}$ верна, тогда R имеет биномиальное распределение, отвечающее n испытаниям и вероятности успеха p. Для альтернатив $\xi_p > \xi_{p0}$ критическая область образуется из малых значений величины R. Используя нормальную аппроксимацию с поправкой на непрерывность, получим, что критическая область размера α образуется из совокупности значений r, для которых

$$\frac{r - np + \frac{1}{2}}{\{np(1-p)\}^{1/2}} < -k_\alpha^*. \tag{1}$$

Далее, событие $R < r'$ эквивалентно тому, что r'-я порядковая статистика $Y_{(r')}$ примет значение, большее ξ_{p0}. Таким образом, данные оказываются в критической области размера α в том и только том случае, если порядковая статистика превысит Y_s, где s — наименьшее значение r, для которого нарушается неравенство (1). Отсюда

$$s = \left[-k_\alpha^* \{np(1-p)\}^{1/2} + np - \frac{1}{2}\right] + 1,$$

где $[a]$ обозначает целую часть числа a. Все те значения, которые не отвергаются для рассматриваемых данных, образуют приближенную одностороннюю доверительную область. Она представляется, таким образом, в виде интервала $[Y_{(s)}, \infty]$. Точные консервативные доверительные интервалы можно получить при использовании биномиального распределения величины R. Верхнюю границу можно получить аналогично.

При таком непараметрическом подходе доверительные границы зависят только от порядковых статистик, близких к $Y_{[(np)]}$. Число таких статистик составляет малую долю данных. Когда рассматривается параметрическое семейство распределений, то подобной зависимости от локализованной части данных, как правило, не наблюдается. Особо критическая ситуация складывается для весьма малых или весьма больших значений p. В зависимости от того, верна или не верна параметрическая модель, описывающая поведение распределения на хвостах, параметрический подход либо будет в значительной степени более точным, либо, по всей вероятности, будет приводить к грубым ошибкам.

[Теоретическая статистика, § 7.2; Рао, § 7b.2, 6f.2]

7.4. Пусть $Y_{(1)}, \ldots, Y_{(n)}$ — порядковые статистики из равномерного распределения на $(\theta, \theta+1)$. Используя безусловный равномерно наиболее мощный односторонний критерий, покажите, что соответствующая оптимальная нижняя доверительная граница имеет вид $T_\alpha = \max(Y_{(1)} - c, Y_{(n)} - 1)$, где $(1-c)^n = 1-\alpha$. Обсудите критически свойства этой границы с учетом того, что

$$\operatorname{pr}(T_\alpha \leqslant \theta \mid Y_{(n)} - Y_{(1)} \geqslant 1-c) = 1.$$

Предложите более разумную нижнюю доверительную границу для θ.

Решение

Пара экстремальных порядковых статистик $(Y_{(1)}, Y_{(n)})$ образует минимальную достаточную статистику, а статистика $A = Y_{(n)} - Y_{(1)}$ является подчиненной, что вытекает из инвариантности статистики A относительно сдвиговых преобразований данных.

Чтобы построить нижнюю доверительную границу, рассмотрим нулевую гипотезу $\theta = \theta_0$ против альтернатив $\theta > \theta_0$. Далее для любого фиксированного $\theta_A > \theta_0$ возможными значениями отношения правдоподобия $\operatorname{lr}_{A0}(y)$ являются 0, 1 и ∞ в зависимости от того, какое из следующих событий осуществилось:

$$\theta_0 \leqslant y_{(1)} < \theta_A, \quad \theta_A \leqslant y_1 \leqslant y_{(n)} \leqslant \theta_0 + 1 \quad \text{и} \quad \theta_0 + 1 < y_{(n)} < \theta_A + 1.$$

При $\theta = \theta_0$ последнее событие произойти не может. Отсюда вытекает, что допустимыми уровнями для критических областей отношения правдоподобия будут

$$1, \ (\theta_0 + 1 - \theta_A)^n, \ 0,$$

причем средний уровень определяется на основе конкретно выбранной альтернативы. Чтобы получить критическую область размера α, необходимо использовать рандомизацию. Наилучший способ проведения рандомизации связан с выбором критической области размера α, которая дает наибольший вклад в мощность при $\theta > \theta_0$. Это требование сразу приводит к критической области равномерно наиболее мощного критерия. Она имеет вид

$$w_\alpha = \{y : y_{(1)} > \theta_0 + c \text{ или } y_{(n)} > \theta_0 + 1\}, \qquad (1)$$

причем непосредственное вычисление вероятности $\operatorname{pr}(Y \in w_\alpha;\, \theta_0)$ показывает, что $\alpha = (1-c)^n$.

Все значения θ_0, для которых данные y не принадлежат критической области (1), образуют искомую доверительную область. Таким образом, она состоит из совокупности значений θ_0, превышающих

$$T_\alpha = \max(y_{(1)} - c,\ y_{(n)} - 1).$$

Критическая область (1) представляется в виде треугольника в верхнем правом углу треугольника, изображающего область возможных значений пары $(y_{(1)}, y_{(n)})$. Вершины последнего треугольника находятся в точках (θ_0, θ_0), (θ_0, θ_0+1) и (θ_0+1, θ_0+1). Область (1) лежит на диагонали $y_{(1)} = y_{(n)}$, левая ее сторона параллельна оси $y_{(n)}$ и проходит через точку (θ_0+c, θ_0+1). Выборки, для которых подчиненная статистика равна a, соответствуют точкам, лежащим на прямой $y_{(n)} = y_{(1)} + a$. Если $a > 1-c$, то область (1) эта прямая не пересекает. Отсюда вытекает, что если $a > 1-c$, то условный уровень области (1) равен нулю, что приводит к искомому результату для нижней доверительной границы.

С нашей точки зрения было бы нецелесообразно приписывать неопределенность безусловным доверительным границам, когда на самом деле известно с уверенностью, что T_α превышает θ, что и происходит при $a > 1-c$. Значение a измеряет степень точности возможной локализации параметра θ. Условное распределение величины $M = \frac{1}{2}(Y_{(1)} + Y_{(n)})$ при заданном $A = a$ — равномерное на $\left[\theta + \frac{1}{2}a,\ \theta + 1 - \frac{1}{2}a\right]$. Соображения, связанные с правдоподобием, наводят на мысль, что весь интервал $\left[M - 1 + \frac{1}{2}a,\ \ M - \frac{1}{2}a\right]$ необходимо рассматривать как доверительный интервал. Наилучшая точная нижняя доверительная граница должна выражаться как $M - \frac{1}{2}a - \alpha(1-a)$.

Недавно Г. А. Барнард показал, что доверительный интервал с наименьшей средней длиной определяется системой условных интервалов с $\alpha = 0$ или $\alpha = 1$ в зависимости от величины значения α. Более ранние ссылки и общее обсуждение проблемы можно найти в работе Пирса (1973).

[Теоретическая статистика,
§ 7. 2; Рао, § 7b. 2]

7.5. Пусть Y_1 и Y_2 — н. о. р. случайные величины, имеющие стандартное распределение Коши с неизвестным параметром сдвига. Покажите, что правдоподобие является бимодальным в том случае, когда $c = y_2 - y_1 > 2$. Исследуйте форму условных доверительных областей для параметра положения в случае $c \leqslant 2$, так и при $c > 2$. Обсудите критически возможность использования значения среднего в качестве статистики положения.

Решение

С точностью до постоянной логарифм правдоподобия записывается в следующем виде:

$$-\log\{1 + (y_1 - \theta)^2\} - \log\{1 + (y_2 - \theta)^2\}.$$

Обозначим $\overline{y}_{.} = \frac{1}{2}(y_1 + y_2)$, $c = y_2 - y_1$ и $\theta = \overline{y}_{.} + \delta$. Логарифм правдоподобия перепишется следующим образом:

$$l(\delta) = -\log\left\{1 + \left(\sigma + \frac{1}{2}c\right)^2\right\} - \log\left\{1 + \left(\delta - \frac{1}{2}c\right)^2\right\},$$

т. е. будет четной функцией от δ.

Непосредственно вычисляется, что

$$l'(\delta) = -2\left\{\frac{\left(\delta + \frac{1}{2}c\right)}{1 + \left(\delta + \frac{1}{2}c\right)^2} + \frac{\left(\delta - \frac{1}{2}c\right)}{1 + \left(\delta - \frac{1}{2}c\right)^2}\right\},$$

$$l''(\delta) = -2\left[\frac{1 - \left(\delta + \frac{1}{2}c\right)^2}{\left\{1 + \left(\delta + \frac{1}{2}c\right)^2\right\}^2} + \frac{1 - \left(\delta - \frac{1}{2}c\right)^2}{\left\{1 + \left(\delta - \frac{1}{2}c\right)^2\right\}^2}\right].$$

В частности, имеем $l''(0) = -4\left(1 - \frac{1}{4}c^2\right)\left(1 + \frac{1}{4}c^2\right)$. Таким образом, получаем, что стационарные точки являются корнями кубического уравнения $l'(\delta) = 0$; один из них равен нулю. Условие бимодальности правдоподобия можно получить либо из требования, чтобы два других корня были действительными, либо из условия, что в нуле достигается минимум правдоподобия, поскольку $l''(0) > 0$.

Условные доверительные интервалы можно получить, шкалируя правдоподобия так, чтобы полный интеграл от них равнялся единице. При $c \leqslant 2$, а также при $c > 2$ и достаточно малых α доверительные области, основанные на правдоподобии, будут интервалами, симметричными относительно $\theta = \overline{y}$. В то же время при $c > 2$ и достаточно больших α доверительные области будут образованы двумя симметрично расположенными относительно точки $\theta = \bar{y}_.$, но не включающими ее интервалами. Итак, если $c \leqslant 2$, то $\bar{y}_.$ в некотором смысле является единственной разумной мерой сдвига. В то же время при $c > 2$ необходимо подробно исследовать функцию правдоподобия, хотя $\bar{y}_.$ по-прежнему остается центром симметрии.

Для выборок объемом больше двух положение дел, конечно, более сложно. Однако для выборок объема 20 и более правдоподобие обычно хорошо аппроксимируется нормальной кривой с центром в точке $\hat{\theta}$ и дисперсией $\{-l''(\hat{\theta})\}^{-1}$. Исходя из этого, можно получить приближенные условные доверительные интервалы. Относительно дальнейших подробностей см. работу Эфрона и Хинкли (1978).

[Теоретическая статистика, § 7.2; Рао, § 7b.2; Барнард и Спрот, 1971]

7.6. Покажите, что если скалярная случайная величина имеет нормальное $N(\mu, 1)$ распределение и если наблюдается $Y = y$, то параметр μ имеет нормальное частотное $N(y, 1)$ распределение в том смысле, который поясняется ниже. Предположим теоретически, что возможно получить пары $(y^{(1)}, \mu^{(1)})$, $(y^{(2)}, \mu^{(2)})$, ..., где каждое $y^{(j)}$ является наблюдением случайной величины $Y^{(j)}$, имеющей распределение $N(\mu^{(j)}, 1)$, а $\mu^{(j)}$ является произвольной величиной. Пусть все пары сдвинуты так, что наблюдаемым значением является y. Это означает, что j-я пара преобразуется к $(y, \mu^{(j)} + y - y^{(j)}) = (y, \mu^{(j)\prime})$, где $\mu^{(j)\prime} = \mu^{(j)} + y - y - y^{(j)}$. Тогда μ имеет частотное распределение $N(y, 1)$. Это теоретическое частотное распределение истинных средних в совокупности повторов опыта, имеющих в качестве наблюденного значения y. Отправляясь от н.о.р. случайных величин $Y_1, \ldots, Y_n$, имеющих нормальное $N(\mu, \sigma^2)$ распределение, используйте аналогичные рассуждения для того, чтобы показать, что μ имеет t-распределение Стьюдента, когда выборки шкалированы так, чтобы иметь наблюденные среднее и среднее квадратическое.

Решение

Математические выкладки весьма просты. Можно написать $Y^{(j)} = \mu^{(j)} + \varepsilon^{(j)}$, где $\varepsilon^{(j)}$ — нормальная $N(0, 1)$ случайная величина. Следовательно, после такого сдвига, что наблюденное значение оказывается фиксированным и равным y, значение параметра оказывается равным $y - \varepsilon^{(j)}$, причем оно имеет нормальное $N(y, 1)$ частотное распределение. Существенный момент состоит в том, что для параметра имеется теоретическое частотное распределение, не зависящее от действительного распределения величин $\mu^{(j)}$. Можно полагать, в частности, что все $\mu^{(j)}$ равны одному и тому же искомому значению μ. Однако в этом нет необходимости.

Когда одновременно неизвестны и масштабный, и сдвиговый параметры, следует рассмотреть линейные преобразования общего вида. Пусть $\bar{y}.$ и $\sqrt{\mathrm{MS}}$ — наблюденные среднее и оценка стандартного уклонения, полученные по n наблюдениям из $N(\mu, \sigma^2)$. Рассмотрим теоретически повторения опыта. Пусть в j-м опыте получено $(\bar{y}^{(j)}, \sqrt{\mathrm{MS}^{(j)}}, \mu^{(j)}, \sigma^{(j)})$. Чтобы привести j-ю совокупность в соответствие с данными, воспользуемся преобразованием y в

$$\sqrt{(\mathrm{MS}/\mathrm{MS}^{(j)})}\, y + \bar{y}. - \sqrt{(\mathrm{MS}/\mathrm{MS}^{(j)})}\, \bar{y}.^{(j)}.$$

Таким образом, $\mu^{(j)}$ преобразуется в

$$\bar{y}. + \sqrt{(\mathrm{MS}/\mathrm{MS}^{(j)})}\,(\mu^{(j)} - \bar{y}.^{(j)}),$$

это означает, что преобразованные средние имеют частотное распределение вида

$$\bar{y}. + T_{n-1}\sqrt{(\mathrm{MS})/n},$$

где T_{n-1} — случайная величина, имеющая t-распределение Стьюдента с $(n-1)$ степенями свободы.

Приведенные соображения (Фрэзер, 1961) придают явную частотную интерпретацию фидуциальному распределению параметра μ.

[Теоретическая статистика, § 7.2;
Рао, § 5.5]

7.7. Параметр μ является неизвестным целым числом. Это число наблюдается с ошибкой, которая с равной вероятностью принимает значение $+1$ и -1. На основе наблюдения y утверждается, что

$$\mathrm{pr}\,(\mu = y - 1) = \mathrm{pr}\,(\mu = y + 1) = \frac{1}{2}.$$

Покажите, что это „вероятностное" утверждение не обладает всеми свойствами вероятностных утверждений. В частности, вскрыть это поможет следующая стратегия заключения пари:

если $y \geqslant 1$, то заключается пари на четную сумму о том, что $\mu = y - 1$;
если $y < 1$, то пари не заключается.

Докажите, что, если $\mu = 0$ или 1, эта стратегия „выигрышна". В то же время, если $\mu \neq 0$ или 1, то она безобидна, т. е. имеет нулевой средний выигрыш. Заметим, что если с вероятностью $\pi_m > 0$ точкой отсечения будет $y = m$, то стратегия выигрышна для всех m, где $m = 0, \pm 1, \ldots$.

Решение

Цель фидуциального подхода, а в некотором смысле и подхода на основе структурных вероятностей (Фрэзер, 1968), состоит в том, чтобы на основе данных получить вероятностное распределение для неизвестных параметров без использования в явном виде предположений байесовской теории. В подходе на основе выборочной теории считается, что хотя доверительные утверждения часто подобны вероятностным утверждениям, однако между ними имеется существенная разница: неизвестные параметры являются единственными в своем роде постоянными и как таковые случайными величинами быть не могут.

Разумно считать, что если неизвестные параметры целесообразно рассматривать как случайные величины, тогда это должно быть возможно и в такой весьма простой ситуации, какая описана здесь. Вычисления показывают, однако, что стратегия заключения пари, при которой держится пари на четную сумму о том, что $\mu = y - 1$, может быть проигрышной. Таким образом, объявленная вероятность, равная 1/2, не имеет одной из обычных „практических" интерпретаций.

Если $\mu = 0$, то ожидаемый выигрыш при размере ставки, равном c, равен

$$c \times \operatorname{pr}(y = +1 \mid \mu = 0) + 0 \times \operatorname{pr}(y = -1 \mid \mu = 0),$$

т. е. положителен. Аналогичные рассуждения справедливы при $\mu = 1$. Для $\mu < 0$ пари не заключается, поскольку y всегда меньше единицы. Для $\mu > 1$ ожидаемый выигрыш равен нулю. Для обобщенной стратегии пусть M — случайная точка отсечения. Тогда при размере ставки, равном c, ожидаемый выигрыш равен

$$\{c \operatorname{pr}(y = \mu + 1 \mid \mu) - c \operatorname{pr}(y = \mu - 1 \mid \mu)\} \operatorname{pr}(M < \mu + 1) +$$
$$+ c \operatorname{pr}(y = \mu + 1 \mid \mu) \operatorname{pr}(M = \mu + 1) = c \operatorname{pr}(y = \mu + 1 \mid \mu) \operatorname{pr}(M = \mu + 1).$$

Он положителен для всех таких μ, что $\mu + 1$ является возможным значением величины M.

[Теоретическая статистика, § 7.2;
Рао, § 5b.5; Бёлер, 1971]

7.9. [1]) Предположим, что необходимо построить $(1-\alpha)$-доверительный интервал заранее заданной ширины $2l$ для неизвестного среднего μ нормального распределения с неизвестной дисперсией. Чтобы достичь этого, берут n_0 наблюдений. Пусть MS_0 — оценка дисперсии, построенная по этим наблюдениям. После этого берутся дополнительные наблюдения, так чтобы общее число наблюдений равнялось

$$\max\left(n_0, \left[4t^{*2}_{n_0-1, \frac{1}{2}\alpha} MS_0/l^2\right]+1\right),$$

где $[x]$ — целая часть числа x. Покажите, что если $\overline{Y}.$ — среднее объединенных данных, то $(\overline{Y}.-l, \overline{Y}.+l)$ определяет консервативный $(1-\alpha)$-доверительный интервал для μ. Доказательство проводите условно относительно значений MS_0. Опишите, как с помощью соответствующей рандомизации консервативный доверительный интервал можно превратить в точный. Обсудите критически возможность более эффективного использования полученных данных при такой процедуре.

Решение

Существует несколько способов обращения того, что можно назвать типичным консервативным интервалом, в точные интервалы. Один из них связан с рандомизацией. Другой определяется следующим образом.

Для каждого n определим постоянные $a_1, \ldots, a_n$, которые зависят от MS_0 и таковы, что

$$a_1+\ldots+a_n=1,\ a_1=\ldots=a_{n_0},$$
$$MS_0(a_1^2+\ldots+a_n^2)=\frac{1}{4}\, l^2/t^{*2}_{n_0-1, \frac{1}{2}\alpha}.$$

Ясно, что подобрать такие a_j возможно. Далее, условно относительно MS_0 величина $(\Sigma a_j Y_j-\mu)/(a_1^2+\ldots+a_n^2)^{1/2}$ имеет нормальное $N(0, \sigma^2)$ распределение. Отсюда вытекает, что безусловное распределение будет таким же и не зависящим от MS_0. Таким образом, статистика

$$T=(\Sigma a_i Y_i-\mu)/\{MS_0\,(a_1^2+\ldots+a_n^2)\}^{1/2}$$

имеет t-распределение Стьюдента с n_0-1 степенями свободы. Отсюда вытекает, что точный доверительный интервал длины $2l$ получен.

[1]) Задача 7.8 опущена

Использование среднего всех наблюдений вместо „неэффективной“ статистики $\Sigma a_j Y_j$ приводит к консервативному интервалу. Более важное замечание касается того, что информация о σ^2, содержащаяся во второй выборке, если таковая берется, никак не используется. На практике наблюдается тенденция использовать в соответствии с внутренней убежденностью в целесообразности этого обычные доверительные интервалы, построенные на основе полных данных, суммирующих всю информацию о μ.

Из рассмотренного примера можно сделать общий вывод, что частотные свойства, которые нельзя получить для выборок фиксированного объема, иногда удается достичь, обращаясь к расширенному выборочному пространству, когда объем выборки может меняться соответствующим образом.

[Теоретическая статистика, § 7.3; Рао, § 7с.6; Стейн, 1945]

7.10. Случайные величины $Y_1, \ldots, Y_n$ независимы и нормально распределены с неизвестной дисперсией σ^2 и $E(Y_j) = \gamma + \beta e^{-\rho x_j}$, где $x_1, \ldots, x_n$ — известные постоянные, а γ, β и ρ — неизвестные параметры. Покажите, что не существует минимальной достаточной статистики, размерность которой была бы меньше размерности выборки. Рассмотрите нулевую гипотезу $\rho = \rho_0$ и покажите, что для локальных уклонений $\rho = \rho_0 + \delta$ модель принимает вид $E(Y_j) = \gamma + \beta e^{-\rho_0 x_j} - \beta\delta x_j e^{-\rho_0 x_j}$. Покажите, используя это, что точный локально наиболее мощный критерий проверки нулевой гипотезы $\rho = \rho_0$ получается при рассмотрении модели $E(Y_j) = \gamma + \beta e^{-\rho_0 x_j} + \psi x_j e^{-\rho_0 x_j}$ и применении теории проверки гипотезы $\psi = 0$ в нормальной модели, для чего используется t-статистика Стьюдента. Получите отсюда $(1-\alpha)$-доверительную область для ρ. Приведите формулировку общего результата, для которого эта задача является частным случаем.

Решение

Правдоподобие принимает вид

$$-\frac{1}{2} n \log(2\pi\sigma^2) - \frac{\Sigma\left(Y_j - \gamma - \beta e^{-\rho x_j}\right)^2}{2\sigma^2}.$$

Таким образом, если ρ — неизвестный параметр, то сокращения данных не происходит.

Для $\rho = \rho_0 + \delta$ можно написать

$$e^{-\rho x_j} = (1 - \delta x_j)\, e^{-\rho_0 x_j} + O(\delta^2).$$

Отсюда получим линеаризованную модель вида

$$E(Y_j)=\gamma+\beta e^{-\rho_0 x_j}+\psi x_j e^{-\rho_0 x_j}+O(\delta^2),$$

где $\psi=-\rho_0\delta$. Таким образом, t-статистика Стьюдента для проверки гипотезы $\psi=0$ будет локально наиболее мощной для проверки гипотезы $\rho=\rho_0$ при $\rho_0\neq 0$. Отметим специально, что критерий является „точным" в том смысле, что статистика критерия имеет при нулевой гипотезе указанное t-распределение Стьюдента. Свойства оптимальности критерия имеют только локальный характер. Хотя это является само собой разумеющимся из-за отсутствия достаточного сокращения данных, в то же время это можно показать, применяя лемму Неймана — Пирсона к распределению данных, условному относительно заданного значения минимальной достаточной статистики при $\rho=\rho_0$.

Точный доверительный интервал для ρ получается как множество значений ρ, не „отвергаемых" введенным выше критерием. Численно решить задачу в принципе не сложно. Однако не очевидно, что доверительная область будет интервалом, хотя, как правило, это будет так.

Общая постановка задачи связана с теорией нормальных моделей вида

$$E(Y_j)=g_j(\beta_1, \ldots, \beta_p, \rho),$$

где зависимость от $\beta_1, \ldots, \beta_p$, но не от ρ, линейна. В более общей ситуации ρ можно считать вектором и точный критерий проверки гипотезы $\rho=\rho_0$ можно построить еще на основе использования F-распределения. Если зависимость от параметров $\beta_1, \ldots, \beta_p$ нелинейна, то точных критериев построить уже нельзя.

Существует возможность перехода к более общим семействам экспоненциального типа, не обязательно нормальным.

[Теоретическая статистика, § 7.3; Рао, § 7b.2; Уильямс, 1962]

7.11. Независимые бинарные испытания проводятся с постоянной, но не известной вероятностью успеха. В n испытаниях наблюдались r успехов. Что можно сказать относительно числа успехов, которые будут наблюдены в дальнейших m испытаниях?

Решение

Рассмотрим независимые случайные величины R и $R^\dagger$, имеющие биномиальные распределения числа успехов в n и m испытаниях с вероятностями успехов θ и $\theta^\dagger$ соответственно. Тогда, чтобы получить нижнюю $(1-\alpha)$-доверительную границу для $R^\dagger$, построим критическую область размера α для проверки гипотезы $\theta=\theta^\dagger$ против альтернатив $\theta>\theta^\dagger$. Совокупность значений $(r, r^\dagger)$, не принадлежащих критической области, определяет значение будущего наблюдения $r^\dagger$, согласующегося с наблюденным значением r. „Точное" решение возможно на основе критерия Фишера. Представим результаты наблюдений так, как если бы речь шла о критерии Фишера для таблиц сопряженности размера 2×2:

Успехи	Неудачи	Всего
r	$n-r$	n
$r^\dagger$	$m-r^\dagger$	m

Критерий Фишера проверяет гипотезу о том, что между строками не существует значимой разности, когда $r^\dagger$ — возможное значение для $R^\dagger$.

Приближенное решение дается нормальной аппроксимацией распределения $R^\dagger/m-R/n$ с использованием поправки на непрерывность. Так, при двусторонней процедуре значения $r^\dagger$ согласуются с r на уровне α, если они удовлетворяют условию

$$\frac{\left\{\left|r^\dagger-\dfrac{m(r+r^\dagger)}{m+n}\right|-\dfrac{1}{2}\right\}^2}{\left\{\dfrac{mn(r+r^\dagger)(m+n-r-r^\dagger)}{(m+n)^3}\right\}}\leqslant k^{*}_{\frac{1}{2}\alpha}.$$

Заметим, что неявная симметрия рассматриваемой задачи состоит в том, что если

$$r=a,\ n=b;\quad r^\dagger=a',\ m=b'$$

согласуются на уровне α, то согласуются также и

$$r=a',\ n=b';\quad r^\dagger=a,\ m=b.$$

В этом смысле процесс обратим во времени.

[Теоретическая статистика, § 7.5]

8. Точечное оценивание

Необходимые сведения

В гл. 7 описывались методы оценивания параметров на основе построения некоторого множества их значений, согласующихся с данными. При точечном оценивании задается одно-единственное значение. Особая осторожность необходима при постановке задачи точечного оценивания. При этом можно различать по крайней мере три типа ситуаций, в которых используются точечные оценки:

(a) Часто, возможно после преобразования параметра, верхняя и нижняя $(1-\alpha)$-доверительные границы представляются приблизительно в виде $t \pm k^*_{\frac{1}{2}\alpha}\sigma_t$. В этом случае значение t, дополненное мерой точности σ_t, представляют собой удобный способ суммирования информации о θ. Однако представляется более целесообразным рассматривать это как способ сжатого представления доверительных границ, а не метод точечного оценивания как таковой.

(b) В задачах принятия решения может потребоваться на самом деле дать одно-единственное значение, например, как основу некоторого критерия контроля. Для того чтобы избежать до некоторой степени произвольности выбора оценки, здесь необходима полная постановка задачи принятия решения (гл. 11).

(c) При анализе больших массивов данных может оказаться разумным вычислить точечную оценку T некоторого параметра θ по каждой совокупности данных, а затем на второй стадии анализа рассмотреть, как θ меняется по всем совокупностям данных.

Если анализ на второй стадии линеен, то необходимо потребовать, чтобы на первой стадии анализа каждое смещение $b(\theta)=E(T;\,\theta)-\theta$ было мало.

В общем случае средняя квадратическая ошибка оценки[1]) определяется как

$$E\{(T-\theta)^2;\ \theta\}=\{b(\theta)\}^2+\operatorname{var}(T;\,\theta).$$

---

[1]) Здесь автором употреблен термин "estimator", т. е. способ оценивания. Если смысл не будет страдать, он будет переводиться так же как термин "estimate"—оценка, т. е. реализованное значение способа оценивания для данных y.— *Прим. перев.*

Один общий подход к построению точечных оценок заключается в том, чтобы ограничить класс рассматриваемых оценок. Это можно сделать, например, требуя от оценок инвариантности относительно некоторого вида преобразований. После этого внутри выделенного класса ищутся оценки с минимальной среднеквадратической ошибкой.

Довольно изящная теория построения оценок связана с требованием несмещенности, т. е. от оценки требуется, чтобы $b(\theta)=0$. В этом случае стремятся к тому, чтобы выбрать T, минимизирующую $\operatorname{var}(T;\ \theta)$, т. е. приводящую к несмещенным оценкам с минимальной дисперсией. В приложениях к таким оценкам надо относиться с известной долей осторожности.

В соответствии с неравенством Крамера — Рао в регулярных задачах с одномерным параметром θ

$$\operatorname{var}(T;\ \theta)\geqslant\{1+b'(\theta)\}^2/i_{.}(\theta),$$

где $i_{.}(\theta)$ — информация по Фишеру выборки (см. гл. 4). Для несмещенных оценок

$$\operatorname{var}(T;\ \theta)\geqslant 1/i_{.}(\theta).$$

Нижняя граница достигается в том и только том случае, если T — линейная функция эффективного вклада $U_{.}(\theta)$. Для векторного параметра насмещенная оценка T компоненты θ удовлетворяет неравенству

$$\operatorname{var}(T_1;\ \theta)\geqslant i_{.}^{11}(\theta),$$

где $i_{.}^{rs}(\theta)$ — $(r,\ s)$-й элемент матрицы $\{i_{.}(\theta)\}^{-1}$. Нижняя граница достигается в том и только том случае, если T_1 — линейная функция компонент вектора $U_{.}(t)$. В более общем случае несмещенная оценка параметрической функции $l^{\mathrm{T}}(\theta)$ имеет дисперсию, не меньшую, чем $l^{\mathrm{T}}\{i_{.}(\theta)\}^{-1}l$.

Если для параметра θ существует полная достаточная статистика S, то несмещенную оценку с минимальной дисперсией можно получить с помощью следующей процедуры, основанной на теореме Рао — Блэкуэлла. Пусть V — любая несмещенная оценка, а $T=E(V\,|\,S)$. Легко показать, что $E(T;\ \theta)=0$, а $\operatorname{var}(T;\theta)\leqslant\operatorname{var}(V;\ \theta)$. Единственность оценки T является следствием полноты достаточной статистики S. Таким образом, любая функция полной достаточной статистики является несмещенной оценкой с минимальной дисперсией своего ожидания.

Имеются два важных метода приближенного исключения смещения. Они связаны с разложениями в ряд и с расщеплением выборки. В первом случае предполагается, что объем n выборки велик и что можно найти такую несмещенную оценку T, что

$$E(T;\ \theta)=\theta+a(\theta)/n+O(n^{-2}). \tag{1}$$

Тогда $T - a(T)/n$ имеет смещение порядка менее n^{-1}. Этот способ часто применяется в ситуации, когда $T = g(S)$, $\theta = g(\varphi)$ и S — несмещенная оценка для φ с дисперсией порядка n^{-1}. В этом случае

$$a(\theta)/n = \frac{1}{2}\{g'(\varphi)\}^2 \operatorname{var}(S;\ \varphi).$$

При использовании метода расщепления выборки обозначим T_n оценку, построенную по $Y_1, \ldots, Y_n$, а $T_{n-1,j}$ — аналогичную оценку по той же выборке, из которой исключено лишь Y_j. Тогда если разложение (1) имеет место для выборок объема n и $(n-1)$, то, положив $T_{n-1,\cdot} = \Sigma T_{n-1,j}/n$, легко показать, что оценка

$$T_n^J = nT_n - (n-1)\,\bar{T}_{n-1,\cdot} \tag{2}$$

имеет смещение в n раз меньшее, чем оценка T_n. Дисперсии $\operatorname{var}(T_n^J;\ \theta)$ и $\operatorname{var}(T_n;\ \theta)$ можно оценить как

$$\frac{n-1}{n}\Sigma(T_{n-1,j} - \bar{T}_{n-1,\cdot})^2. \tag{3}$$

Хотя метод расщепления полезен, когда свойства распределений статистик в точности неизвестны, однако нельзя считать, что он хорошо работает для существенно нелинейных статистик или чрезвычайно неоднородных данных.

Тематика устойчивого оценивания возникает в том случае, когда требуется, чтобы оценки не были чувствительны к едва различимым уклонениям от предположений о виде распределений, например когда в выборке присутствует малая доля посторонних наблюдений. При оценивании параметров сдвига в теории устойчивых оценок в качестве замены выборочного среднего рекомендуется использовать статистики, в которых уменьшено влияние или отображены соответствующим образом (например, симметрично) крайние порядковые статистики.

Задачи

8.1. Предположим для простоты, что имеется скалярный параметр. Пусть правдоподобие достигает своего наибольшего значения в точке, в которой существует вторая производная. Назовем значение, в котором достигается максимум правдоподобия, оценкой максимального правдоподобия. Для достаточно больших α доверительный интервал, основанный на правдоподобии, имеющий доверительный уровень $1-\alpha$, приближенно симметричен относительно оценки максимального правдоподобия. В соответствии с точкой зрения, что точечные оценки иногда полезны как величины, указывающие на центр доверительных интервалов, оценки максимального правдоподобия обладают и этим специфическим свойством. Какие дальнейшие свойства функции правдоподобия сделали бы это еще более убедительным?

Решение

Высказанные соображения дают слабое обоснование целесообразности использования оценок максимума правдоподобия в том случае, когда абсолютный максимум правдоподобия достигается в стационарной точке. Если функция правдоподобия унимодальна и приближенно симметрична относительно точки максимума, тогда обоснование можно усилить. Дело в том, что доверительные области, основанные на правдоподобии и имеющие уровень $1-\alpha$ для всех α, будут приближенно симметричными относительно оценки максимального правдоподобия.

Заметим, что в таких случаях для любой конкретной выборочной точки y функцию правдоподобия можно сделать в точности симметричной соответствующими преобразованиями параметра. Как правило, такие преобразования зависят от y. Одно из преобразований, не зависящее от данных, при котором наблюдается тенденция к симметризации правдоподобия, определяется следующим образом: при этом преобразовании третья производная логарифма правдоподобия имеет нулевое ожидаемое значение в точке, отвечающей истинному значению параметра.

[Теоретическая статистика, §8.1; Рао, § 5.3; Силвей, гл. 4]

8.2. Для непрерывных распределений оценку T называют медианно смещенной оценкой для θ, если для всех θ

$$\mathrm{pr}\,(T<\theta;\ \theta)=\mathrm{pr}\,(T>\theta;\ \theta)=\frac{1}{2}.$$

Докажите, что если $g(\ )$ строго монотонна, то $g(T)$ медианно несмещена для $g(\theta)$. Пусть SS_d — остаточная сумма квадратов в нормальной теории, имеющая d степеней свободы. Покажите, что медианно несмещенная оценка дисперсии имеет вид SS_d/c_d, где $c_d\simeq d-\frac{2}{3}$. Покажите также, что то же самое значение величины c_d используется как приближенное, если оценка есть то значение $\tilde{\theta}$, для которого $\mathrm{pr}\,(T\leqslant t_{\text{набл}};\ \tilde{\theta})=\frac{1}{2}$, что приводит к другой медианно несмещенной оценке. Обсудите критически условия, при которых эти требования могут оказаться целесообразными.

Решение

Факт инвариантности медианной несмещенности относительно строго монотонных преобразований вытекает из эквивалентности событий $T < \theta$ и $g(T) < g(\theta)$, где без потери общности в качестве $g(\cdot)$ взята возрастающая функция. Указанная инвариантность делает медианную несмещенность на первый взгляд привлекательной в качестве одного из понятий, на основе которого будет принудительно обеспечена единственность в задаче точечного оценивания. Тем не менее представляется, что это понятие не имеет особой ценности, в частности, когда точечная оценка используется для представления в компактном виде доверительных интервалов или для доверительного сокращения. Отметим в то же время, что если необходимо проверить гипотезу H_0: $\theta = \theta_0$ на основе большого числа независимых массивов данных, то сокращение каждого массива до медианно несмещенной оценки позволило бы просто применить критерий знаков.

Аналогичные замечания относятся и ко второму критерию. Спрашивают: при каком значении параметра θ величина $t_{\text{набл}}$ соответствует центру своего выборочного распределения? Хотя это свойство оценки имеет определенную интуитивную привлекательность, однако оно представляется не более чем математическим приемом принудительного обеспечения единственности при выборе оценки.

Чтобы из SS_d образовать медианно несмещенную оценку, используем тот факт, что SS_d/σ^2 имеет хи-квадрат распределение с d степенями свободы. Если медиану этого распределения обозначить $c^*_{d,\frac{1}{2}}$, тогда требуемой оценкой для σ^2 будет $T = SS_d/c^*_{d,\frac{1}{2}}$. Для нахождения величин $c^*_{d,\frac{1}{2}}$ можно использовать таблицы. С другой стороны, можно воспользоваться разложением Эджворта для хи-квадрат распределения со средним d, дисперсией $2d$ и коэффициентом асимметрии $\gamma_1 = 2\sqrt{2}/\sqrt{d}$. На основе обращения Фишера—Корниша[1]) медиана равна приближенно

$$d - \frac{1}{6}\gamma_1\sqrt{2d} = d - \frac{2}{3}.$$

На самом деле это приближение на удивление точное: для $d = 2$ и $d = 10$ истинное значение медиан равно 1.39 и 9.34 соответственно.

---

[1]) См. стр. 510 русского издания „Теоретической статистики". — *Прим. перев.*

[Теоретическая статистика, § 8.2]

8.3. Покажите, что если $Y_1, \ldots, Y_n$ — и. о. р. с плотностью

$$\exp\{a(\theta)\,b(y)+c(\theta)+d(y)\},$$

тогда $\frac{1}{n}\Sigma b(Y_j)$ достигает нижней границы Крамера — Рао для несмещенных оценок своего ожидания.

Решение

Общий эффективный вклад равен

$$U_{.}(\theta)=a'(\theta)\,\Sigma b(Y_j)+nc'(\theta).$$

Поскольку $n^{-1}\Sigma b(Y_j)$ является линейной функцией от $U_{.}(\theta)$, то в неравенстве Крамера — Рао достигается равенство.

[Теоретическая статистика, § 8.3; Рао, § 5.2; Силвей, § 2.10]

8.4. Объясните на качественном уровне присутствие выражения $\{1+b'(\theta)\}^2$ в общей форме неравенства Крамера — Рао. В процессе этого:

(a) объясните, почему появляется $b'(\cdot)$, а не $b(\cdot)$;

(b) объясните, почему граница должна обращаться в нуль, когда $b'(\theta)=-1$;

(c) объясните, почему упомянутое выше выражение возводится в квадрат, а не в первую степень.

Обсудите, почему обращение в нуль смещения $b(\cdot)$ в изолированной точке θ_0 недостаточно для обоснования упрощенного варианта границы в точке θ_0.

Решение

(a) Добавление произвольной постоянной к оценке:
(i) оставляет дисперсию без изменения,
(ii) не изменяет $b'(\cdot)$,
(iii) произвольно меняет $b(\cdot)$.

Все это наводит на мысль, что граница зависит от $b'(\cdot)$, а не от $b(\cdot)$.

(b) Оценка, принимающая постоянное значение, имеет нулевую дисперсию; для нее $b'(\theta)=-1$.

(c) Граница должна быть неотрицательной.

При оценивании среднего μ нормального $N(\mu, 1)$ распределения вырожденная оценка, постоянно равная нулю, имеет нулевое смещение при $\mu=0$. Таким образом, при $\mu=0$ нулевая дисперсия противоречит простой нижней границе n^{-1} для дисперсии несмещенной оценки. Тщательное исследование доказательства показывает, что простая граница предполагает обращение в нуль смещения в некоторой открытой окрестности рассматриваемого параметрического значения.

[Теоретическая статистика, § 8.3;
Рао, § 5.2; Силвей, § 2.10]

8.5. При выполнении некоторых условий регулярности уравнение $k(Y,\ \tilde{\theta})=0$, рассматриваемое как уравнение относительно $\tilde{\theta}$, называют уравнением оценивания для θ, если $E\{k(Y;\theta);\theta\}=0$ для всех θ. Такое уравнение называют оптимальным, если оно имеет наименьшее возможное значение индекса

$$E\{k^2(Y;\ \theta);\ \theta\}/E[\{\partial k(Y,\ \theta)/\partial\theta\}^2;\ \theta].$$

Попытайтесь обосновать это распределение. Покажите, что значение индекса не менее $1/i\ (\theta)$, нижняя граница достигается в том и только том случае, когда $k(.\ ,\ .)$ пропорционально эффективному вкладу. Обобщите рассмотренные понятия и результаты на случай многомерного параметра.

Решение

Наличие быстрых итеративных процедур решения нелинейных уравнений означает, что оценки в явном виде определить нельзя. Тем самым представляется естественной необходимость изучения уравнений оценивания как таковых. Условие $E\{k(Y,\theta);\ \theta\}=0$ является нормирующим условием. Оно гарантирует, грубо говоря, возможность правильного восстановления значения параметра θ в присутствии пренебрежимо малой случайной вариации. Если $k(Y,\ \theta)$ в точности или приближенно линейно по θ для каждого Y, тогда можно написать

$$k(Y,\ \theta)+(\tilde{\theta}-\theta)\frac{\partial k(Y,\ \theta)}{\partial\theta}=0. \tag{1}$$

Предположим, что S.D. $\{\partial k(y,\ \theta)/\partial\theta;\ \theta\}\ll|E\{\partial k(Y;\ \theta)/\partial\theta;\ \theta\}|$. Отсюда вытекает, что предложенный выше индекс с точностью до членов первого порядка приближенно равен $\operatorname{var}(\tilde{\theta};\ \theta)$, минимизация которой представляется весьма разумной.

Продифференцируем по θ обе части равенства

$$E\{k(Y;\ \theta);\ \theta\}=0.$$

Получим

$$\int k(y;\ \theta)\frac{1}{f_Y(y;\ \theta)}\frac{\partial f_Y(y;\ \theta)}{\partial\theta}f_Y(y;\ \theta)\,dy+\int\frac{\partial k(y;\ \theta)}{\partial\theta}f_Y(y;\ \theta)\,dy=0.$$

Требуемая нижняя граница найдется непосредственно из неравенства Шварца:

$$\operatorname{var}\{k(Y,\ \theta);\ \theta\} \operatorname{var}\{U.(\theta);\ \theta\} \geqslant [\operatorname{cov}\{k(Y,\ \theta),\ U.(\theta);\ \theta\}]^2$$

и предшествующего равенства.

Когда θ — вектор-столбец с q компонентами, требуется иметь q оценивающих функций $k_1(Y,\ \theta),\ \ldots,\ k_q(Y,\ \theta)$. Если из этих функций образовать вектор $k(Y;\ \theta)$ и обозначить $g(Y;\ \theta)$ матрицу с элементами $g_{rs}(Y;\ \theta) = \partial k_r(Y,\ \theta)/\partial\theta_s$, тогда соотношение (1) примет вид

$$k(Y,\ \theta) + \mathbf{g}(Y,\ \theta)(\bar{\theta} - \theta) = 0$$

и

$$\bar{\theta} - \theta = -\{\mathbf{g}(Y,\ \theta)\}^{-1} k(Y,\ \theta).$$

Если предположить, что S.D. $\{g_{rs}(Y;\ \theta);\ \theta\} \ll |E\{g_{rs}(Y;\ \theta);\ \theta\}|$, то соответствующим индексом для оценивания компоненты θ_s будет, таким образом,

$$\sum_{r,t} E\{k_r(Y,\ \theta) k_t(Y,\ \theta);\ \theta\} E\{g^{rs}(Y,\ \theta);\ \theta\} E\{g^{ts}(Y,\ \theta);\ \theta\}.$$

[Теоретическая статистика, §8.3;
Рао, § 5а.1.2; Годемб, 1960]

8.6. Рассмотрим частный случай ситуации, обсуждавшейся в задаче 8.5, когда оценка получается в явном виде из уравнения $g_1(Y) - \bar{\theta} g_2(Y) = 0$, где статистики $g_j(\cdot)$ таковы, что $E\{g_1(Y);\ \theta\} = \theta E\{g_2(Y);\ \theta\}$ для всех θ. Соответствующее уравнение называют уравнением несмещенного оценивания. Как можно было бы образовать объединенную оценку, если даны несколько совокупностей данных, порожденных общим значением θ?

Решение

Условие $E\{g_1(Y);\ \theta\} = \theta E\{g_2(Y);\ \theta\}$, связанное с уравнением несмещенного оценивания, является частным случаем требования $E\{k(Y;\ \theta);\ \theta\} = 0$, рассмотренного в задаче 8.5.

Если $g_{1j}(Y_j) - \bar{\theta} g_{2j}(Y_j) = 0$ является оценивающим уравнением для j-й совокупности данных Y_j $(j = 1,\ \ldots,\ m)$, то несмещенное оценивание уравнения примет вид

$$\Sigma w_j g_{1j}(Y_j) - \bar{\theta} \Sigma w_j g_{2j}(Y_j) = 0,$$

где w_j — произвольные постоянные. Это уравнение соответствует уравнению, рассмотренному в предыдущей задаче при

$$k(Y,\ \theta) = \Sigma w_j g_{1j}(Y_j) - \theta \Sigma w_j g_{2j}(Y_j).$$

Отсюда если Y_j независимы, то индекс для задачи 8.5 принимает вид

$$\frac{\Sigma w_j^2 \operatorname{var}\{g_{1j}(Y_j) - \theta g_{2j}(Y);\, \theta\}}{[\Sigma w_j E\{g_{2j}(Y_j);\, \theta\}]^2}.$$

В частных случаях постоянные w_j можно выбрать так, чтобы точно минимизировать индекс (1). Рассмотрим общий случай. Пусть массивы данных имеют схожую структуру. А именно j-я совокупность данных содержит n_j наблюдений, а $g_{1j}(\cdot)$ и $g_{2j}(\cdot)$ эквивалентны суммам однородных наблюдений. Тогда индекс (1) приближенно пропорционален

$$\Sigma n_j w_j^2/(\Sigma n_j w_j)^2.$$

Последнее отношение минимизируется при $w_j = \text{const}$, что эквивалентно простому сложению отдельных оценивающих уравнений.

[Теоретическая статистика, § 8.3;
Рао, § 6а, 6а.2; Дурбин, 1960]

8.7. Авторегрессионный процесс первого порядка Y_j, $j = 0, 1, \ldots$, определим как последовательность

$$Y_{j+1} - \theta Y_j = \varepsilon_{j+1},$$

где $Y_0 = y_0$ — постоянная, а н. о. р. случайные величины $\varepsilon_1, \ldots, \varepsilon_n$ имеют нормальное распределение с нулевыми средними. Покажите, что

$$\Sigma Y_j Y_{j+1} - T \Sigma Y_j^2 = 0$$

является уравнением несмещенного оценивания, для которого достигается нижняя граница индекса, определенная в задаче 8.5.

Решение

Имеем

$$Y_j = \theta^j y_0 + \theta^{j-1}\varepsilon_1 + \ldots + \varepsilon_j$$

для $j = 1, \ldots, n$. Таким образом, если $\operatorname{var}(\varepsilon_j) = \sigma^2$, то

$$E(Y_j Y_{j+r}) = \theta^{2j+r} y_0^2 + \sigma^2 (1 + \theta^2 + \ldots + \theta^{2j-2})\theta^r \quad (r = 0,\ 1).$$

Отсюда непосредственно вытекает, что

$$E\left(\sum_{j=0}^{n-1} Y_j Y_{j+1}\right) = \theta E\left(\sum_{j=0}^{n-1} Y_j^2\right),$$

т. е. уравнение оценивания является несмещенным.

Поскольку процесс является марковским, то логарифм правдоподобия имеет вид

$$\log f_Y(y;\ \theta) = \sum_{j=0}^{n-1} \log f_{Y_{j+1}\mid Y_j}(y_{j+1}\mid y_j;\ \theta) =$$

$$= -\frac{1}{2}\, n \log \sigma^2 - \frac{1}{2} \sum_{j=0}^{n-1} (y_{j+1} - \theta y_j)^2/\sigma^2.$$

Дифференцируя его по θ, получим эффективный вклад

$$\frac{1}{2} \sum_{j=0}^{n-1} y_j\,(y_{j+1} - \theta y_j)/\sigma^2.$$

Поскольку эффективный вклад пропорционален оценивающей функции, то из задачи 8.5 получаем, что нижняя граница дисперсии достигается.

Для больших n влиянием случайности величины y_0 можно пренебречь.

[Теоретическая статистика, § 8.3; Рао, § 5а; Силвей, гл. 2; Дурбин, 1960]

8.8[1]. Вычислите эффективный вклад и информацию для н. о. р. случайных величин, имеющих показательную плотность, параметризованную сначала с помощью среднего μ как $\mu^{-1} \exp\{-\mu^{-1}y\}$, а затем через $\rho = \mu^{-1}$ как $\rho e^{-\rho y}$.

Решение

Если $Y_1, \ldots, Y_n$ — н. о. р. случайные величины, имеющие показательное распределение со средним μ, то п. р. в. одного наблюдения имеет вид $\mu^{-1} \exp\{-\mu^{-1}y\}$. Отсюда доля наблюдения Y_j в общем эффективном вкладе выборки равна

$$U_j(\mu) = \frac{\partial}{\partial\mu}(-\log \mu - Y_j/\mu) = -\mu^{-1} + \mu^{-2} Y_j,$$

а следовательно,

$$U_.(\mu) = -n\mu^{-1} + \mu^{-2} \Sigma Y_j.$$

Далее, информация, отвечающая Y_j, равна

$$E\{-\partial U_j(\mu)/\partial\mu;\ \mu\} = -\mu^{-2} + 2\mu^{-3} E(Y_j;\ \mu) = \mu^{-2},$$

а общая информация выборки составляет, таким образом, $i_.(\mu) = n\mu^{-2}$.

---

[1]) Улучшенный вариант.

Однако если п. р. в. записывается в виде $\rho e^{-\rho y}$, то можно вычислить эффективный вклад и информацию для ρ. Соответствующие характеристики, вычисленные для ρ, будем снабжать верхним знаком *. Получим

$$U_j^*(\rho)=\frac{\partial}{\partial\rho}(\log\rho-Y_j)=\frac{1}{\rho}-Y_j;\ \partial U_j^*(\rho)/\partial\rho=-\rho^{-2},$$

откуда

$$U_.^*(\rho)=n\rho^{-1}-\Sigma Y_j,\ i_.^*(\rho)=n\rho^{-2}.$$

Эти соотношения иллюстрируют общую связь репараметризованных характеристик, рассмотренную в гл. 4. В нашем частном случае она принимает вид

$$U^*=U\,(d\mu/d\rho),\ i^*=i\,(d\mu/d\rho)^2,$$

где $d\mu/d\rho=-\rho^{-2}$.

[Теоретическая статистика, § 8.3, 4.8; Рао, § 5а; Силвей, § 2.8]

8.10[1]). Пусть $Y_1, \ldots, Y_n$ — и. о. р. случайные величины, имеющие показательную плотность $\rho e^{-\rho y}$. Покажите, что несмещенная оценка с минимальной дисперсией для функции распределения величин Y_j, вычисленная в точке y, равна

$$1-\{1-\min(y,\ S)/S\}^n,$$

где $S=\Sigma Y_j$.

Решение.

Воспользуемся общим результатом теоремы Рао — Блекуэлла. Будем исходить из полной достаточной статистики $S=\Sigma Y_j$ и простой несмещенной оценки V, определяемой как

$$V=\begin{cases}1 & (Y_1\leqslant y),\\ 0 & (Y_1>y).\end{cases}$$

Чтобы вычислить $T=E(V\,|\,S)$ — несмещенную оценку с минимальной дисперсией, необходимо получить условную плотность случайной величины Y_1 при заданном значении $S=s$. Эта плотность вычисляется как

$$\frac{f_{Y_1}(y;\ \rho)\,f_{S-Y_1}(s-y;\ \rho)}{f_S(s;\ \rho)}=\frac{e^{-\rho y}\rho\,\{\rho\,(s-y)\}^{n-1}\rho\exp\{-\rho\,(s-y)\}/(n-1)!}{\rho\,(\rho s)^n\,e^{-\rho s}/n!}=$$
$$=n\,(1-y/s)^{n-1}\qquad(0\leqslant y\leqslant s).$$

[1]) Задача 8.9 опущена.

Отсюда вытекает, что

$$T = E(V \mid S) = \mathrm{pr}(V = 1 \mid S) = \begin{cases} 1 & (y \geqslant S), \\ 1 - (1 - y/s)^n & (y < S). \end{cases}$$

Для больших n оценка T равна приближенно $1 - \exp\{-ny/S\}$. Последнее выражение совпадает с оценкой максимального правдоподобия.

[Теоретическая статистика, § 8.4; Рао, § 5а.2; Силвей, § 2.4, 2.6]

8.11. В связи с оцениванием дисперсии σ^2 нормального распределения постройте несмещенную оценку для $\log \sigma$, используя: (а) точные и (b) приближенные методы. Когда целесообразно использовать полученные результаты? Подскажите, как можно применить рассмотренные методы оценивания для анализа данных наблюдений пуассоновского процесса.

Решение

Пусть в нормальной модели MS оценка для σ^2 основана на d степенях свободы. Тогда MS d/σ^2 имеет хи-квадрат распределение с d степенями свободы. Откуда

$$E\left(\frac{1}{2}\log \mathrm{MS}\right) = \log \sigma - \frac{1}{2}\log d + $$

$$+ \int_0^\infty \left(\frac{1}{2}\log x\right) \frac{\frac{1}{2}\left(\frac{1}{2}x\right)^{\frac{1}{2}d-1} e^{-\frac{1}{2}x}}{\Gamma\left(\frac{1}{2}d\right)}\, dx =$$

$$= \log \sigma - \frac{1}{2}\log d + \Gamma'\left(\frac{1}{2}d\right) / \Gamma\left(\frac{1}{2}d\right).$$

Следовательно, требуемая оценка имеет вид $\frac{1}{2}\log \mathrm{MS} + \frac{1}{2}\log d - \psi\left(\frac{1}{2}d\right)$. Имеются таблицы и приближения для ψ-функции $\psi = \Gamma'/\Gamma$ (см. Абрамович и Стеган, 1965).

При использовании приближенных методов заметим, что из свойств хи-квадрат распределения имеем $E(\mathrm{MS}) = \sigma^2$ и $\mathrm{var}(\mathrm{MS}) = 2\sigma^4/d$. Отсюда, используя разложение (1) из задачи 8.12 с $\theta = \sigma^2$ и $g(\theta) = \frac{1}{2}\log \theta$, получим

$$E\left(\frac{1}{2}\log \mathrm{MS}\right) = \log \sigma - 1/d + \ldots .$$

Таким образом, искомая оценка равна приближенно $\frac{1}{2}\log \mathrm{MS} + 1/d$.

Преимущество второго подхода, кроме его простоты, состоит в том, что он применим к построению оценок, основанных на данных, порожденных не обязательно нормальными моделями.

Полученные результаты находят применение в ситуации, когда имеется несколько таких оценок и необходимо представить $\log \sigma$ в виде линейной модели от контролируемых переменных, т. е. речь идет о мультипликативной модели для σ. Поправка на смещение может иметь важное значение в том случае, когда оценки основаны на разных степенях свободы.

Если ρ — интенсивность пуассоновского процесса, а T_m — время до осуществления m-го события после „данного" момента, то $2\rho T_m$ имеет хи-квадрат распределение с $2m$ степенями свободы. Таким образом, обсужденные выше оценки применимы и для пуассоновской модели с $d = 2m$, $\mathrm{MS} = T_m/m$ и $\sigma^2 = 1/\rho$.

[Теоретическая статистика, § 8.4; Рао, § 6а, 6d]

8.12. Используя разложение Тейлора функции $g(t)$ в окрестности $t = 0$, покажите, что если $E(T;\ \theta) = 0$, а $\mathrm{var}(T;\ \theta)$ мала, то

$$\mathrm{var}\{g(T);\ \theta\} \simeq \{g'(\theta)\}^2\, \mathrm{var}(T;\ \theta).$$

Получите отсюда способ преобразования независимых оценок нескольких параметров $\theta_1, \ldots, \theta_m$ с дисперсиями $v(\theta_1), \ldots, v(\theta_m)$ в величины с приближенно постоянной изменчивостью.

Решение

Заметим, что разложение функции $g(\cdot)$ в ряд Тейлора имеет вид

$$g(T) = g(\theta) + (T - \theta)\, g'(\theta) + \frac{1}{2}(T - \theta)^2 g''(\theta) + \ldots .$$

Отсюда сначала получаем

$$E\{g(T);\ \theta\} = g(\theta) + \frac{1}{2}\mathrm{var}(T;\ \theta)\, g''(\theta) + \ldots . \tag{1}$$

Далее,

$$[g(T) - E\{g(T);\ \theta\}]^2 = (T - \theta)^2 \{g'(\theta)\}^2 + $$
$$+ \{(T - \theta)^3 - (T - \theta)\,\mathrm{var}(T;\ \theta)\}\, g'(\theta)\, g''(\theta) + \ldots .$$

Возьмем ожидание от обеих частей равенства

$$\mathrm{var}\{g(T);\ \theta\} = \{g'(\theta)\}^2\, \mathrm{var}(T;\ \theta) + \ldots . \tag{2}$$

Второе слагаемое в равенстве (1) представляет собой поправку на смещение, в то время как соотношение (2) дает искомое выражение для дисперсии с точностью до членов первого порядка. Можно вычислить также и дальнейшие члены разложения. Строгое доказательство полученных соотношений потребует непосредственного использования свойств предельных операций.

Обобщить полученные результаты на случай векторных статистик T совсем просто. Они могут быть использованы, в частности, для получения соответствующих полезных результатов для отношений.

Наиболее простой путь для получения приближенно постоянной дисперсии состоит в том, чтобы найти достаточно простую форму связи между $v(\theta)$ и θ. Часто это можно сделать, составляя график значений $\log v(\bar{\theta}_i)$ для соответствующих оценок $\log \bar{\theta}_i$. При этом можно попытаться подобрать зависимость вида $v(\theta) = a\theta^b$ для соответствующего значения b. Если $g(T)$ приближенно имеет постоянную дисперсию, например равную 1, то из соотношения (2) будем иметь

$$1 = \{g'(\theta)\}^2 v(\theta).$$

Откуда

$$g(\theta) = \int^{\theta} dx/\sqrt{v(x)}.$$

В частном случае $v(\theta) = a\theta^b$ это приводит к преобразованию

$$g(T) \propto \begin{cases} \log T & (b = 2), \\ T^{1-\frac{1}{2}b} & (b \neq 2). \end{cases}$$

Стандартные преобразования, полученные на основе рассмотренных доводов, включают преобразование корень квадратный из пуассоновских величин, логарифмическое преобразование оценки дисперсии для нормальных моделей и обратный гиперболический тангенс для выборочного коэффициента корреляции.

[Теоретическая статистика, § 8.4;
Рао, § 6а.2,6]

8.13. Говорят, что случайная величина Y распределена логарифмически нормально, если случайная величина $Z = \log Y$ распределена нормально, например со средним λ и дисперсией τ^2. Покажите, что $E(Y) = \exp\left(\lambda + \frac{1}{2}\tau^2\right)$. Для совокупности н. о. р. случайных величин $Y_1, \ldots, Y_n$, имеющих логарифмически нормальное распределение, достаточными статистиками являются $\bar{Z}. = \frac{1}{n}\Sigma \log Y_j$ и $MS_Z = \Sigma(\log Y_j - \bar{Z}.)^2/(n-1)$ — оценки среднего и дисперсии величины Z. Отсюда приближенно несмещенной оценкой для $E(Y)$ является $\exp\left\{\bar{Z}. + \frac{1}{2}\mathrm{MS}_Z\right\}$. Рассмотрите метод устранения смещения в применении к этой задаче.

Решение

Отметим два простых, но важных результата: если Z и Q_d распределены соответственно нормально $N(\mu, \sigma^2)$ и хи-квадрат с d степенями свободы, то

$$E(e^{sZ}) = \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right), \quad E(e^{sQ}) = (1-2s)^{-\frac{1}{2}d}\text{ [1]}.$$

Отсюда если $Z = \log Y$ имеет нормальное $N(\lambda, \tau^2)$ распределение, то из производящей функции моментов получим

$$E(Y) = E(e^Z) = \exp\left(\lambda + \frac{1}{2}\tau^2\right). \tag{1}$$

Минимальной достаточной статистикой, построенной по $Y_1, \ldots, Y_n$, будет $(\bar{Z}., \mathrm{MS}_Z)$, причем $E(\bar{Z}.) = \lambda$ и $E\mathrm{MS}_Z = \tau^2$. Отсюда оценкой первого порядка для EY будет $T = \exp\left\{\bar{Z}. + \frac{1}{2}\mathrm{MS}_Z\right\}$.

Поскольку $\bar{Z}.$ и MS_Z независимы, то легко вычислить точное ожидание T. Действительно, $\bar{Z}.$ нормальна $N(\lambda, \tau^2/n)$, а $\mathrm{MS}_Z = \tau^2 Q_{n-1}/(n-1)$. Отсюда

$$E(e^{\bar{Z}.}) = \exp\left(\lambda + \frac{1}{2}\tau^2/n\right), \quad E\left(e^{\frac{1}{2}\mathrm{MS}_Z}\right) = \{1 - \tau^2/(n-1)\}^{-\frac{1}{2}(n-1)}.$$

Таким образом,

$$\begin{aligned} E(T) &= \exp\left[\lambda + \frac{1}{2}\tau^2/n - \frac{1}{2}(n-1)\log\{1 - \tau^2/(n-1)\}\right] = \\ &= \exp\left\{\lambda + \frac{1}{2}\tau^2 + \frac{1}{2}\tau^2/n + \frac{1}{4}\tau^4/n + O(1/n^2)\right\}. \end{aligned} \tag{2}$$

Другой подход связан с векторной формой разложения (1) задачи 8.12, эквивалентной разложению Тейлора статистики T как функции от $(\bar{Z}., \mathrm{MS}_Z)$ и (λ, τ^2).

---

[1] Отметим, что $E\left\{\exp\frac{1}{2}\mathrm{MS}_Z;\ \tau^2\right\}$ не существует при $\tau^2 \geqslant n-1$. Это необходимо учитывать при обосновании разложения (2) и в дальнейших выкладках. — *Прим. перев.*

Исследование представления (2) показывает, что смещение порядка n^{-1} устраняется при переходе к оценке

$$T \exp\left\{-\left(\frac{1}{2}\,\mathrm{MS}_Z + \frac{1}{4}\,\mathrm{MS}_Z^2\right)\Big/ n\right\}.$$

Хотя в качестве альтернативного метода устранения смещения можно было бы использовать метод расщепления выборки, однако общие соображения, связанные с достаточностью, позволяют предположить преимущество рассмотренного выше метода.

В принципе можно найти $E(Y_1 \mid \overline{Z}_., \mathrm{MS}_Z)$ — точную несмещенную оценку с минимальной дисперсией. Другой подход (см. работу Неймана и Скотта, 1960) связан с разложением в ряд

$$\exp\left(\lambda + \frac{1}{2}\,\tau^2\right) = \sum_{r,\,s=0}^{\infty} \frac{\lambda^r}{r!}\,\frac{\tau^{2s}}{2^s s!}$$

и последующим несмещенным оцениванием величин $\lambda^r \tau^{2s}$ через $\overline{Z}_.$ и MS_Z.

[Теоретическая статистика, § 8.4; Рао, § 6a]

8.14. Исследуйте свойства оценки по методу расщепления выборки (см. (2) раздела „Необходимые сведения") в применении к задаче оценивания параметра равномерного распределения на $(0, \theta)$. В качестве T_n возьмите максимум всех наблюдений в выборке. Сравните полученную оценку с оценкой, основанной на достаточной статистике. Покажите, что в рассматриваемом случае весьма плохо использовать точечную оценку в качестве суммарной характеристики доверительных областей.

Решение

Пусть $Y_{(1)} \leqslant \ldots \leqslant Y_{(n)}$ — вариационный ряд, отвечающий выборке $Y_1, \ldots, Y_n$. Тогда $T_n = Y_{(n)}$ со средним, равным $n\theta/(n+1)$. Таким образом, метод расщепления выборки для устранения смещения применим, хотя смещение проще полностью устранить, приняв в качестве оценки $(n+1)\,T_n/n$.

Когда опускается одно из наблюдений, то $(n-1)$ раз вновь будет получена величина T_n, а в оставшемся случае будет получена величина $Y_{(n-1)}$. Отсюда оценка по методу расщепления имеет вид

$$nY_{(n)} - \frac{(n-1)}{n}\,\{(n-1)\,Y_{(n)} + Y_{(n-1)}\} = \frac{2n-1}{n}\,Y_{(n)} - \frac{n-1}{n}\,Y_{(n-1)}$$

с ожиданием, равным $\theta[1 - \{n(n+1)\}^{-1}]$.

Метод расщепления выборки для получения оценки дисперсии T_n неприменим, поскольку T_n существенно нелинейна. В любом случае доверительные интервалы для θ вида $T_n \pm k\{\mathrm{var}(T_n)\}^{1/2}$ для рассматриваемой задачи мало подходят. Дело в том, что T_n — достоверная нижняя граница для θ, а правдоподобие для θ убывает при возрастании $\theta - T_n$. Таким образом, целесообразно доверительные интервалы искать в виде $[T_n, (1+c)T_n]$.

[Теоретическая статистика, § 8.4]

8.15. Докажите, что если при оценивании дисперсии исходить из смещенной оценки $\Sigma(Y_j - \bar{Y}_.)^2/n$, то оценка метода расщепления выборки (2) совсем устраняет смещение. Исследуйте целесообразность использования оценки (3) как оценки дисперсии.

Решение

Будем исходить из оценки $T_n = \Sigma(Y_j - \bar{Y}_.)^2/n$. Непосредственные вычисления просты, но довольно трудоемки. Их можно избежать, привлекая соображения, связанные с симметрией. Оценка по методу расщепления выборки T_n^{J} обязательно симметрична, квадратична и инвариантна относительно добавления постоянной ко всем наблюдениям. Таким образом,

$$T_n^{\mathrm{J}} = a_n \Sigma (Y_j - \bar{Y}_.)^2.$$

Чтобы определить постоянную a_n, рассмотрим какой-либо частный случай, например $Y_1 = 1$, $Y_j = 0$ $(j \geqslant 2)$. Тогда для этого случая $\Sigma(Y_j - \bar{Y}_0)^2 = (n-1)/n$, $T_n = (n-1)/n^2$ и $T_{n-1,j} = (n-2)/(n-1)^2$, за исключением $j = 1$, $T_{n-1,1} = 0$. Отсюда следует, что $a_n = 1/(n-1)$.

Если $Y_1, \ldots, Y_n$ — н. о. р. случайные величины с дисперсией σ^2 и коэффициентом эксцесса γ_2, то

$$\mathrm{var}\left\{\frac{\Sigma(Y_j - \bar{Y}_.)^2}{n-1}\right\} \sim \frac{2\sigma^4}{n}\left(1 + \frac{1}{2}\gamma_2\right). \tag{1}$$

Чтобы получить оценку дисперсии статистики T_n^{J} на основе метода расщепления, рассмотрим „псевдозначения“

$$T_j^{\mathrm{P}} = nT_n - (n-1)T_{n-1,j}$$

и оценим дисперсию оценки величины σ^2 как

$$V^{\mathrm{J}} = \Sigma(T_j^{\mathrm{P}} - T_n^{\mathrm{J}})^2/\{n(n-1)\}.$$

Соображения, связанные с учетом степени полинома V^{J}, симметрией и инвариантностью относительно сдвига Y_j на постоянную, показывают, что

$$V^{\mathrm{J}} = b_n \Sigma(Y_j - \bar{Y}_.)^4 + c_n\{\Sigma(Y_j - \bar{Y}_.)^2\}^2.$$

Постоянные b_n и c_n получают, вычисляя значения V^J для двух простых частных случаев. Получающееся значение V^J асимптотически согласуется с выражением (I).

Таким образом, в рассматриваемом случае оценивание дисперсии методом расщепления выборки не имеет очевидных преимуществ в сравнении с непосредственным использованием стандартных оценок для σ^2 и γ_2 при оценивании $\operatorname{var}(T_n)$.

Отметим также, что при оценивании $\operatorname{var}(T_n)$ посредством приближенно нормальных доверительных границ лучше работать с приближенно нормальным вариантом оценки T_n. Итак, в данном случае логарифм или кубический корень из выборочной дисперсии следует предпочесть самой выборочной дисперсии.

[Теоретическая статистика, § 8.4]

8.16. Наблюдения $y_{jk}(j=1, \ldots, m;\ k=1, \ldots, r)$ отвечают случайным величинам Y_{jk} с неизвестными параметрами положения μ_j и неизвестной постоянной дисперсией. Возможно, что значения параметров μ_j близки друг другу. Линейную комбинацию

$$\tilde{\mu}_j(a) = a\bar{y}_{j.} + (1-a)\bar{y}_{..}$$

рассматривают как подходящую оценку для μ_j. Покажите, что квадратичный критерий ошибки предсказания $\Sigma\Sigma\{y_{jk} - \tilde{\mu}_j(a)\}^2$ минимизируется, когда $a = 1$.

Обозначьте оценку $\tilde{\mu}_j(a)$, построенную по данным, в которых опущено наблюдение y_{jk}, как $\tilde{\mu}_{j,k}(a)$. Тогда ошибку предсказания для y_{jk} можно измерить посредством „псевдозначения“ $y_{jk} - \tilde{\mu}_{j,k}(a)$. Отсюда приходим к полному квадратичному критерию ошибки $\Sigma\Sigma\{y_{jk} - \tilde{\mu}_{j,k}(a)\}^2$. Покажите, что значение a, которое минимизирует этот критерий, является монотонно возрастающей функцией от $\mathrm{MS}_b/\mathrm{MS}_w$. Что вы можете сказать относительно явной формы окончательной оценки для μ_j и о методе вывода этой оценки?

Решение

Для первой части задачи имеем

$$\Sigma\Sigma\{y_{jk} - \tilde{\mu}_j(a)\}^2 = \Sigma\Sigma\{(y_{jk} - \bar{y}_{j.}) + (1-a)(\bar{y}_{j.} - \bar{y}_{..})\}^2 =$$
$$= \Sigma\Sigma(y_{jk} - \bar{y}_{j.})^2 + r(1-a)^2\Sigma(\bar{y}_{j.} - \bar{y}_{..})^2.$$

Второе положительное слагаемое пропадает, когда $a = 1$. В этом случае получаем обычную линейную оценку $\bar{y}_{j.}$.

Если μ_j взять в качестве наилучшего прогноза значения Y_{jk}, то представляется целесообразным измерить, насколько хорошо оценки для μ_j предсказывают независимые дополнительные наблюдения. Это приводит естественным образом к сравнению $\tilde{\mu}_{j.k}(a)$ с y_{jk} и к симметрическому сбалансированному критерию

$$S(a) = \Sigma\Sigma\{y_{jk} - \tilde{\mu}_{j.k}(a)\}^2.$$

Величины $\tilde{\mu}_{j.k}(a)$ можно выразить следующим образом:

$$\frac{a}{r-1}(r\bar{y}_{j.} - y_{jk}) + \frac{1-a}{rm-1}(rm\bar{y}_{..} - y_{jk}).$$

Отсюда вытекает, что

$$S(a) = \left\{\frac{ar}{r-1} + \frac{(1-a)\,rm}{rm-1}\right\}^2 \mathrm{SS}_w + \frac{(1-a)^2\,(rm)^2}{(rm-1)^2}\mathrm{SS}_b,$$

где $\mathrm{SS}_w = \Sigma\Sigma(y_{jk} - \bar{y}_{j.})^2$, $\mathrm{SS}_b = \Sigma(\bar{y}_{j.} - \bar{y}_{..})^2$. Квадратичная форма $S(a)$ минимизируется при

$$a^* = \frac{F-1}{F + \dfrac{m-1}{m(r-1)}},$$

где

$$F = \left\{\frac{r\mathrm{SS}_b}{(m-1)}\right\} \Big/ \left\{\frac{\mathrm{SS}_w}{m(r-1)}\right\}.$$

Когда $F = 1$, это обычно рассматривается как свидетельство приближенного равенства средних, $a^* = 0$, и все оценки равны величине $\bar{y}_{..}$. Когда F весьма велико, что свидетельствует о принятии величинами μ_j различных значений, $a^* \simeq 1$, и тогда μ_j оцениваются через индивидуализированные средние $\bar{y}_{j.}$. Эти результаты аналогичны полученным при эмпирическом байесовском подходе, когда μ_j имеют общее нормальное распределение с неизвестными параметрами.

Хотя представляется, что рассмотренный метод почти не включает формальных предположений, однако неявно предполагается: (i) наличие аддитивной модели для μ_j и (ii) однородная изменчивость для y_{jk}. Тем не менее общие методы внутренних сравнений, используемые с необходимой осторожностью, могут оказаться весьма информативными как вспомогательное средство при принятии модели и оценивания степени адекватности модели. Подробное обсуждение этих методов проводится в работе Стоуна (1974, 1977).

[Теоретическая статистика,
§ 8.4, 8.5]

8.17. Рассмотрим стационарный пуассоновский процесс с интенсивностью ρ. Предположим, что по наблюдению траектории процесса до момента t желательно оценить величину

$\theta = \exp\{-\rho x\}$. Если $N(t)$ — число событий в $(0, t)$, тогда „естественной" оценкой для θ будет

$$\bar{\theta} = \exp\{-N(t)\, x/t\}.$$

Разобьем теперь временной интервал на r подынтервалов равной длины t/r. Вычислим для θ оценку по методу расщепления выборки $\tilde{\theta}_r$, которая получается последовательным исключением одного из подынтервалов. Покажите, что когда r возрастает, а t фиксировано, то

$$\tilde{\theta}_r \to \bar{\theta}\{1 - N(t)(e^{x/t} - 1 - x/t)\},$$

т. е. при очень мелком разбиении интервала $(0, t)$ оценка по методу расщепления $\tilde{\theta}_r$ эквивалентна оценке $\bar{\theta}$ с поправкой на смещение до первого порядка.

Решение

Пусть число интервалов r велико, так что в любом интервале осуществляется самое большее одно событие. Тогда при пропуске j-го интервала получим оценку вида

$$\exp\{-(n - \delta_j)\, x/(t - t/r)\},$$

где $N(t)$ из δ_j равны единице, а все остальные δ_j равны нулю. Таким образом, оценкой по методу расщепления будет

$$\exp\left\{-\frac{N(t)\, xr}{(r-1)\, t}\right\}\left[1 - \frac{N(t)}{r} + \frac{N(t)}{r}\exp\left\{\frac{rx}{(r-1)\, t}\right\}\right].$$

Она приближенно равна

$$\exp\left\{-\frac{N(t)\, x}{t}\right\}\{1 - N(t)(e^{x/t} - 1 - x/t)\}.$$

Более простой подход связан с использованием разложения (1) из задачи 8.12. При этом полагают $T = N(t)/t$, $\theta = \rho$, $g(T) = e^{-Tx}$. Итак, поскольку $N(t)$ имеет пуассоновское распределение со средним и дисперсией, равными ρt, то

$$E(\bar{\theta}) \sim e^{-\rho x} + \frac{1}{2}\rho t^{-1} x^2 e^{-\rho x}.$$

Подставляя $N(t)/t$ вместо ρ, придем к оценке с уменьшенным смещением:

$$\hat{\theta} = \left\{1 - \frac{1}{2} N(t)\, x^2/t^2\right\}.$$

Она эквивалентна оценке по методу расщепления выборки, когда для $e^{x/t}$ используются три члена разложения.

[Теоретическая статистика,
§ 8.4; Гейвер в Хоуэл, 1970]

8.18. Пусть T_n — оценка по полной выборке из n наблюдений, а $T'_{\frac{1}{2}n}$ и $T''_{\frac{1}{2}n}$ обозначают соответствующие оценки, построенные по двум равным частям данных, когда данные случайно разделяются на две равные по объему совокупности. Если для случайной выборки объема n

$$E(T_m;\ \theta) = \theta + a_1(\theta)/m + a_2(\theta)/m^2 + \ldots,$$

то покажите, что оценка

$$2T_n - \frac{1}{2}\left(T'_{\frac{1}{2}n} + T''_{\frac{1}{2}n}\right)$$

имеет смещение, по порядку меньшее, чем смещение оценки T_n. Обсудите критически метод разделения данных.

Решение

Требуемый результат вытекает непосредственно из того факта, что обе оценки, и $T'_{\frac{1}{2}n}$, и $T''_{\frac{1}{2}n}$, есть просто оценки, построенные по случайным выборкам объема $\frac{1}{2}n$.

Заметим, что конкретные неслучайные методы разделения данных могут и не привести к такому же результату. Например, если $T_n = \overline{Y}$, и если упорядоченная выборка $Y_{(1)} \leqslant \ldots \leqslant Y_{(n)}$ разбивается на совокупности

$$\left\{Y_{(2j-1)};\ j = 1, \ \ldots, \ \frac{1}{2}n\right\}, \quad \left\{Y_{(2j)};\ j = 1, \ \ldots, \ \frac{1}{2}\right\},$$

тогда средние по этим совокупностям являются смещенными оценками для $E(Y)$.

[Теоретическая статистика,
§ 8.4; Кенуй, 1949, 1956]

8.19. Линейную модель $E(Y) = \mathbf{x}\beta$ называют квадратически сбалансированной, если все диагональные элементы матрицы $\mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T$ равны. Пусть ошибки являются независимыми с постоянной дисперсией и постоянным коэффициентом эксцесса γ_2. Докажите, что если модель квадратически сбалансирована либо $\gamma_2 = 0$, то из всех квадратичных форм, которые являются несмещенными оценками дисперсии ошибки, обычная остаточная сумма квадратов имеет минимальную дисперсию.

Решение

Начнем исследование с общей квадратичной формы $Q = Y^{\mathsf T}\mathbf{a}Y$, для которой

$$E(Q) = E\{(Y - \mathbf{x}\beta)^{\mathsf T}\mathbf{a}(Y - \mathbf{x}\beta)\} + \beta^{\mathsf T}\mathbf{x}^{\mathsf T}\mathbf{a}\mathbf{x}\beta = \sigma^2 \operatorname{tr}(\mathbf{a}) + \beta^{\mathsf T}\mathbf{x}^{\mathsf T}\mathbf{a}\mathbf{x}\beta.$$

Оценка Q не смещена в том случае, если $\operatorname{tr}(\mathbf{a}) = 1$ и $\mathbf{x}^{\mathsf T}\mathbf{a}\mathbf{x} = 0$, т. е. если $\mathbf{a}\mathbf{x} = 0$.

Когда ранг матрицы $\mathbf{x}$ равен p, оценка $MS_{\text{ост}}$ — остаточная средняя сумма квадратов — соответствует

$$\mathbf{a}_{\text{ост}} = \frac{1}{n-p}\{\mathbf{I} - \mathbf{x}(\mathbf{x}^{\mathsf T}\mathbf{x})^{-1}\mathbf{x}^{\mathsf T}\}.$$

Для матрицы $\mathbf{a}_{\text{ост}}$ условие несмещенности выполнено.

В общем случае дисперсию квадратичной формы Q можно выразить в виде

$$2\sigma^4\left\{\operatorname{tr}(\mathbf{a}^2) + \frac{1}{2}\gamma_2\mathbf{d}^{\mathsf T}\mathbf{d}\right\} + 4\sigma^2\beta^{\mathsf T}\mathbf{x}^{\mathsf T}\mathbf{a}^2\mathbf{x}\beta + 4\gamma_1\sigma^3\beta^{\mathsf T}\mathbf{x}^{\mathsf T}\mathbf{a}\mathbf{d},$$

где $\mathbf{d} = \operatorname{diag}(\mathbf{a})$, а γ_1 и γ_2 — коэффициенты асимметрии и эксцесса компонент вектора Y. Требование несмещенности оценки Q означает, что $\mathbf{a}\mathbf{x} = 0$, а следовательно, два последних слагаемых в выражении для дисперсии величины Q равны нулю и дисперсия принимает вид $2\sigma^4\left\{\operatorname{tr}(\mathbf{a}^2) + \frac{1}{2}\gamma_2\mathbf{d}^{\mathsf T}\mathbf{d}\right\}$. Теперь можно написать $\mathbf{a} = \mathbf{a}_{\text{ост}} + \mathbf{b}$, заметив при этом, что $\mathbf{b}\mathbf{x} = 0$, $\operatorname{tr}(\mathbf{b}) = 0$, $\mathbf{a}_{\text{ост}}\mathbf{b} = \mathbf{b}/(n-p)$, а $\operatorname{diag}(\mathbf{a}) = \operatorname{diag}(\mathbf{a}_{\text{ост}}) + \operatorname{diag}(\mathbf{b})$. Отсюда вытекает, что $\operatorname{tr}(\mathbf{a}^2) = 1 + \operatorname{tr}(\mathbf{b}^2)$. Итак, дисперсия несмещенной оценки Q принимает вид

$$\operatorname{var}(MS_{\text{ост}}) + 2\sigma^4\left\{\Sigma\Sigma\, b_{ij}^2 + \gamma_2\left(\Sigma b_{ii}a_{\text{ост},ii} + \frac{1}{2}\Sigma b_{ii}^2\right)\right\}. \qquad (1)$$

Если $\gamma_2 = 0$, то оптимальность оценки $MS_{\text{ост}}$ становится очевидной. Далее, если диагональные элементы матрицы $\mathbf{x}(\mathbf{x}^{\mathsf T}\mathbf{x})^{-1}\mathbf{x}^{\mathsf T}$ равны, то каждый диагональный элемент матрицы $\mathbf{a}_{\text{ост}}$ равен n^{-1} и выражение (1) принимает вид

$$\operatorname{var}(MS_{\text{ост}}) + 2\sigma^4\left\{\left(1 + \frac{1}{2}\gamma_2\right)\Sigma b_{ii}^2 + \sum_{i \neq j} b_{ij}^2\right\},$$

поскольку $\operatorname{tr}(\mathbf{b}) = 0$. Оптимальность оценки $MS_{\text{ост}}$ вытекает теперь из того факта, что $1 + \frac{1}{2}\gamma_2 \geqslant 0$. При $E(Z) = 0$ это эквивалентно неравенству $E(Z^4) \geqslant \{E(Z^2)\}^2$.

[Теоретическая статистика, § 8.5; Рао, § 4а.5; Атикулла, 1962]

8.20. Докажите теорему Гаусса о том, что для линейной модели с некоррелированными ошибками, имеющими постоянную дисперсию, оценки метода наименьших квадратов имеют минимальную дисперсию среди всех линейных несмещенных оценок. Сравните это со свойством оптимальности оценок метода наименьших квадратов для независимых нормально распределенных оценок. Покажите, что средняя квадратическая ошибка смещенных оценок на всем параметрическом пространстве неограниченна. Поэтому требование несмещенности в теореме Гаусса можно было бы заменить на требование ограниченности средней квадратической ошибки.

Решение

Большинство вводных курсов теоретической статистики содержат теоремы Гаусса (или Гаусса — Маркова). По этой причине развернутого ответа здесь приведено не будет. Существуют различные методы доказательства этой теоремы. В частности, метод минимизации дисперсии при условии выполнения требований линейной несмещенности, вводимых с помощью множителей Лагранжа, матричные методы доказательства и геометрический метод доказательства. В последнем случае берется $l^{\mathrm{T}}Y$ — произвольная линейная комбинация наблюдений и рассуждают следующим образом:

(a) $l = l_{\mathbf{x}} + l_{\perp\,\mathbf{x}}$, где $l_{\mathbf{x}}$ и $l_{\perp\,\mathbf{x}}$ — векторы, соответственно принадлежащие и перпендикулярные пространству, натянутому на столбцы матрицы $\mathbf{x}$;

(b) $E\left(l_{\perp\,\mathbf{x}}^{\mathrm{T}}Y\right) = 0$;

(c) $\operatorname{cov}\left(l_{\mathbf{x}}^{\mathrm{T}}Y,\ l_{\perp\,\mathbf{x}}^{\mathrm{T}}Y\right) = 0$, отсюда $\operatorname{var}(l^{\mathrm{T}}Y) = \operatorname{var}\left(l_{\mathbf{x}}^{\mathrm{T}}Y\right) + \operatorname{var}\left(l_{\perp\,\mathbf{x}}Y\right)$;

(d) из пунктов (b) и (c) вытекает, что несмещенная оценка с минимальной дисперсией определяется некоторым вектором l, принадлежащим пространству, натянутому на вектор-столбцы матрицы $\mathbf{x}$;

(e) поскольку искомые векторы имеют вид $a^{\mathrm{T}}\mathbf{x}^{\mathrm{T}}$, то соответствующие линейные функции от Y являются линейными комбинациями правых частей уравнения наименьших квадратов;

(f) существует самое большее одна комбинация правых частей уравнений наименьших квадратов, оценивающая заданное β_r.

При сравнении этого результата с результатом, вытекающим из неравенства Крамера — Рао, следует отметить прежде всего, что в обоих случаях от оценок требуется несмещенность, а величина ошибки измеряется ожиданием квадратической ошибки. Первое требование является, конечно, довольно искусственным. В теореме Гаусса предполагается линейность оценок, но требование к типу распределения ошибок оказывается довольно слабым. В неравенстве Крамера — Рао предполагается нормальность ошибок, а линейность оптимальной оценки получается как следствие. В общем случае оптимальная оценка необязательно линейна.

Смещенные линейные оценки привносят в среднеквадратическую ошибку вклад, равный квадрату смещения, который можно выразить в виде квадратичной формы от β. Поскольку параметрическое пространство неограниченно, то неограничен и указанный вклад. Отсюда заключаем, что из ограниченности среднеквадратической ошибки вытекает несмещенность оценки.

[Теоретическая статистика, § 8.5; Рао, § 4а; Силвей, § 3.5; Барнард, 1963]

8.21. Две случайные величины X и Y таковы, что можно наблюдать маргинальное распределение величины Y и условное распределение величины Y при заданном X. Покажите, что эти два распределения задают условное распределение величины X при заданном Y в том и только том случае, если определенное интегральное уравнение Фредгольма имеет единственное решение. Покажите, что если условное распределение величины Y при заданном $X=x$ нормально со средним $\gamma+\beta x$ и постоянной дисперсией, то условное распределение величины X при заданном Y всегда определяется. Рассмотрите более подробно частный случай, когда маргинальное распределение величины Y тоже нормально. Обсудите потенциальную важность этих результатов в задаче оценивания значения $x^\dagger$, соответствующего новому наблюдению $Y^\dagger$ при заданных наблюдениях $(x_1, Y_1), \ldots, (x_n, Y_n)$ линейной нормальной модели $E(Y) = \gamma+\beta x$.

Решение

Совместная п.р.в. случайных величин X и Y выражается в следующих эквивалентных формах:

$$f_X(x) f_{Y|X}(y \mid x) \quad \text{и} \quad f_Y(y) f_{X|Y}(x \mid y),$$

откуда

$$f_{X|Y}(x \mid y) = f_X(x) f_{Y|X}(y \mid x)/f_Y(y),$$

причем в правой части равенства неизвестна лишь $f_X(x)$. Далее, поскольку интегрирование левой части приводит к единице, то будем иметь

$$\int f_X(x) f_{Y|X}(y \mid x)\, dx = f_Y(y).$$

Это не что иное, как общее уравнение Фредгольма

$$\int a(x)\,k(x,\ y)\,dx = b(y),$$

которое необходимо решить относительно $a(x)$. Если для $a(x) = f_X(x)$ можно получить единственное решение, то условная плотность $f_{X\mid Y}(x\mid y)$ определяется.

Предположим теперь, что $f_{Y\mid X}(y\mid x)$ имеет плотность нормального $N(\gamma+\beta x,\ 1)$ распределения. Подстановка ее в интегральное уравнение придает ему следующий вид

$$\int \left\{ f_X(x) \exp\left(-\frac{1}{2}\beta^2 x^2 - \gamma\beta x\right)\right\} e^{\beta y x}\,dx = (2\pi)^{\frac{1}{2}} e^{\frac{1}{2}y^2-\gamma y} f_Y(y).$$

Это уравнение представимо в форме

$$\int g(x)\,e^{sx}\,dx = h(s),$$

т. е. $h(s)$ — преобразование Лапласа функции $g(\cdot)$. Таким образом, $g(x)$ определяется как единственный прообраз для $h(\cdot)$.

Для частного случая нормальной плотности $f_Y(y)$ легко показать, что если $\beta \neq 0$, то $f_X(x)$ также является плотностью нормального распределения, т. е. X и Y имеют двумерное нормальное распределение. Чтобы рассмотреть более общий случай, предположим, что Y при заданном $X = x$ имеет нормальное $N(\gamma+\beta x,\ \sigma^2)$ распределение, а случайная величина Y нормально $N(\mu,\ \tau^2)$ распределена. Тогда поскольку $E(Y) = E\{E(Y\mid X)\}$, то будем иметь $\mu + \gamma + \beta E(X)$. Далее, поскольку $\operatorname{var}(Y) = E\{\operatorname{var}(Y\mid X)\} + \operatorname{var}\{E(Y\mid X)\}$, то имеем, что $\tau^2 = \sigma^2 + \beta^2 \operatorname{var}(X)$. Эти соотношения показывают, что случайная величина X нормально распределена со средним $(\mu-\gamma)/\beta$ и дисперсией $(\tau^2-\sigma^2)/\beta^2$, когда $\beta \neq 0$.

Если ковариация между X и Y равна δ, то

$$E(Y\mid X=x) = \text{const} + \frac{\delta}{\operatorname{var}(X)}x,\quad E(X\mid Y=y) = \text{const} + \frac{\delta}{\operatorname{var}(Y)}y,$$

откуда получаем, что $\delta = (\tau^2-\sigma^2)/\beta$. Отсюда

$$E(X\mid Y=y) = (\mu-\gamma)/\beta + (\tau^2-\sigma^2)(y-\mu)/(\beta\tau^2).$$

Дисперсия условного нормального распределения величины X при заданном $Y = y$ равна

$$\operatorname{var}(X\mid Y=y) = \operatorname{var}(X) - \operatorname{var}\{E(X\mid Y)\} = \sigma^2(\tau^2-\sigma^2)/(\beta^2\tau^2).$$

В задаче калибровки в ее наиболее простой форме имеются данные $(x_1,\ Y_1),\ \ldots,\ (x_n,\ Y_n)$, причем случайные величины Y_j независимы и условно нормально $N(\gamma+\beta x_j,\ \sigma^2)$ распределены. Задача состоит в том, чтобы оценить значение x' величины x, соответствующее будущему наблюдению $Y^\dagger$, имеющему нормальное $N(\gamma+\beta x^\dagger, \sigma^2)$ распределение. Если значения $x_1, \ldots, x_n$ образуют случайную выборку из той же самой популяции, что $x^\dagger$, тогда можно обоснованно говорить о п. р. в. $f_X(x)$ для случайной величины X. Вне зависимости от вида п. р. в. $f_X(x)$ можно, в принципе, для заданных $f_Y(y)$ и $f_X(x)$ получить $f_{X|Y}(x|y)$. Необходимо отметить, что условное ожидание величины X при заданном $Y=y$ является линейным по y только в том случае, если маргинальным распределением случайной величины Y будет нормальное распределение. На основе данных обычным образом оценивают γ, β и σ^2. Значения $(x_1, \ldots, x_n)$ и $(y_1, \ldots, y_n)$ можно использовать для оценивания $f_X(x)$ и $f_Y(y)$. При этом, возможно, придется использовать и непараметрические методы.

Конечно, если данные определенно говорят в пользу нормального маргинального распределения, тогда было бы целесообразно работать непосредственно с нормальной линейной регрессионной моделью величины X на Y и стоило бы использовать стандартные методы интервального оценивания для $E(X|Y=y^\dagger)$. Однако предположение, что $x_1, \ldots, x_n$ и $x^\dagger$ порождаются одним и тем же распределением, часто совершенно неподходящее.

[Теоретическая статистика,
§ 8.5; Таллис, 1969]

9. Асимптотическая теория

Необходимые сведения

В гл. 4—8 рассматривались критерии значимости, интервальные оценки и точечные оценки. Было показано, как получать процедуры, удовлетворяющие различным критериям оптимальности. Однако имеется много задач, в которых данные ранее рекомендации не проходят. Причиной этого могут быть как сложность нахождения соответствующих распределений, так и невозможность применить методы исключения мешающих параметров. Ниже будут рассмотрены приближенные процедуры, основанные на асимптотических свойствах функции правдоподобия.

Предположим, что θ имеет заданную размерность, обозначаемую $\dim(\theta)$, и что число наблюдений n неограниченно растет. Тогда при весьма общих условиях оценка максимального правдоподобия (о.м.п.) $\hat{\theta}$, максимизирующая логарифм правдоподобия $l_Y(\theta'; y)$, сходится по вероятности к θ. Это свойство оценки будет называться состоятельностью. Чтобы воспользоваться в дальнейшем результатами общей теории, необходимо потребовать, чтобы правдоподобие удовлетворяло некоторым довольно сильным условиям регулярности. В частности, θ должно быть внутренней точкой параметрического пространства, а общая информация $i_.(\theta) = \operatorname{var}\{U_.(\theta); \theta\}$ должна быть положительно определена, неограниченно возрастать с ростом n и должна совпадать со средним значением второй производной по θ от $l_Y(\theta; Y)$, умноженной на минус единицу. О. м. п. асимптотически эквивалентна состоятельному решению уравнения правдоподобия $U_.(\theta') = 0$, а разложение Тейлора разности $U_.(\hat{\theta}) - U_.(\theta)$ приводит к аппроксимации первого порядка

$$i_.(\theta)(\hat{\theta} - \theta) \sim U_.(\theta).$$

Если центральная предельная теорема применима к $U_.(\theta)$ как к сумме н.о.р. случайных величин, то отсюда можно вывести, что для больших n оценка $\hat{\theta}$ асимптотически нормально $N\{\theta, i_.^{-1}(\theta)\}$ распределена. Чтобы получить состоятельную нормальную аппроксимацию, вместо $i_.(\theta)$ можно взять со знаком минус наблюденное значение второй производной от $l_Y(\theta'; y)$ в точке $\theta' = \hat{\theta}$.

Соответствующее свойство правдоподобия заключается в том, что для значений θ', близких к θ,

$$\mathrm{lik}_Y(\theta'; Y) \sim \mathrm{lik}_Y(\hat{\theta}; Y) \exp\left\{-\frac{1}{2}(\hat{\theta}-\theta')^{\mathrm{T}}\, \mathrm{i}.(\hat{\theta})(\hat{\theta}-\theta')\right\}. \quad (1)$$

Это соотношение указывает на приближенную достаточность о.м.п. для больших n. При этом можно использовать состоятельную аппроксимацию для $\mathrm{i}.(\theta)$.

Для моделей, зависящих от неограниченно большого числа параметров, полученные выше результаты в общем случае несправедливы. Однако для моделей, правдоподобие которых представимо в виде

$$\mathrm{lik}_Y(\theta; y) = \prod_{j=1}^{n} \mathrm{lik}_{Y_j \mid S_j}(\psi; y_j \mid s_j)\, \mathrm{lik}_{S_j}(\psi, \lambda_j; s_j),$$

изложенная теория применима для условного правдоподобия

$$\mathrm{lik}_{Y \mid S}(\psi; y) = \prod_{j=1}^{n} \mathrm{lik}_{Y_j \mid S_j}(\psi; y_j \mid s_j).$$

Для полиномиальных моделей с вероятностями исходов $\pi_j(\theta)$ и выборочными частостями $N_j\ (j=1, \ldots, m)$, $\sum N_j = n$ при минимизации, например, любой из следующих статистик:

$$\sum \{N_j - n\pi_j(\theta)\}^2/\{n\pi_j(\theta)\}, \quad \sum \{N_j - n\pi_j(\theta)\}^2/N_j$$

получают оценки, эквивалентные о. м. п.

Более точную нормальную аппроксимацию для $\hat{\theta}$ можно получить на основе разложений высоких порядков разности $U.(\hat{\theta}) - U.(\theta)$. Конечная информационная разность $i.(\theta) - i_{\hat{\theta}}(\theta)$ может быть вычислена, а тем самым получена соответствующая аппроксимация второго порядка дисперсии оценки $\hat{\theta}$.

Для регулярных задач асимптотическая теория критериев значимости (а следовательно, и интервального оценивания) строится на основе приведенного выше нормального представления (1) для правдоподобия. Для простой нулевой гипотезы H_0: $\theta = \theta_0$ при альтернативе H_A: $\theta \neq \theta_0$ подход на основе отношения максимума правдоподобия приводит к статистике критерия $W = 2\{l_Y(\hat{\theta}; Y) - l_Y(\theta_0; Y)\}$. Используя разложение Тейлора, или обращаясь к нормальной аппроксимации, будем иметь соответственно

$$W \sim W_e = (\hat{\theta} - \theta_0)^{\mathrm{T}}\, \mathrm{i}.(\theta)(\hat{\theta} - \theta_0)$$

и

$$W \sim W_u = U^{\mathrm{T}}(\theta_0)\, \mathrm{i}^{-1}(\theta_0)\, U.(\theta_0). \quad (2)$$

При H_0 каждая из статистик имеет асимптотически хи-квадрат распределение с $\dim(\theta)$ степенями свободы. Критерии, использующие критические области, основанные на этих статистиках, имеют локально оптимальные значения и мощности, что по существу является обобщением теоремы Неймана—Пирсона гл. 4. Привлекательность статистики W_u состоит в том, что нет необходимости вычислять $\hat{\theta}$. Статистики W и W_u, но не W_e, являются полностью инвариантными относительно репараметризации.

Наиболее часто встречается тип сложных гипотез, при котором требуется, чтобы значение θ принадлежало некоторому подпространству, которое в наиболее простом случае является линейным. Так, при $\theta = (\psi, \lambda)$ имеем H_0: $\psi = \psi_0$ при альтернативе H_A: $\psi \neq \psi_0$ и λ — мешающий параметр. Подход на основе максимума отношения правдоподобия приводит к статистике $W = 2\{l_Y(\hat{\psi}, \hat{\lambda}; Y) - l_Y(\psi_0, \hat{\lambda}_0, Y)\}$, где $\hat{\lambda}_0$ — о. м. п. параметра λ при условии $\psi = \psi_0$. Рассмотренные ранее методы разложения приводят к асимптотически эквивалентным статистикам

$$W_e = (\hat{\psi} - \psi_0)^{\mathrm{T}} \{i_{\cdot}^{\psi\psi}(\psi_0, \hat{\lambda}_0)\}^{-1} (\hat{\psi} - \psi_0)$$

и

$$W_u = U_{\cdot\psi}^{\mathrm{T}}(\psi_0, \hat{\lambda}_0)\, i_{\cdot}^{\psi\psi}(\psi_0, \hat{\lambda}_0)\, U_{\cdot\psi}(\psi_0, \hat{\lambda}_0). \tag{3}$$

При H_0 их асимптотическим распределением будет хи-квадрат распределение с $\dim(\psi)$ степенями свободы. Таким образом, критерии, определяемые этими статистиками, являются асимптотически подобными. Отметим, что W_u не требует вычисления о. м. п. $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$, а W и W_u, но не W_e, вновь инвариантны относительно репараметризации.

Соответствующие статистики для нелинейных теоретических ограничений на значение параметра θ описываются в задаче 10 данной главы.

Для гипотез, связанных с критерием согласия для полиномиальных моделей, статистики W_u и W_e приводят к хи-квадрат статистикам.

Асимптотическая локальная мощность статистик W, W_u и W_e определяется параметром нецентральности $(\psi - \psi_0)^{\mathrm{T}} \{i_{\cdot}^{\psi\psi}(\psi, \lambda)\}^{-1} (\psi - \psi_0)$ соответствующего нецентрального хи-квадрат распределения. Асимптотическая эффективность двух статистик, имеющих при нулевой гипотезе в качестве предельного хи-квадрат распределение с одной и той же степенью свободы, определяется как отношение соответствующих параметров нецентральности. В частности, для скалярного параметра θ и статистик критерия T_1, T_2, которые приближенно нормально $N(\mu_j(\theta), \sigma_j^2(\theta))$ распределены ($j = 1, 2$), питменовская асимптотическая относительная эффективность $e(T_1:T_2)$ равна отношению

$$[\{\mu_1'(\theta_0)\}^2/\sigma_1^2(\theta)]/[\{\mu_2'(\theta_0)\}^2/\sigma_2^2(\theta_0)].$$

Укажем два типа сложных гипотез, сильно отличающихся от рассмотренных выше:

(i) гипотеза H_0 образует подмножество параметрического пространства Ω, не являющееся подпространством, как, например, совокупность $|\theta| \leqslant c$;

(ii) гипотезы H_0 и H_A определяют разделенные семейства распределений.

Для гипотез типа (i) можно использовать асимптотические аппроксимации для правдоподобия, в то время как для гипотез типа (ii) это невозможно. Пусть в H_0 и H_A задаются отдельные регулярные семейства распределений. Чтобы вывести асимптотические свойства статистик критерия, предопределенные свойствами оценок максимального правдоподобия, можно снова раскладывать в ряд максимум логарифма отношения правдоподобия.

Большинство из рассмотренных ниже задач имеют отношение к развитию асимптотической теории и до некоторой степени к нестандартным приложениям. Необходимо подчеркнуть, что в приложениях многого можно достичь осторожным и критическим использованием нескольких простых результатов. Упомянем, в частности, результаты о выражении для асимптотической дисперсии оценок максимума правдоподобия и результаты о критериях, основанных на сравнении максимизированных логарифмических правдоподобий.

Задачи

9.2[1]. Пусть $Y_1, \ldots, Y_n$ — н. о. р. случайные величины, п. р. в. которых зависит от скалярного параметра θ. Предположим, что выполнены следующие условия регулярности:

(a) параметрическое пространство Ω является замкнутым подмножеством вещественной прямой, и истинное значение параметра является внутренней точкой пространства Ω;

(b) вероятностные распределения, отвечающие любым двум не совпадающим значениям параметра θ, различны;

(c) три первые производные логарифмического правдоподобия $l(\theta; Y)$ по θ существуют в окрестности истинного значения параметра почти всюду.

Далее, в упомянутой окрестности абсолютное значение третьей производной, умноженное на n^{-1}, ограничено сверху

[1]) Задача 9.1 опущена.

функцией от Y, ожидание которой существует;

$$\text{(d)}\quad i_{.}(\theta) = \operatorname{var}\{U_{.}(\theta);\ \theta\} = E\{-\partial U_{.}(\theta)/\partial\theta;\ \theta\}$$

— конечно и положительно в окрестности истинного значения параметра.

Покажите, что уравнение правдоподобия $U_{.}(\theta') = 0$ имеет состоятельное решение.

Решение

Отметим специально различие, существующее между утверждением в рассматриваемой задаче и состоятельностью о. м. п. В задаче утверждается на качественном уровне, что для весьма больших n уравнение правдоподобия будет иметь решение, близкое истинному значению параметра. При этом оставляется открытой возможность, что в отдалении от θ могут существовать решения, для которых правдоподобие достигает большее значение.

Подробное доказательство приводиться не будет. Будет достаточно показать, что при $n \to \infty$ для любых фиксированных $\delta > 0$

$$\operatorname{pr}\{U_{.}(\theta-\delta) > 0 > U_{.}(\theta+\delta);\ \theta\} \to 1.$$

Чтобы показать это, запишем

$$U_{.}(\theta') = U_{.}(\theta) + (\theta'-\theta)\,U'_{.}(\theta) + \frac{1}{2}(\theta'-\theta)^2\,U''_{.}(\bar{\theta}),\ .$$

где $\bar{\theta}$ лежит между θ и θ'. Заметим, что из слабого закона больших чисел

$$U_{.}(\theta)/n \to 0,\ \ U'_{.}(\theta)/n \to -i(\theta),\ \ U''_{.}(\theta)/n - M(\theta) \to 0$$

по вероятности. Здесь $M(\theta)$ — некоторая ограниченная функция, существование которой в последнем условии регулярности относительно третьей производной исходной плотности распределения. Отсюда легко показать, что обе вероятности

$$\operatorname{pr}\{U_{.}(\theta-\delta) < 0;\ \theta\}\ \text{ и }\ \operatorname{pr}\{U_{.}(\theta+\delta) > \theta;\ \theta\}$$

стремятся к нулю при $n \to \infty$.

В случае векторного параметра θ конечной размерности и при дополнительных условиях в случае зависимых и неодинаково распределенных случайных величин работает рассмотренный выше метод, только разложение ведется покомпонентно. Этот метод не проходит для компонент с плотностью

$\frac{1}{2}\exp(-|y-\theta|)$, поскольку необходимое разложение относительно θ сделать нельзя.

[Теоретическая статистика, § 9.2; Рао, § 5.7; Силвей, § 4.5; Крамер, 1975, стр.]

9.3. Пусть Y_{jk} $(k=1, \ldots, r_j;\ j=1, \ldots, m)$ независимы и нормально $N(\mu, \sigma_j^2)$ распределены. Здесь все параметры неизвестны. Минимальной достаточной статистикой является $(\bar{Y}_{1.}, \ldots, \bar{Y}_{m.}, SS_1, \ldots, SS_m)$, где, как обычно, $SS_j = \sum (Y_{jk} - \bar{Y}_{j.})^2$. Для того чтобы оценить μ, проверьте, что уравнение правдоподобия является частным случаем уравнения

$$\sum_{j=1}^{m} \frac{a_j(\bar{Y}_{j.} - \mu')}{SS_j + r_j(\bar{Y}_{j.} - \mu')^2} = 0.$$

Заметьте, что каждое слагаемое в левой части уравнения имеет нулевое ожидание, когда $\mu' = \mu$, т. е. равно истинному значению. Далее, предположим, что r_1, r_2 фиксированы, а m становится большим. Обобщая результат задачи 9.2, покажите, что решение $\bar{\mu}_a$ оценивающего уравнения состоятельно пир соответствующих условиях на a_j и r_j. Далее, покажите, что при этих условиях оценка $\bar{\mu}_a$ имеет предельное нормальное распределение, дисперсия которого при соответствующем выборе постоянных a_j может быть сделана равномерно меньше, чем дисперсия для предельного нормального распределения о. м. п.

Прокомментируйте результат и его связь с подобным критерием Бартлета проверки гипотезы H_0: $\mu = \mu_0$ (рассмотренных на стр. 166 Теоретической статистики).

Решение

Логарифм функции правдоподобия для $(\mu, \sigma_1^2, \ldots, \sigma_m^2)$ можно записать следующим образом:

$$\text{const} - \sum r_j \log \sigma_j - \frac{1}{2} \sum r_j (\bar{y}_{j.} - \mu)^2/\sigma_j^2 - \frac{1}{2} \sum SS_j/\sigma_j^2.$$

Отсюда получаем уравнения правдоподобия

$$\sum r_j (\bar{y}_{j.} - \hat{\mu})/\hat{\sigma}_j^2 = 0, \quad r_j\hat{\sigma}_j^2 = SS_j - r_j(\bar{y}_{j.} - \hat{\mu})^2,$$

приводящие к уравнению для $\hat{\mu}$

$$\sum \frac{r_j^2(\bar{y}_{j.} - \hat{\mu})}{SS_j + r_j(\bar{y}_{j.} - \hat{\mu})^2} = 0,$$

которое является частным случаем упомянутого при формулировке задачи уравнения с $a_j = r_j^2$.

Для простоты обозначим

$$B_j(\mu) = \frac{a_j(\overline{Y}_{j.} - \mu)}{SS_j + r_j(\overline{Y}_{j.} - \mu)^2}.$$

Отсюда общее уравнение оценивания принимает вид $\sum B_j(\tilde{\mu}_a) = 0$. Доказательство состоятельности решения проводится по аналогии с доказательством, приведенном при решении задачи 9.2.

Для того чтобы применить закон больших чисел к $\sum B_j(\mu)/m$, необходимо показать, что

$$E\{B_j(\mu);\ \theta\} = 0,\ \operatorname{var}\{m^{-1} \sum B_j(\mu);\ \theta\} \to 0,$$

где $\theta = (\mu, \sigma_1^2, \ldots)$. В дополнение будут использоваться условия регулярности типа

$$m^{-1} \sum B_j'(\mu) = K + o_p(1),\ \ K \neq 0,$$

где $o_p(1)$ обозначают случайную величину, сходящуюся по вероятности к нулю. Другое условие состоит в том, что величина $m^{-1} |\sum B_j''(\mu)|$ ограничена функцией, не зависящей от θ и имеющей ограниченное среднее. Эти условия аналогичны условиям для $U(\theta)$, использованным в связи с исследованием свойств о. м. п.

Разложение состоятельного оценивающего уравнения в ряд Тейлора приводит к следующему выражению:

$$m^{-1} \sum B_j(\tilde{\mu}_a) = m^{-1} \sum B_j(\mu) + (\tilde{\mu}_a - \mu)\{m^{-1} \sum B_j'(\mu) + o_p(1)\}.$$

Отсюда вытекает, что применение центральной предельной теоремы к $m^{-1} \sum B_j(\mu)$ и слабого закона больших чисел к сумме $m^{-1} \sum B_j'(\mu)$ позволяет получить для $\tilde{\mu}_a - \mu$ в качестве предельного нормальное распределение с нулевым средним и дисперсией

$$V_a = \frac{\operatorname{var}\{\sum B_j(\mu);\ \theta\}}{[E\{\sum B_j'(\mu);\ \theta\}]^2} = \frac{\sum a_j^2/\{r_j^2(r_j - 2)\sigma_j^2\}}{\{\sum a_j/(r_j\sigma_j^2)\}^2}.$$

Из ограниченности величины V_a вытекает, что $a_j = 0$, если $r_j = 1, 2$. Легко проверить, что V_a минимизируется при $a_j = r_j(r_j - 2)$. В этом случае V_a равномерно меньше значений V_a при $a_j = r_j^2$, соответствующих состоятельной о. м. п. Тщательное исследование показывает, что необходимые условия регулярности $B_j''(\mu)$ будут выполнены, если показать, что $|B_j''(\mu)| \leqslant 8a_j r_j^2 SS_j^{-3}$.

Итак, в этом конкретном случае наличие бесконечно большого числа мешающих параметров приводит к неэффективности о. м. п. Отметим, что подстановка некоторого конкретного значения μ_0 в оценивающее уравнение приводит к статистике критерия для проверки гипотезы H_0: $\mu=\mu_0$. Использование в оценивающем уравнении весов $a_j=r_j(r_j-2)$ в точности соответствует весам, которые использованы в статистике Бартлетта (1936) для проверки гипотезы H_0. Смысл такого выбора неясен. Если $r_j=1$ или $r_j=2$, то статистика $\bar{Y}_{j.}$ информации о μ в статистику критерия не вносит. Это оставляет открытой возможность для некоторых оценок, не принадлежащих рассмотренному здесь классу, получить улучшение свойств статистики критерия. Сказанное касается в первую очередь данных, в которых имеется значительная доля массивов наблюдений с $r_j=2$.

В приложениях, в которых дисперсия не стабильна, а средние меняются, важно рассмотреть возможность того, что дисперсия связана со средним. В других случаях для σ_j^2 можно использовать эмпирическую байесовскую модель. При соответствующих предположениях это приводит к обобщению рассмотренного здесь уравнения (см. задачу 10.12).

[Теоретическая статистика, § 5.2, 9.2; Рао, § 5 6а.2; Силвей, гл. 7; Нейман и Скотт, 1948; Бартлетт, 1936]

9.4. При обычных предположениях о линейной модели оценка регрессионного параметра β имеет ковариационную матрицу $(\mathbf{x}^T\mathbf{x})^{-1}\sigma^2$. Покажите, что наблюдаемый в произвольной точке $\bar{x}$ факторпространства отклик можно предсказать со среднеквадратической ошибкой $\sigma^2\{1+x^T(\mathbf{x}^T\mathbf{x})^{-1}\bar{x}\}$, где появление второго слагаемого связано с неопределенностью, возникающей из-за незнания параметра β.

Рассмотрите теперь более общую ситуацию. Пусть точка x из факторпространства, которой соответствует отклик Y_x, имеет п. р. в. $f_{Y_x}(y; x, \theta)$. Из некоторых данных параметр θ оценивается посредством метода максимума правдоподобия с асимптотической ковариационной матрицей $\mathbf{i}_.^{-1}(\theta)$. Требуется измерить неопределенность, возникающую в предсказании отклика для произвольной точки $\bar{x}$ факторпространства и приписываемую отсутствию сведений о параметре θ. Как обосновать измерение неопределенности посредством $E\{-\log f_{Y_x}\times$ $\times(Y; \bar{x}, \theta); \theta\}$ в случае известного θ? Покажите, что когда параметр θ оценивается, то разумной мерой общей неопределенности будет $E\{-\log f_{Y_x}(Y; x, \theta); \theta\}$. Определите компоненту неопределенности, приписываемой отсутствию сведений о θ, посредством разности этих величин. С помощью обычных асимптотических аппроксимаций покажите, что эта разность равна $\frac{1}{2}\operatorname{tr}\{\mathbf{i}(\theta;\tilde{x})\,\mathbf{i}^{-1}(\theta)\}$, где $\mathbf{i}(\theta;\tilde{x})$ — информационная матрица, вычисленная по одному наблюдению в точке x. Проверьте, что все это приводит к результатам, совпадающим с полученными выше для линейной модели.

Решение

Для наблюденного вектора $Y=\mathbf{x}\beta+\varepsilon$ оценка наименьших квадратов $b=(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^TY$ имеет ковариационную матрицу $(\mathbf{x}^T\mathbf{x})^{-1}\sigma^2$. Отсюда получаем, что $\hat{Y}=\tilde{x}^Tb$ имеет дисперсию $\tilde{x}^T\operatorname{var}(b)\tilde{x}=\tilde{x}^T(\mathbf{x}^T\mathbf{x})^{-1}\tilde{x}\sigma^2$ и является несмещенной оценкой среднего в модели $\tilde{Y}=\tilde{x}^T\beta+\tilde{\varepsilon}$. Поскольку $\tilde{Y}$ и $\hat{Y}$ независимы,

$$E(\tilde{Y}-\hat{Y})^2=\operatorname{var}(\tilde{Y})+\operatorname{var}(\hat{Y})=\sigma^2\{1+\tilde{x}^T(\mathbf{x}^T\mathbf{x})^{-1}\tilde{x}\}.$$

В нелинейной задаче, когда Y_x имеет п. р. в. $f_{Y_x}(y;x,\theta)$, мера

$$E\{-\log f_{Y_x}(Y;x,\theta);\theta\}$$

есть не что иное, как энтропия — понятие, происходящее из теории информации. Энтропия аддитивна при объединении независимых наблюдений. Она равна нулю в том и только том случае, когда Y_x постоянна; для сдвигово-масштабного семейства энтропия пропорциональна параметру масштаба. Далее, она обладает свойством, что на ней минимизируется перекрестная энтропия или информационное расстояние $E\{-\log f_{Y_x}(Y;x,\theta');\theta\}$ (см. задачу 4.16). Это последнее замечание связано также с использованием величины $K(\hat{\theta},\theta;x)=E\{-\log f_{Y_x}(Y;x,\hat{\theta});\theta\}$ как общей меры неопределенности: эта величина также аддитивна и мала, когда оценка $\hat{\theta}$ близка к значению θ.

Положительность разности $K(\hat{\theta},\theta;\tilde{x})-K(\theta,\theta;\tilde{x})$ можно показать следующим образом:

$$E_{\hat{\theta}}E_{\tilde{Y}}\{-\log f_{\tilde{Y}_x}(\tilde{Y};\tilde{x},\hat{\theta})+\log f_{Y_x}(\tilde{Y};\tilde{x},\theta);\theta\}=$$
$$=E_{\hat{\theta}}E_{\tilde{Y}}[-\log f_{Y_x}(\tilde{Y};\tilde{x},\theta)+(\hat{\theta}-\theta)^TU(\theta;\tilde{x})-$$
$$-\frac{1}{2}(\hat{\theta}-\theta)^T\{\partial U(\theta;\tilde{x})/\partial\theta\}(\hat{\theta}-\theta)+o_p(n^{-1})+\log f_{Y_{\tilde{x}}}(\tilde{Y};\tilde{x},\theta);\theta]=$$
$$=-\frac{1}{2}E_{\hat{\theta}}E_{\tilde{Y}}[\operatorname{tr}\{(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T\partial U(\theta;\tilde{x})/\partial\theta\}]+o(n^{-1}).$$

Из независимости величины $\tilde{Y}$ и оценки $\hat{\theta}$ вытекает, что асимптотически правая часть полученного соотношения равна

$$\frac{1}{2}\operatorname{tr}\{\mathbf{i}_{\cdot}^{-1}(\theta)\,\mathbf{i}(\theta;\ \tilde{x})\}.$$

В случае линейной модели результаты будут теми же самыми или другими в зависимости от того, является ли $\sigma^2 = \operatorname{var}(Y)$ компонентой параметра θ или не является. Дело в том, что информационные матрицы имеют диагональную блочную структуру. Легко проверить, что если оценка θ вычислена для модели $Y = \mathbf{x}\theta + \varepsilon$, а $\tilde{Y}_{\tilde{x}} = \tilde{x}^{\mathrm{T}}\theta + \tilde{\varepsilon}$, то $\mathbf{i}_{\cdot}(\theta) = (\mathbf{x}^{\mathrm{T}}\mathbf{x})\,\sigma^2$ и $\mathbf{i}(\theta;\ \tilde{x}) = \tilde{x}\tilde{x}^{\mathrm{T}}\sigma^2$, а потому

$$\operatorname{tr}\{\mathbf{i}^{-1}(\theta)\,\mathbf{i}(\theta;\ \tilde{x})\} = \operatorname{tr}\{\tilde{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\tilde{x}\}.$$

Здесь использовано равенство $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CAB})$. Отметим, что рассмотренная мера не зависит от σ^2 и является, таким образом, в точности относительной мерой, основанной на средних квадратах.

[Теоретическая статистика, § 9.2; Рао, § 5; Силвей, гл. 4; Уайт, 1973]

9.5. Пусть $Y_1, \ldots, Y_n$ независимы и Y_j имеет п. р. в. $f(y;\ \psi,\ \lambda_j)$. Предположим, что S_j — минимально достаточна для λ_j при фиксированном ψ. Причем в этом случае S_j функционально от ψ не зависит. Это задача с несущественными параметрами. Покажите, что если ψ и λ_j скалярны, то нижняя граница Крамера — Рао для дисперсий несмещенных оценок параметра ψ с $\theta = (\psi, \lambda_1, \ldots, \lambda_n)$ имеет вид

$$\{i_{\cdot\psi\psi}(\theta) - \sum i_{\cdot\psi j}^2(\theta)\, i_{\cdot jj}^{-1}(\theta)\}^{-1},$$

где

$$i_{\cdot\psi\psi}(\theta) = E\left\{-\frac{\partial^2 l(\theta;\ Y)}{\partial\psi^2};\ \theta\right\},\quad i_{\cdot\psi j}(\theta) = E\left\{-\frac{\partial^2 l(\theta;\ Y)}{\partial\psi\,\partial\lambda_j};\ \theta\right\}$$

и

$$i_{\cdot jj}(\theta) = E\left\{-\frac{\partial^2 l(\theta;\ Y)}{\partial\lambda_j^2};\ \theta\right\}\quad (j = 1, \ldots, n).$$

Можно показать, что эта нижняя граница асимптотически достигается для условной о. м. п. параметра ψ в том случае, если п. р. в. принадлежит экспоненциальному семейству, а S_j таковы, что их распределения зависят от λ_j, но не от ψ, а условное распределение величины Y_j при заданном S_j зависит только от ψ.

Исследуйте асимптотическую эффективность условной о. м. п. параметра ψ для случаев, когда Y_j — двумерные нормальные случайные векторы со средними λ_j и $\lambda_j+\psi$ и единичной ковариационной матрицей.

Решение

Поскольку Y_j независимы, то эффективный вклад можно записать следующим образом:

$$U_{\cdot}(\psi, \lambda_1, \ldots, \lambda_n)=\sum_{j=1}^{n} U_j(\psi, \lambda_j),$$

где

$$U_j^{\mathrm{T}}=(U_{j\circ}, 0, \ldots, U_{jj}, 0, \ldots, 0) \quad (j=1, \ldots, n),$$

поскольку от λ_j зависит только Y_j и не зависят другие Y_k. Отсюда без труда получаем, что

$$\mathbf{i}_{\cdot}(\theta)=\begin{bmatrix} a & b^{\mathrm{T}} \\ b & \mathbf{d} \end{bmatrix},$$

где

$$a=\sum i_{j,\,00}(\theta), \quad b^{\mathrm{T}}=\{i_{1,\,01}(\theta), \ldots, i_{n,\,0n}(\theta)\},$$
$$d=\operatorname{diag}\{i_{1,\,11}(\theta), \ldots, i_{n,\,nn}(\theta)\}.$$

Нижняя граница Крамера — Рао для ψ, приведенная в разделе "Необходимые сведения" гл. 8, в рассматриваемом случае принимает вид $i^{00}(\theta)=(a-b^{\mathrm{T}}\mathbf{d}^{-1}b)^{-1}$, тем самым получен искомый результат.

Заметим далее, что если S_j — подчиненные статистики относительно ψ в обобщенном смысле, т. е.

$$f_{Y_j}(y_j; \theta)=f_{S_j}(S_j; \lambda_j) f_{Y_j \mid S_j}(y_j \mid s_j; \psi),$$

то условные о. м. п. совпадают с безусловными о. м. п. Случай, когда S_j не только подчинена, но и достаточна для λ_j при фиксированном ψ, более интересен, чем рассмотренный выше. Итак,

$$f_{Y_j}(y_j; \theta)=f_{S_j}(s_j; \theta) f_{Y_j \mid S_j}(y_j \mid s_j; \psi).$$

Используя очевидные обозначения, получим

$$i^{Y}_{.00}(\theta)-\sum\{i^{Y}_{.jj}(\theta)\}^{-1}\{i^{Y}_{.0j}(\theta)\}^2=i^{Y\mid S}_{00}+i^{S}_{00}-\sum(i^{S}_{.jj})^{-1}(i^{S}_{.0j})^2.$$

Таким образом, условные о. м. н., полученные на основе Y при заданном значении $S=s$, будут асимптотически нормальны, если

$$i^{S}_{.00}(\theta)-\sum\{i^{S}_{jj}(\theta)\}^{-1}\{i^{S}_{0j}(\theta)\}^2=0. \qquad (1)$$

На самом деле, нуль можно заменить на $o(i^{\psi\,|\,S}_{.00})$. При некоторых условиях на λ можно получить здесь и более общий результат. Заметим, что условие (1) для $i^S(\theta)$ эквивалентно тому, что вклад $U_{.j}(\theta)$ является линейной функцией от $U_{.j}(\theta)$. Таким образом, в предположении независимости S_j, необходимое и достаточное условие состоит в том, что $U^S_{j0}(\theta)$ — линейная функция от $U^S_{j1}(\theta)$, т. е.

$$\frac{\partial}{\partial\psi}\log f_{S_j}(s_j;\theta)=\gamma_j(\theta)\frac{\partial}{\partial\lambda_j}\log f_{S_j}(s_j;\theta)\quad(j=1,\ \ldots,\ n),\qquad(2)$$

где $\gamma_j(\theta)$ не зависят от s_j. Этот результат является главным в проведенном Андерсеном (1970) обстоятельном анализе задачи.

В частном случае $Y_{j1}+Y_{j2}=S_j$ достаточна для λ_j при заданном ψ. Условное правдоподобие при заданных S_j эквивалентно, конечно, правдоподобию величин $\{Y_{j2}-Y_{j1},\ j=1,\ \ldots,\ n\}$, сводящемуся к правдоподобию на основе $\bar{\psi}^* =\bar{Y}_{.2}-\bar{Y}_{.1}$, которое на самом деле эквивалентно правдоподобию на основе $\bar{\psi}$. Поскольку S_j нормально $N(\psi+2\lambda_j,\ 2)$ распределены, то отсюда незамедлительно следует, что равенство (2) выполняется с $\gamma_j(\theta)=\frac{1}{2}$. Таким образом, асимптотическая эквивалентность проверяется общим методом.

[Теоретическая статистика, § 9.2; Рао, § 5а; Силвей, гл. 4]

9.6. Пусть $Y_1,\ \ldots,\ Y_n$ — последовательные состояния стационарной марковской цепи с переходными вероятностями $\mathrm{pr}(Y_j=1\,|\,Y_{j-1}=1)=\lambda$ и стационарными вероятностями $\psi=\mathrm{pr}(Y_j=1)$. Покажите, что дисперсия предельного нормального распределения для о. м. п. $\hat{\psi}$ имеет вид $\psi(1-\psi)\times\times(1-2\psi+\lambda)/\{n(1-\lambda)\}$. Последнее выражение является также дисперсией среднего $\bar{Y}_.$. Исходя из этого, проверьте, что если в явное выражение о. м. п. параметра λ при фиксированном ψ вместо ψ подставить $\bar{Y}_.$, то оценка останется асимптотически эффективной.

Решение

Первый шаг состоит в том, чтобы получить равновесное распределение марковской цепи с двумя состояниями и переходными вероятностями $p_{st}=\mathrm{pr}(Y_{j+1}=t\,|\,Y_j=s)$, т. е. необходимо показать, что в новой параметризации

$$p_{00}=(1-2\psi+\lambda\psi)/(1-\psi)$$

и, конечно, $p_{10}=1-\lambda$ и $p_{11}=\lambda$. Для $0<\lambda$, $\psi<1$ условия регулярности для состоятельности и асимптотической нормальности оценки $\hat{\theta}=(\hat{\psi}, \hat{\lambda})$ выполняются.

Далее,

$$U_j^T(\theta)=\left(\frac{\partial}{\partial\psi}, \frac{\partial}{\partial\lambda}\right)\log f(Y_j \mid Y_{j-1}; \theta),$$

$$i_.(\theta)=\sum_{j=2}^{n} \operatorname{var}\{U_j(\theta); \theta\}+O(1)\sim n \operatorname{var}\{U_2(\theta); \theta\}.$$

Четыре возможных значения каждой компоненты величин $U_j(\theta)$ учитываются при дифференцировании элементов переходной матрицы. Вычислив дисперсии, получим информационную матрицу для одного наблюдения:

$$\begin{bmatrix} \dfrac{(1-\lambda)^2}{(1-\psi)^4}\left(\dfrac{1}{p_{00}}+\dfrac{1}{p_{11}}\right) & \dfrac{(1-\lambda)\psi}{(1-\psi)^3}\left(\dfrac{1}{p_{00}}+\dfrac{1}{p_{01}}\right) \\ & \dfrac{\psi^2}{(1-\psi)^2}\left(\dfrac{1}{p_{00}}+\dfrac{1}{p_{01}}\right)+\dfrac{1}{p_{10}}+\dfrac{1}{p_{11}} \end{bmatrix}.$$

Умножив ее на n и затем обратив, получим, в частности, что главный член асимптотической дисперсии оценки $\hat{\psi}$ имеет вид

$$\psi(1-\psi)(1-2\psi+\lambda)(1-\lambda)^{-1}n^{-1}.$$

Далее, ясно, что $E(\bar{Y}_.)=\psi$ и что центральная предельная теорема к $\bar{Y}_.$ применима. Найдем дисперсию для $\bar{Y}_.$. Будем опираться при этом на метод, применимый к среднему любого стационарного процесса:

$$\operatorname{var}(\bar{Y}_.)=\frac{1}{n^2}\sum \operatorname{var}(Y_j)+\frac{1}{n^2}\sum_{j\neq k}\operatorname{cov}(Y_j, Y_k)=$$

$$=\frac{\sigma^2}{n}+\frac{2\sigma^2}{n^2}\sum_{h=1}^{n-1}(n-h)\gamma_h,$$

где $\gamma_h=\operatorname{cov}(Y_j, Y_{j+h})$. Для нашего конкретного процесса

$$\sigma^2=\psi(1-\psi),\ \gamma_h=\operatorname{pr}(Y_j=Y_{j+h}=1)-\psi^2.$$

Далее,

$$\operatorname{pr}(Y_j=Y_{j+h}=1)=\operatorname{pr}(Y_j=1)\operatorname{pr}(Y_{j+h}=1 \mid Y_j=1)=$$
$$=\psi\{\psi+(1-\psi)(\lambda-\psi)^h/(1-\psi)^h\}.$$

Последнее равенство получается при вычислении переходной матрицы цепи на h шагов. Используя полученный результат, полностью определяем вид $\operatorname{var}(\bar{Y}_.)$.

Итак, показано, что $\hat{\psi}$ и $\bar{Y}_.$ асимптотически нормальны с одними и теми же средними и дисперсиями. Отсюда можно строго показать, что $\overline{Y}. = \tilde{\psi} + o_p\left(n^{-\frac{1}{2}}\right)$. Не строго, это вытекает из задачи 9.9, поскольку вектор $(\overline{Y}., \tilde{\psi})$ в качестве предельного имеет вырожденное нормальное распределение с ковариационной матрицей $\mathbf{i}.(\theta) \otimes \mathbf{I}$. Покажем теперь, что $\tilde{\lambda}$ — о.м.п. параметра λ при $\psi = \overline{Y}.$ является асимптотически эффективной.

Отметим, что $\tilde{\lambda}$ является решением уравнения правдоподобия

$$U._\lambda(\lambda', \overline{Y}.) = 0.$$

Раскладывая его в ряд Тейлора, получим

$$U._\lambda(\tilde{\lambda}, \overline{Y}.) = U._\lambda(\hat{\lambda}, \hat{\psi}) + (\tilde{\lambda} - \hat{\lambda})U._{\lambda\lambda}(\lambda^*, \psi^*) + (\overline{Y}. - \hat{\psi})U._{\psi\lambda}(\lambda^*, \psi^*), \tag{1}$$

где

$$|\lambda^* - \hat{\lambda}| < |\tilde{\lambda} - \hat{\lambda}|, \; |\psi^* - \hat{\psi}| < |\overline{Y}. - \hat{\psi}| \text{ и } U._\lambda(\hat{\lambda}, \hat{\psi}) = 0.$$

Из состоятельности оценок (λ^*, ψ^*) и закона больших чисел имеем

$$n^{-1}U._{\lambda\lambda}(\lambda^*, \psi^*) \to i_{\lambda\lambda}(\lambda, \psi) = \lim n^{-1} i._{\lambda\lambda}(\lambda, \psi),$$
$$n^{-1}U._{\lambda\psi}(\lambda^*, \psi^*) \to i_{\lambda\psi}(\lambda, \psi) = \lim n^{-1} i._{\lambda\psi}(\lambda, \psi).$$

Поскольку $\overline{Y}. - \hat{\psi} = o_p(n^{-1/2})$, то из (1) получаем

$$\tilde{\lambda} - \hat{\lambda} = o_p(n^{-1/2}).$$

Отсюда следует, что $\sqrt{n}(\tilde{\lambda} - \lambda)$ и $\sqrt{n}(\hat{\lambda} - \lambda)$ имеют одно и то же предельное нормальное распределение.

Легко проверить, что при $\psi = \overline{Y}.$ оценка $\tilde{\lambda}$ является решением квадратного уравнения, в то время как $\hat{\lambda}$ и $\hat{\psi}$ получаются как решение уравнения четвертой степени. Таким образом, оценка $(\overline{Y}., \tilde{\lambda})$ вычисляется проще.

[Теоретическая статистика § 9.2;
Рао, § 5; Силвей, гл. 4; Клошн,
1973]

9.7. Пусть $Y_1, \ldots, Y_n$ независимы и каждое Y_j имеет распределение Пуассона со средним $\theta_1 + \theta_2 x_j$, где x_j фиксированы. На значения параметра (θ_1, θ_2) наложено ограничение, чтобы каждое из упомянутых средних было положительно. Покажите, что о.м.п. параметра (θ_1, θ_2) асимптотически эквивалентна оценке взвешенных наименьших квадратов. Проведите сравнение свойств этой оценки с оценкой обычных наименьших квадратов. Рассмотрите, в частности, случай $\theta_2 = 0$.

Решение

Логарифм функции правдоподобия имеет вид

$$\text{const} - \sum_{j=1}^{n} (\theta_1 + \theta_2 x_j) + \sum_{j=1}^{n} Y_j \log (\theta_1 + \theta_2 x_j),$$

а вектор вклада

$$U.(\theta) = \begin{bmatrix} -n + \Sigma Y_j (\theta_1 + \theta_2 x_j)^{-1} \\ -\Sigma x_j + \Sigma x_j Y_j (\theta_1 + \theta_2 x_j)^{-1} \end{bmatrix}.$$

Таким образом, формально уравнение правдоподобия $U.(\hat{\theta}) = 0$ имеет тот же самый вид, что и уравнение наименьших квадратов, получаемое при минимизации взвешенных сумм квадратов $\Sigma \omega_j (Y_j - \theta_1 - \theta_2 x_j)^2$ с подстановкой в уравнение наименьших квадратов весов $\omega_j = (\theta_1 + \theta_2 x_j)^{-1} = \text{var}(Y_j)^{-1}$. В том случае, когда объем выборки велик, а x_j и значения параметров таковы, что все средние отделены от нуля, оценки максимума правдоподобия с большей вероятностью будут решениями уравнения правдоподобия. Использование „эмпирических весов“ $\tilde{\omega}_j = Y_j^{-1}$ с соответствующими модификациями для $Y_j = 0$ приводит к оценкам, эквивалентным оценкам максимального правдоподобия при $n \to \infty$, если все $E(Y_j)$ велики.

Уравнения правдоподобия необходимо решать методом итераций. В качестве начального значения можно взять оценки невзвешенных наименьших квадратов:

$$\tilde{\theta}_1 = \bar{Y}. - \tilde{\theta}_2 \bar{x}., \quad \tilde{\theta}_2 = \Sigma Y_j (x_j - \bar{x}.) / \Sigma (x_j - x.)^2.$$

Одна из итеративных процедур для решения рассматриваемой и других аналогичных задач, связанных с пуассоновскими и биномиальными распределениями, состоит в том, чтобы использовать метод взвешенных квадратов, определяя веса на основе оценки параметров на предшествующих этапах итерационного процесса.

Вычисление информационной матрицы на основе значений $\text{var}\{U.(\theta)\}$ приводит к следующему выражению:

$$\text{i.}(\theta) = \begin{bmatrix} \Sigma (\theta_1 + \theta_2 x_j)^{-1} & \Sigma x_j (\theta_1 + \theta_2 x_j)^{-1} \\ . & \Sigma x_j^2 (\theta_1 + \theta_2 x_j)^{-1} \end{bmatrix}.$$

Вычисляется обратная матрица $\text{i}_.^{-1}(\theta)$. После этого можно непосредственно сравнить полученные величины с дисперсиями оценок наименьших квадратов. Оказывается, например, что приближенная асимптотическая дисперсия для $\hat{\theta}_2$ с точностью до $O(\theta_2)$ равна дисперсии для $\tilde{\theta}_2$.

[Теоретическая статистика, § 9.2; Рао, § 5; Силвей, гл. 4]

9.8. Асимптотическая достаточность о.м.н. $\hat{\theta}$ в регулярных задачах означает, что потеря информации при сокращении Y посредством $\hat{\theta}$ асимптотически пренебрежима по сравнению с $i.(\theta)$. Покажите, что для одномерных θ потеря информации равна $E_{\hat{\theta}} \operatorname{var}\{U.(\theta) \mid \hat{\theta}; \theta\}$. С точностью до членов первого порядка это можно вывести из условной дисперсии величины $U'(\theta)$ при условии $U.(\hat{\theta}) = 0$. Аналогичные выкладки будут справедливы также для общих уравнений состоятельного оценивания.

Рассмотрим, в частности, регулярную полиномиальную модель, где $N_1, \ldots, N_m$ — частости, отвечающие выборке объема n с вероятностями исходов $\pi_1(\theta), \ldots, \pi_m(\theta)$. Заметим, что на N_j можно смотреть как на независимые случайные пуассоновские величины со средними $n\pi_j(\theta)$, на которые наложена связь $\Sigma N_j = n$, а $U.(\hat{\theta}) = 0$ можно считать приближенно эквивалентной другой линейной связи для N_j. Исходя из этого, получите явное выражение для $i.(\theta) - i_{\hat{\theta}}(\theta)$.

На основе этих же самых соображений получите величины потери информации при использовании общего асимптотически эффективного уравнения оценивания. Обсудите эти результаты и их связь с приближениями второго порядка для дисперсий предельных нормальных распределений эффективных оценок.

Решение

Обсуждение поставленной задачи будет проводиться здесь в рамках нестрогих рассуждений Фишера (1925), которые впоследствии были превращены в строгие Рао (1961) и Эфроном (1975).

Выражая плотность величины Y в виде произведения условной плотности для Y при заданном $\tilde{\theta}$ и плотности для $\tilde{\theta}$, получим, что

$$U.(\theta; Y) = U(\theta; Y \mid \tilde{\theta}) + U(\theta; \tilde{\theta}).$$

Слагаемые, стоящие в правой части равенства, некоррелированы. Отсюда информационные потери от замены Y на $\tilde{\theta}$ равны

$$\operatorname{var}\{U(\theta; Y \mid \hat{\theta}); \theta\} = \operatorname{var}\{U.(\theta; Y) - U(\theta; \tilde{\theta}); \theta\} =$$
$$= E[\operatorname{var}\{U.(\theta; Y) \mid \tilde{\theta}; \theta\}]. \qquad (1)$$

В последующих обсуждениях $\tilde{\theta}$ может быть любой оценкой. Однако для оценок максимума правдоподобия условие $\tilde{\theta} = t$ означает, что $U.(t; Y) = 0$. В этом случае разложение Тейлора с точностью до членов первого порядка дает

$$\operatorname{var}\{U.(\theta; Y) \mid \bar{\theta} = t; \theta\} = (\theta - t)^2 \operatorname{var}\{U'_{.}(\theta; Y) \mid U.(t; Y) = 0; \theta\}. \quad (2)$$

В общем случае пришлось бы использовать свойства совместного асимптотически нормального распределения величин $U.(\theta; Y)$ и $U'_{.}(\theta; Y)$. Основополагающий результат состоит в том, что для скалярной величины X_2

$$\operatorname{var}(X_2 \mid X_1 = x_1) = \operatorname{var}(X_2) - \\ - \{\operatorname{cov}(X_1, X_2)\}^{\mathrm{T}} \{\operatorname{var}(X_1)\}^{-1} \operatorname{cov}(X_1 X_2). \quad (3)$$

Положив $X_2 = U'_{.}(\theta)$, $X_1 = U.(\theta)$, представление (3) используем при нахождении приближения первого порядка для $U.(t)$ в равенстве (2).

Для полиномиальной задачи используем соотношение (3) в связи с пуассоновским представлением величины N_j, при котором N_j являются независимыми пуассоновскими величинами со средним $n\pi_j(\theta)$, на которые наложена связь $\Sigma N_j = n$. Приближенная связь $U.(\theta) = 0$ означает

$$\Sigma N_j \pi_j'(\theta)/\pi_j(\theta) = 0$$

и

$$U'_{.}(\theta) = \Sigma N_j \{\pi_j''(\theta)/\pi_j(\theta) - [\pi_j'(\theta)/\pi_j(\theta)]^2\}.$$

Итак, применим соотношение (3), где X_1 — некоторая пара линейных комбинаций величин N_j, а X_2 — другая линейная комбинация. Подробные вычисления показывают, что

$$i.(\theta) - i_{\hat{\theta}}(\theta) \sim \\ \sim \frac{\Sigma\{\pi_j''/\pi_j - (\pi_j'/\pi_j)^2\}}{\Sigma(\pi_j'/\pi_j)^2} - \Sigma\frac{(\pi_j')^2}{\pi_j} - \frac{[\Sigma\pi_j'\{\pi_j''/\pi_j - (\pi_j'/\pi_j)^2\}]^2}{\{\Sigma(\pi_j')^2/\pi_j\}^2}.$$

В случае оценивающего уравнения общего вида $g(T, n^{-1}N_1, \ldots, n^{-1}N_m) = 0$ проводится аналогичный анализ, чтобы определить $i.(\theta) - i_T(\theta)$. Такое уравнение приводит к эффективным оценкам, т. е. к эквивалентным оценкам $\hat{\theta}$ с точностью до величин первого порядка, если функция $g(t, p_1, \ldots, p_m)$ удовлетворяет условиям

$$\left[\frac{\partial g}{\partial t} \Big/ \frac{\partial g}{\partial p_j}\right]_{t=\theta,\ p_j=\pi_j(\theta)} = -\frac{\pi_j'(\theta)}{\pi_j(\theta)}\{i.(\theta)\}^{-1}.$$

Частные детали выкладок приводятся в работе Рао (1961), где показано, что потери информации минимальны при $T = \hat{\theta}$.

Когда асимптотические оценки сравниваются на основе более точных вычислений, то становится ясным, что необходимо проявлять осторожность в связи с потерями информации при выборе статистики для построения доверительных интервалов и критериев значимости. Как показал Эфрон (1975), результаты потери информации непосредственно связаны со свойствами асимптотической дисперсии в том смысле, что для скорректированной на смещенность эффективной оценки T_n

$$\operatorname{var}(T_n;\theta) \sim i_{T_n}^{-1}(\theta) + n^{-2}c(\theta) + o(n^{-2}),$$

где $c(\theta)$ не зависит от метода оценивания.

[Теоретическая статистика, § 9.2;
Рао, § 5, 6e]

9.9. [1]) Предположим, что $Y_1, \ldots, Y_n$ — н.о.р. с непрерывной п.р.в.

$$f(y;\theta)=\begin{cases} c(\theta)\,d(y) & \text{при } (a\leqslant y\leqslant b(\theta)),\\ 0 & \text{в противном случае,}\end{cases}$$

где $b(\theta)$ — монотонная функция одномерного параметра θ. Покажите, что о.м.п. для θ имеет вид $b^{-1}(Y_{(n)})$. Исходя из этого, покажите, что критерий м.о.п. (максимума отношения правдоподобий) для проверки гипотезы H_0: $\theta=\theta_0$ против двусторонней альтернативы $\theta\neq\theta_0$ задается с помощью статистики

$$W=-2n\log\int_a^{Y_{(n)}} c(\theta_0)\,d(y)\,dy,$$

имеющей при H_0 в точности хи-квадрат распределение с двумя степенями свободы.

Решение

Предположим для определенности, что $b(\theta)$ возрастает. Тогда

$$\operatorname{lik}(\theta;Y)=\left\{\int_a^{b(\theta)} d(y)\,dy\right\}^{-n}\prod_{j=1}^{n} d(Y_j)\quad \text{для } Y_{(n)}\leqslant b(\theta).$$

Максимум правдоподобия достигается минимизацией интеграла по области

$\{\theta: Y_{(n)}\leqslant b(\theta)\}$, т. е. при выборе $b(\hat\theta)=Y_{(n)}$ и $\hat\theta=b^{-1}(Y_{(n)})$.

Максимальное значение правдоподобия равно

$$\left\{\int_a^{Y_{(n)}} d(y)\,dy\right\}^{-n}\Pi d(Y_j).$$

[1]) Улучшенный вариант.

Если $Y_{(n)} \leqslant b(\theta_0)$, то правдоподобие значения θ_0 равно нулю. При H_0 это невозможно. При $Y_{(n)} < b(\theta_0)$

$$\left\{ \int\limits_a^{b(\theta_0)} d(y)\,dy \right\}^{-n} \Pi d(Y_j) = \{c(\theta_0)\}^n \Pi d(Y_j),$$

откуда

$$W = -2n \log \int\limits_a^{Y_{(n)}} c(\theta_0)\, d(y)\, dy.$$

Далее, для любой непрерывной случайной величины X с функцией распределение F случайная величина $V = F(X)$ имеет равномерное распределение на (0, 1). Таким образом, статистика W равна умноженному на минус два логарифму от $V^n_{(n)}$, где $V_{(n)}$ — максимум из n независимых равномерных случайных величин. Отсюда вытекает[1]), что W равна умноженному на минус два логарифму равномерно распределенной на (0, 1) случайной величины. Следовательно, функция распределения статистики W имеет вид $1 - e^{-\frac{1}{2}z}$, что соответствует функции распределения хи-квадрат случайной величины с двумя степенями свободы.

[Теоретическая статистика, § 9.3; Рао, § 6е; Силвей, гл. 7; Хогг, 1956]

9.10. Пусть $Y_1, \ldots, Y_n$ — н.о.р. случайные величины с регулярной п.р.в. $f(y; \theta)$, где $\dim(\theta) > 1$. Предположим, что на значения θ наложено ограничение вида $\zeta(\theta) = 0$, где функция $\zeta(\cdot)$ дифференцируема. Используйте метод множителей Лагранжа для вывода уравнения правдоподобия с ограничением $\zeta(\theta) = 0$. Получите предельное нормальное распределение вида $\tilde{\theta}$ — о.м.п. с ограничениями.

Пусть ограничение $\zeta(\theta) = 0$ является нулевой гипотезой при альтернативе $\zeta(\theta) \neq 0$. Покажите, что статистики критерия м.о.п. асимптотически эквивалентны нормированной функции от $\zeta^2(\hat{\theta})$, где $\hat{\theta}$ — обычная о.м.п. без ограничений. Покажите, что соответствующий вариант статистики вклада W_u основывается на $U_.(\tilde{\theta})$.

---

[1]) n-я степень максимума из n независимых равномерно на (0, 1) распределенных случайных величин распределена равномерно на (0, 1). — *Прим. перев.*

Решение

Предположим для простоты, что ограничение $\zeta(\theta)=0$—скалярное. Приведенный ниже метод решения задачи легко обобщается на векторный случай. Имеется несколько асимптотически эквивалентных способов решения задачи. В принципе одним из наиболее простых среди них, но часто не подходящим для конкретных приложений, является метод репараметризации. В этом случае $\zeta(\theta)$ становится одной из компонент параметра. Другой подход заключается в том, чтобы работать с асимптотически нормальной формой правдоподобия (см. (1) в разделе „Необходимые сведения") при локально линеаризованной форме ограничений, т. е. с ограничениями на значения параметра вида

$$(\theta-\theta_0)^T(\partial\zeta/\partial\theta_0)=0.$$

Если задачу рассматривать лишь как задачу максимизации с ограничениями, то необходимо взять $l(\theta')-\lambda\zeta(\theta')$ и решить систему

$$U.(\bar\theta)-\bar\lambda\partial\zeta(\bar\theta)/\partial\bar\theta=0,\quad \zeta(\bar\theta)=0$$

относительно θ и λ. Разлагая уравнения системы относительно истинного значения параметра, получим

$$U.(\theta)+\mathbf{i}.(\theta)(\bar\theta-\theta)-\bar\lambda\partial\zeta(\theta)/\partial\theta=0,$$
$$(\bar\theta-\theta)^T(\partial\zeta/\partial\theta)=0.$$

Таким образом, с точностью до величин первого порядка малости имеем

$$\begin{bmatrix} U.(\theta) \\ 0 \end{bmatrix}=\begin{bmatrix} \mathbf{i}.(\theta) & -\partial\zeta/\partial\theta \\ (-\partial\zeta/\partial\theta)^T & 0 \end{bmatrix}\begin{bmatrix} \bar\theta-\theta \\ \bar\lambda \end{bmatrix}=\begin{bmatrix} \mathbf{P} & Q \\ Q^T & R \end{bmatrix}\begin{bmatrix} \bar\theta-\theta \\ \bar\lambda \end{bmatrix}.$$

Вектор, стоящий в левой части равенства, асимптотически $MN\{0, \mathbf{i}^{-1}(\theta)\}$ нормален. Таким образом, вектор $\bar\theta-\theta$ нормально $MN\{0, \mathbf{P}\mathbf{i}^{-1}(\theta)\mathbf{P}^T\}$ распределен, т. е. имеет сингулярное нормальное распределение, подчиняющееся локально линейному ограничению $\zeta(\bar\theta)=0$.

Статистикой отношения правдоподобия для проверки гипотезы H_0: $\zeta(\theta)=0$ является

$$W=2\{l(\hat\theta)-l(\bar\theta)\}\simeq(\hat\theta-\bar\theta)^T\,\mathbf{i}.(\theta)(\hat\theta-\bar\theta),$$

где правое представление статистики W получается после разложения $l(\bar\theta)$ около $(\hat\theta)$. Имеем также

$$\zeta(\hat\theta)=\zeta(\bar\theta)+(\hat\theta-\bar\theta)^T\{\partial\zeta(\hat\theta)/\partial\theta\},$$

где $\zeta(\bar{\theta}) = 0$. Следовательно, асимптотически

$$W \simeq \{\zeta(\hat{\theta})\}^2 [\{\partial\zeta(\bar{\theta})/\partial\theta\}^{\mathrm{T}}\, \mathbf{i}_{.}^{-1}(\theta)\, \{\partial\zeta(\bar{\theta})/\partial\theta\}]^{-1}.$$

В качестве предельного распределения правая часть последнего соотношения имеет хи-квадрат распределение с одной степенью свободы при H_0.

Преимущество последней формы статистики критерия состоит в том, что нет необходимости в явном виде вычислять $\hat{\theta}$. Однако другая форма статистики критерия имеет вид

$$U_{.}^{\mathrm{T}}(\bar{\theta})\, \mathbf{i}_{.}^{-1}(\bar{\theta})\, U_{.}(\bar{\theta}).$$

В общем случае, если ζ — q-мерное соотношение, то аналогичные рассмотренные квадратичные формы от $\zeta(\hat{\theta})$ и $U_{.}(\bar{\theta})$ имеют в качестве предельного хи-квадрат распределение с q степенями свободы.

[Теоретическая статистика, § 9.2, 9.3; Рао, § 6е; Силвей, гл. 7; Эйтчисон и Силвей, 1958]

9.11. Рассмотрим двумерную таблицу сопряженности признаков с частостями N_{jk} ($j = 1, \ldots, r$; $k = 1, \ldots, c$) и соответствующими вероятностями π_{jk}. Для нулевой гипотезы независимости строк от столбцов при общей альтернативе получите соответствующие формы статистик W, W_u и W_e.

Решение

Имеется несколько подходов к решению этой задачи. Один из них состоит в том, чтобы исследовать результаты, полученные в задаче 9.10, с $\theta = \{\pi_{jk}\}$ и ограничениями $\zeta(\theta) = 0$, эквивалентными ограничениям на π_{jk}, предопределяемым независимостью строк и столбцов. Тогда статистика, основанная на $\zeta(\hat{\theta})$, как увидим ниже, будет квадратичной формой от уклонений наблюденных частостей от принятых частостей, т. е. от разностей $N_{jk} - N_{j+}N_{+k}/n$, где N_{j+} и N_{+k} обозначают суммы по строке и столбцу соответственно.

Непосредственно действовать со статистиками W_u и W_e в том виде, как они определены равенством (3) из раздела „Необходимые сведения“, в общем случае довольно сложно. Ограничимся поэтому частным случаем $r = c = 2$. В этом случае произведем репараметризацию

$$\pi_{11} = K^{-1}e^{\lambda_1+\lambda_2+\psi}, \quad \pi_{12} = K^{-1}e^{\lambda_1-\lambda_2-\psi},$$
$$\pi_{21} = K^{-1}e^{-\lambda_1+\lambda_2-\psi}, \quad \pi_{22} = K^{-1}e^{-\lambda_1-\lambda_2+\psi},$$

где K таково, что сумма вероятностей равна единице. Гипотеза независимости строк и столбцов эквивалентна тому, что $\psi=0$.

Из логарифма функции правдоподобия $\Sigma\Sigma N_{jk}\log\pi_{jk}$ получим

$$U_{\cdot\lambda_1}=-n\alpha_1+(N_{1+}-N_{2+}),$$
$$U_{\cdot\lambda_2}=-n\alpha_2+(N_{+1}-N_{+2}),$$
$$U_{\cdot\psi}=-n\beta+(N_{11}+N_{22}-N_{12}-N_{21}),$$

где

$$\alpha_1=\pi_{11}+\pi_{12}-\pi_{21}-\pi_{22},$$
$$\alpha_2=\pi_{11}-\pi_{12}+\pi_{21}-\pi_{22},$$
$$\beta=\pi_{11}+\pi_{22}-\pi_{12}-\pi_{21}.$$

Информационная матрица равна матрице вторых производных от логарифма правдоподобия со знаком минус:

$$\bar{\mathbf{i}}_{\cdot}=n\begin{bmatrix}1-\alpha_1^2 & \beta-\alpha_1\alpha_2 & \alpha_2-\alpha_1\beta\\ \cdot & 1-\alpha_2^2 & \alpha_1-\alpha_2\beta\\ \cdot & \cdot & 1-\beta^2\end{bmatrix}.$$

Чтобы вычислить статистику W_u, заметим, что при гипотезе H_0: $\psi=0$ о.м.п. для π_{jk} имеет вид $N_{j+}N_{+k}/n^2$. Таким образом, принятые частости равны

$$\tilde{N}_{jk}=N_{j+}N_{+k}/n,$$

а

$$U_{\cdot\psi}(0,\tilde{\lambda}_0)=(N_{11}-\tilde{N}_{11})+(N_{22}-\tilde{N}_{22})-(N_{12}-\tilde{N}_{12})-(N_{21}-\tilde{N}_{21}).$$

Отсюда с очевидностью вытекает, что

$$W_u=\Sigma\Sigma(N_{jk}-\tilde{N}_{jk})^2/\tilde{N}_{jk},$$

что является общим результатом.

Статистика W_e пропорциональна квадрату величины

$$\hat{\psi}=\log\{(N_{11}N_{22})/(N_{12}N_{21})\}.$$

Используя формулу $\log(1+x)=x-\frac{1}{2}x^2+\ldots$ при разложении статистики W_e в случае справедливости гипотезы H_0, придем к эквивалентной форме

$$W_e\simeq\Sigma\Sigma(N_{jk}-\tilde{N}_{jk})^2/N_{jk}.$$

Обе статистики, и W_e, и W_u, при H_0 в качестве предельного имеют хи-квадрат распределение с $(r-1)(c-1)$ степенями свободы для произвольных r и c.

Более подробный анализ отношения правдоподобия и эквивалентности статистик типа хи-квадрат для таблиц сопряженности общего типа приводится в работе Бишона и др. (1975). Аналогичная проблематика рассматривается в задачах 9.15 и 9.17.

[Теоретическая статистика, § 9.3; Рао, § 6b, 6c; Силвей, гл. 7]

9.12. Предположим, что $Y_1, \ldots, Y_n$ независимы и Y_j нормально $N(\mu_j, \sigma^2)$ распределены. Здесь $\mu_j = \beta w_j$ или $\mu_j = \gamma z_j$ $(j = 1, \ldots, n)$, параметры β, γ и σ^2 неизвестны, w_j и z_j — известные постоянные. Для того чтобы выяснить, какая из моделей верна, рассмотрим для Y плотность типа искусственной экспоненциальной смеси

$$f_Y(y;\ \psi, \beta, \gamma, \sigma^2 \mid w, z) \propto \{g(y;\ \beta w, \sigma^2)\}^{\psi} \{g(y;\ \gamma z, \sigma^2)\}^{1-\psi},$$

где $g(y;\ \mu, \sigma^2)$ — плотность нормального $N(\mu, \sigma^2)$ распределения. Постройте статистику вклада W_n для проверки гипотезы H_0: $\psi = 1$ против альтернативы H_A: $\psi < 1$. Сравните полученный критерий с критерием обычного точного анализа рассматриваемой задачи.

Решение

После нормирования экспоненциальной смеси логарифм правдоподобия для одного наблюдения с точностью до постоянной равен

$$-\log\sigma - \frac{1}{2\sigma^2}\{\psi(y-\beta w)^2 + (1-\psi)(y-\gamma z)^2 - \\ -\psi(1-\psi)(\beta w - \gamma z)^2\}. \qquad (1)$$

Заметим, что, когда $\psi = 1$, параметр γ оценить нельзя. Если ввести обозначение $\lambda = (\beta, \gamma, \sigma^2)$, то статистика эффективного вклада для проверки гипотезы H_0 определяется в виде нормированной квадратичной формы от $U_{.\psi}(1, \lambda) - b_{.\psi\lambda} U_{.\lambda}(1, \lambda)$, где $b_{.\psi\lambda}$ — вектор коэффициентов регрессии $U_{.\psi}$ на $U_{.\lambda}$, рассчитанных для $(1, \lambda)$. В этом частном случае компоненты величины $U_{.\lambda}$, соответствующие γ, равны нулю при $\psi = 1$.

Оставшиеся компоненты также будут нулевыми, если вместо β и σ^2 подставить их о. м. п., вычисленные в предположении справедливости гипотезы H_0. Следовательно, статистика вклада, соответствующая соотношению (3) раздела „Необходимые сведения", представляется в виде нормированной квадратичной формы от $U_{.\psi}(1, \lambda)$, вычисленной для $\beta = \hat{\beta}_0$ и $\sigma^2 = \hat{\sigma}_0^2$. Далее, если γ фиксировано, то из (1) найдем, что

$$U_{.\psi}(1, \hat{\beta}_0, \gamma, \hat{\sigma}_0^2) = -\gamma(\Sigma z_j w_j - \hat{\beta}_0 \Sigma z_j w_j)/\hat{\sigma}_0^2.$$

Когда эта статистика нормируется, то ясно, что γ пропадает. Оказывается, что нормированная статистика равна

$$\sqrt{W_u} = \frac{\Sigma z_j y_j - \hat{\beta}_0 \Sigma z_j w_j}{\hat{\sigma}_0 \sqrt{\{\Sigma (z_j - dw_j)^2\}}},$$

где $d = \Sigma z_j w_j / \Sigma w_j^2$. Полученная статистика непосредственно связана со стандартным анализом рассматриваемой задачи, будучи эквивалентна критерию остаточной регрессии величины y на z после согласования их регрессии на w. Единственное различие заключается в том, что $\hat{\sigma}_0^2$ заменяется на среднеквадратический остаток одновременно на w и z. В этом случае $\sqrt{W_u}$ будет иметь в точности t-распределение Стьюдента.

Аткинсон (1970) приводит общее обсуждение использования экспоненциальных смесей в модели сравнений и связывает этот подход с методом обобщенного отношения правдоподобий, использованного Коксом (1961, 1962).

[Теоретическая статистика, §9,3; Рао, § 6е; Силвей, гл. 7]

9.13. Говорят, что случайный вектор (X, Y) имеет двумерное пуассоновское распределение, если его компоненты представляются в следующем виде: $X = U + V$, $Y = U + W$, где U, V и W — независимые пуассоновские случайные величины. Заметим, что маргинальные распределения компонент X и Y являются пуассоновскими, а X и Y независимы тогда и только тогда, когда $E(U) = 0$. Постройте асимптотический критерий независимости для независимых пар $(X_1, Y_1), \ldots, (X_n, Y_n)$. Покажите, что с вычислительной точки зрения статистику критерия м. о. п. найти нелегко. В то же время статистика эффективного вклада W_u с мешающими параметрами $E(X)$ и $E(Y)$, которые оцениваются соответственно посредством $\bar{X}.$ и $\bar{Y}.$, приводит к рассмотрению статистики критерия $\Sigma(X_j - \bar{X}.) \times (Y_j - \bar{Y}.)/\sqrt{(\bar{X}.\bar{Y}.n)}$, имеющей при нулевой гипотезе асимптотически нормальное распределение.

Решение

Пусть пуассоновские случайные величины в качестве средних имеют ψ, λ_v и λ_w соответственно. По определению X и Y

$$\mathrm{pr}(X = x,\ Y = y;\ \theta) = e^{\epsilon} = \sum_{u=0}^{\min(x,\ y)} = \frac{\lambda_v^{x-u} \lambda_w^{y-u} \psi^u e^{-(\psi + \lambda_v + \lambda_w)}}{u!\,(x-u)!\,(y-u)!}.$$

На основе этого вычисляем компоненты вектора индивидуальных вкладов при $\psi=0$:

$$(\partial l/\partial\psi)_{\psi=0}=-1+(xy)/(\lambda_v\lambda_w),$$
$$(\partial l/\partial\lambda_v)_{\psi=0}=-1+x/\lambda_v,$$
$$(\partial l/\partial\lambda_w)_{\psi=0}=-1+y/\lambda_w.$$

Отсюда информационная матрица отдельных наблюдений при $\psi=0$ имеет вид

$$\mathbf{i}(0,\ \lambda)=\begin{bmatrix}\lambda_v^{-1}+\lambda_w^{-1}+(\lambda_v\lambda_w)^{-1} & \lambda_v^{-1} & \lambda_w^{-1}\\ \lambda_v^{-1} & \lambda_v^{-1} & 0\\ \lambda_w^{-1} & 0 & \lambda_w^{-1}\end{bmatrix}.$$

При гипотезе H_0 случайные величины X и Y независимы, соответствующие им параметры оцениваются как $\hat{\lambda}_{v0}=\bar{X}.$ и $\hat{\lambda}_{w0}=\bar{Y}.$. Поэтому

$$U_{.\psi}(0,\ \hat{\lambda}_0)=-n+(\sum X_jY_j)/(\bar{X}.\bar{Y}.),$$
$$i_{.}^{\psi\psi}(0,\ \hat{\lambda}_0)=\hat{\lambda}_{v0}\hat{\lambda}_{w0}/n=\bar{X}.\bar{Y}./n,$$

что приводит непосредственно к статистике вклада

$$W_u=\{U_{.\psi}(0,\ \hat{\lambda}_0)\}^2\,i^{\psi\psi}(0,\ \hat{\lambda}_0)=$$
$$=\{\Sigma(X_j-\bar{X}.)(Y_j-\bar{Y}.)\}^2/(n\bar{X}.\bar{Y}.).$$

Заметим, что статистики $\bar{X}.$ и $\bar{Y}.$ — минимальные достаточные при H_0. Отсюда вытекает, что статистика W_u определяет, в принципе, подобный критерий, использующий условное распределение статистики ΣX_jY_j при заданном значении $\bar{X}.=\bar{x}.$, $\bar{Y}.=\bar{y}.$. Точная условная дисперсия статистики $\Sigma(X_j-\bar{X}.)\times$ $\times(Y_j-\bar{Y}.)$ при заданном значении $\bar{X}.=\bar{x}.$ и $\bar{Y}=\bar{y}.$ равна $(n-1)\bar{x}.\bar{y}.$.

[Теоретическая статистика, § 9.3; Рао, § 6е; Силвей, гл. 7; Нейман, 1959]

9.14. Пусть $Y_1,\ \ldots,\ Y_n$ — н.о.р. с плотностью $f(y;\ \psi,\ \lambda)$. Рассмотрим сложную гипотезу H_0: $(\psi,\ \lambda)\in\Omega_{\psi0}\times\Omega_\lambda$ при альтернативе $(\psi,\ \lambda)\in(\Omega_\psi-\Omega_{\psi0})\times\Omega_\lambda$. Покажите, что критерий м.о.п. для проверки гипотезы H_0 в том и только том случае асимптотически эквивалентен критерию проверки гипотезы $H_{0\lambda}$: $(\psi,\ \lambda)\in\Omega_{\psi,0}\times\{\lambda\}$, когда $\mathbf{i}_{.\psi\lambda}=0$. Обсудите критически возможные применения указанных результатов.

Решение

Неявно предполагается, что H_0 — подпространство гипотез, т. е. что Ω_{ψ_0} — подпространство пространства Ω_{ψ_0}. Предположим, что параметр ψ изоморфен паре (ξ, η), причем из $\psi \in \Omega_{\psi_0}$ вытекает $\xi = \xi_0$ и наоборот. Положим в этом случае $\theta = (\xi, \eta, \lambda)$ и $\phi = (\eta, \lambda)$. Тогда

$$\mathbf{i}.(\theta) = \begin{bmatrix} \mathbf{i}_{.\xi\xi} & \mathbf{i}_{.\xi\phi} \\ \mathbf{i}^{\mathrm{T}}_{.\xi\phi} & \mathbf{i}_{.\phi\phi} \end{bmatrix}.$$

Известно, что асимптотически статистика критерия отношения правдоподобия эквивалентна

$$W_e = (\hat{\xi} - \xi_0)^{\mathrm{T}} \mathbf{i}.(\hat{\xi} : \hat{\phi}) (\hat{\xi} - \xi_0),$$

где

$$\mathbf{i}.(\xi : \phi) = \mathbf{i}_{.\xi\xi} - \mathbf{i}_{.\xi\phi} \mathbf{i}^{-1}_{.\phi\phi} \mathbf{i}^{\mathrm{T}}_{.\xi\phi}.$$

Кроме того, о.м.п. параметра $\psi = (\xi, \eta)$, когда $\lambda = \bar{\lambda}$ будет $\hat{\psi}$, поэтому статистика отношения правдоподобия с заданным значением λ, равным $\bar{\lambda}$, имеет вид

$$\tilde{W}_e = (\hat{\xi} - \xi_0)^{\mathrm{T}} \mathbf{i}.(\hat{\xi} : \bar{\eta}) (\hat{\xi} - \xi_0),$$

где

$$\mathbf{i}.(\xi : \phi) = \mathbf{i}_{.\xi\xi} - \mathbf{i}_{.\xi\phi} \mathbf{i}^{-1}_{.\phi\phi} \mathbf{i}^{\mathrm{T}}_{.\xi\phi}.$$

Рассматриваемые квадратические формы эквивалентны, если $\mathbf{i}_{.\xi\phi} \mathbf{i}^{-1}_{.\phi\phi} \mathbf{i}^{\mathrm{T}}_{.\xi\phi} = \mathbf{i}_{.\xi\eta} \mathbf{i}^{-1}_{.\eta\eta} \mathbf{i}^{\mathrm{T}}_{.\xi\eta}$, откуда следует, что $\mathbf{i}_{.\psi\lambda} = 0$. Это равенство является и необходимым условием эквивалентности. В этом можно убедиться, или раскладывая левую часть равенства, или обращаясь к методам, связанным с линейной регрессией. Указанный результат можно рассматривать как существенное обобщение результатов регрессионной теории в том, что частный регрессионный коэффициент имеет дисперсию, равную дисперсии общего коэффициента регрессии только при выполнении условия ортогональности.

Отметим, что из условия $i_{.\psi\lambda} = 0$ вытекает локальная факторизация асимптотически нормального правдоподобия на Ω_{ψ_0}, приводящая к исчезновению в отношении правдоподобия членов, зависящих от λ.

Указанный результат полезен в ситуациях, когда λ легко оценить для всего Ω. Именно этот случай имеет место при оценивании параметров масштаба, когда ψ состоит из параметров сдвига.

В качестве конкретного применения полученных результатов в многомерном анализе укажем задачу проверки гипотезы о том, что m векторов средних лежат в q-мерном подпространстве ($q < m$). Частный случай $q=0$ соответствует ситуации равенства векторов. Если о.м.п. ковариационный матрицы по всему параметрическому пространству рассматривать как фиксированную матрицу, то это приведет к значительным упрощениям и к стандартному критерию, основанному на канонических переменных (Рао, § 8с.6)

[Теоретическая статистика, § 9.3; Рао, § 6е; Силвей, гл. 7]

9.15. Рассмотрим стационарный дискретный случайный процесс с r состояниями. Пусть M_{jk} — число одношаговых переходов из состояния j в состояние k на протяжении n последовательных наблюдений. Покажите, что статистика критерия м.о.п. для проверки нулевой гипотезы о последовательной независимости против альтернативы марковской зависимости первого порядка асимптотически эквивалентна хи-квадрат статистике

$$T=\Sigma\Sigma\frac{(M_{jk}-\tilde{M}_{jk})^2}{\tilde{M}_{jk}},$$

где $\tilde{M}_{jk}=M_{j.}M_{.k}/(n-1)$. Получите непосредственно соответствующую W_u-статистику.

Решение

Число переходов M_{jk} соответствует марковским переходным вероятностям

$$\theta_{jk}=\mathrm{pr}\,(Y_i=k \mid Y_{i-1}=j) \quad (j,\ k=1,\ \ldots,\ r).$$

Нулевая гипотеза состоит в том, что $\theta_{jk}=\lambda_k$. Пусть

$$M_{j.}=\sum_{k=1}^{r} M_{jk}, \quad M_{.k}=\sum_{j=1}^{r} M_{jk},$$

а N_j — частости попадания в состояние j в последовательности из последних $(n-1)$ наблюдений. Здесь будет предполагаться, что первое наблюдение Y_1 фиксировано, хотя при больших n это не имеет значения. Итак, используя функцию правдоподобия

$$l(\theta)=\Sigma\Sigma\, M_{jk}\log\theta_{jk},$$

легко показать, что $\hat{\theta}_{jk}=M_{jk}/M_{j.}$, в то время как при нулевой гипотезе о независимости $\hat{\theta}_{jk,0}=\hat{\lambda}_k=N_k/(n-1)$.

Статистика отношения правдоподобия определяется следующим образом:

$$W = 2\{l(\hat{\theta}) - l(\hat{\theta}_0)\} = \\ = 2[\Sigma\Sigma M_{jk} \log(M_{jk}/M_{j.}) - \Sigma N_k \log\{N_k/(n-1)\}],$$

где, в соответствии с принятыми обозначениями, $N_k = M_{.k}$. Отсюда

$$W = 2\Sigma\Sigma M_{jk} \log\left\{1 + \frac{(n-1)M_{jk} - M_{j.}M_{.k}}{M_{j.}M_{.k}}\right\}.$$

Используя разложение $\log(1+x) = x - \frac{1}{2}x^2 + O(x^3)$, равенство $(n-1)\Sigma\Sigma M_{jk} = \Sigma M_{j.}\Sigma M_{.k}$ и тот факт, что при нулевой гипотезе из

$$M_{jk}/M_{j.} = \lambda_k + o_p(1), \quad M_{.k}/(n-1) = \lambda_k + o_p(1)$$

вытекает $(n-1)M_{jk}/(M_{j.}M_{.k}) = 1 + o_p(1)$, получим

$$W = \Sigma\Sigma(M_{jk} - \tilde{M}_{jk})^2/\tilde{M}_{jk} + o_p(1) = T + o_p(1).$$

Если N_k считать, как обычно, частостями в полной последовательности длины n, то должен появиться еще один член $o_p(1)$ и аппроксимирующая хи-квадрат статистика останется без изменения. Из стандартной теории статистик W следует, что предельное хи-квадрат распределение этих эквивалентных статистик имеет $(r-1)^2$ степеней свободы.

Статистика вклада T есть не что иное, как статистика вклада W_u. Прямой вывод этой статистики во многом подобен тому, как была получена на основе использования $U.(\theta_0)$ статистика вклада в задаче 9.11.

[Теоретическая статистика, § 9.3; Рао, § 6е; Силвей, гл. 7; Андерсон и Гудман, 1957]

9.16. В регулярных параметрических задачах состоятельная оценка $T = t(Y)$ в качестве предельного имеет нормальное $MN_q\{0, \mathbf{v}(\theta)\}$ распределение. Покажите, что T и м.о.п. $\hat{\theta}$ имеет предельное совместное нормальное распределение с матрицей перекрестных ковариаций вида $\mathbf{i}_.^{-1}(\theta)$. Далее, для проверки нулевой гипотезы H_0: $\theta = \theta_0$ против альтернативы $\theta \neq \theta_0$ рассмотрите статистику W_T, определенную равенством

$$\exp\left(\frac{1}{2}W_T\right) = \operatorname{lik}(T; Y)/\operatorname{lik}(\theta_0; Y).$$

Сравнивая W_T со статистикой м.о.п., покажите, что предельное распределение статистики W_T при H_0 такое же, как у суммы

$$\sum_{j=1}^{q} Z_j^2 - \sum_{j=1}^{q} \alpha_j(\theta) Z_{q+j}^2,$$

где $Z_1, \ldots, Z_{2q}$ — н.о.р. случайные величины и имеют нормальное $N(0, 1)$ распределение, а $\alpha_j(\theta)$ — собственные значения матрицы $\mathbf{i}.(\theta)\, \mathbf{v}(\theta) - \mathbf{I}$.

Предположим, что $Y_1, \ldots, Y_n$ — н.о.р. случайные величины с непрерывной функцией распределения $f_Y(y; \theta)$. Пусть значения указанных непрерывных случайных величин сгруппированы в m интервалов, соответствующие вероятности попаданий в них равны $\pi_j(\theta)$, а частности N_j, $j = 1, \ldots, m$. Пусть $\hat{\theta}_Y$ — о.м.п., полученная на основе негруппированных данных. Используйте соображения, аналогичные примененным ранее для предельного распределения при нулевой гипотезе хи-квадрат статистики согласия

$$\sum_{j=1}^{m} \frac{\{N_j - n\pi_j(\hat{\theta}_Y)\}^2}{n\pi_j(\hat{\theta}_Y)}.$$

Решение

Приведем только набросок решения, опуская подробные математические выкладки, которые легко найти в указываемой литературе.

Для простоты предположим сначала, что θ — скаляр. Тогда для любой постоянной a статистика $T + (1-a)\hat{\theta}$ имеет асимптотически среднее θ и дисперсию $\{a^2 v(\theta) + 2a(1-a)\operatorname{cov}(T, \hat{\theta}) + (1-a)^2/i.\}$. Отсюда заключаем, что для того, чтобы эта дисперсия имела минимум при $a=0$, необходимо, чтобы $\operatorname{cov}(T, \hat{\theta}) = 1/i.$ Другими словами, $T - \hat{\theta}$ и $\hat{\theta}$ асимптотически должны быть некоррелированы.

Для исследования W_T произведем разложение

$$W_T = 2[\{l_Y(T) - l_Y(\hat{\theta})\} + \{l_Y(\hat{\theta}) - l_Y(\theta_0)\}] \sim$$
$$\sim (\hat{\theta} - \theta_0)^{\mathrm{T}}\, \mathbf{i}.(\theta_0)\, (\hat{\theta} - \theta_0) - (T - \hat{\theta})^{\mathrm{T}}\, \mathbf{i}.(\theta_0)\, (T - \hat{\theta}).$$

В силу приведенных соображений второе слагаемое асимптотически не зависит от первого. Используем теперь следующее свойство (Рао, § 8а): если X — нормально $MN(0, \mathbf{a})$ распределенный вектор, то $X^{\mathrm{T}}\mathbf{B}X = \Sigma\alpha_j Z_j^2$, где Z_j — н.о.р. случайные величины, имеющие нормальное $N(0, 1)$ распределение, а α_j — собственные значения матрицы $\mathbf{ba}$. Положим

$$X^{\mathrm{T}} = (\hat{\theta}, T - \hat{\theta})^{\mathrm{T}}, \text{ а } \mathbf{a} = \operatorname{diag}\{\mathbf{i}.^{-1}(\theta)\}.$$

Это приводит к статистике

$$W_T \sim \sum_{j=1}^{q} Z_j^2 - \sum_{k=1}^{q} \alpha_k Z_{q+k}^2 \qquad (2)$$

с α_k, являющимся собственными значениями матрицы $\mathbf{i}.(\theta)\{\mathbf{v}(\theta)-\mathbf{i}^{-1}(\theta)\}=\mathbf{i}.(\theta)\mathbf{v}(\theta)-\mathbf{I}$. Заметим, что собственные значения α_k неотрицательны. Это и следовало ожидать, поскольку имелось соотношение $W_T \leqslant W = Z_1^2 + \ldots + Z_q^2$.

Чтобы применить этот подход к критериям согласия, предположим, что имеется n н.о.р. случайных величин $Y_1, \ldots, Y_n$ с п.р.в. $f(y;\theta)$. Если сгруппировать их значения в m интервалов с вероятностями попадания значения Y_j в соответствующие интервалы $\pi_1(\theta), \ldots, \pi_m(\theta)$, то критерии согласия могут быть основаны только на соответствующих частостях $N_1, \ldots, N_m$. При этом необходимо использовать $\hat{\theta}_N$ — оценку максимального правдоподобия, вычисленную только на основе этих частостей.

В обычных обозначениях статистика отношения правдоподобий имеет вид

$$W - 2[l_N(\hat{\pi}) - l_N\{\pi(\hat{\theta}_N)\}] = 2\Sigma N_j \log[N_j/\{n\pi_j(\hat{\theta}_N)\}],$$

а ее асимптотически эквивалентные формы

$$W_u = \Sigma\{N_j - n\pi_j(\hat{\theta}_N)\}^2/\{n\pi_j(\hat{\theta}_N)\},$$
$$W_e \simeq \Sigma\{N_j - n\pi_j(\hat{\theta}_N)\}^2/N_j.$$

Каждая из этих статистик в качестве предельного имеет хи-квадрат распределение с $m-1-\dim(\theta)$ степенями свободы.

Часто, однако, было бы проще и, в принципе, лучше оценить θ исходя из исходных значений Y. Например, можно использовать соответствующую о.м.п. $\hat{\theta}_Y$. Возникающая при этом статистика отношения правдоподобия

$$\bar{W} = 2[l_N(\hat{\pi}) - l_N\{\pi(\hat{\theta}_Y)\}]$$

или эквивалентные ей хи-квадрат статистики имеют в качестве предельного распределения не типа хи-квадрат. Раскладывая статистику $\bar{W}$ аналогично тому, как это сделано для статистики W_T в (1), убедимся, что предельным для $\bar{W}$ будет распределение типа (2), зависящее от собственных значений матрицы $\operatorname{var}(\hat{\theta}_Y)\{\operatorname{var}(\hat{\theta}_N)\}^{-1}$, причем $\bar{W} \geqslant W$. Ватсон (1958) обсуждает задачу, в которой границы интервалов группировки оцениваются квантилями принятого распределения, т. е. распределения с плотностью $f(y;\hat{\theta}_Y)$. Далее, он, а также Чернов и Леман (1954) приводят подробное и строгое обсуждение затронутых выше вопросов. В общем случае, если m не слишком мало, например более 3 или 4, то эффект от использования $\hat{\theta}_N$ вместо $\hat{\theta}_Y$ пренебрежимо мал.

[Теоретическая статистика, § 9.3; Рао, § 6b, 6e; Силвей, гл. 7]

9.17[1]). Две случайные величины U и V, принимающие целые неотрицательные значения, называют квазинезависимыми, на множестве $\mathscr{S}$, если для некоторых функций $\alpha(\cdot)$ и $\beta(\cdot)$

$$\mathrm{pr}\,(U=u,\ V=v)=\alpha(u)\,\beta(v)\quad (u,\ v\in\mathscr{S}).$$

Семейство случайных величин $X^{(n)}$ и $Y^{(n)}$, принимающих значение в области $0\leqslant x+y\leqslant n$, называют (F)-независимыми, если для фиксированных функций $\alpha'(\cdot)$, $\beta'(\cdot)$, $\gamma'(\cdot)$ и $\delta'(\cdot)$, не зависящих от n,

$$\mathrm{pr}\,(X^{(n)}=x,\ Y^{(n)}=y)=\alpha'(n)\,\beta'(x)\,\gamma'(y)\,\delta'(n-x-y)$$

для $n=1,\ \ldots,\ N$. Обсудите, имеется ли между этими определениями связь. Приведите примеры ситуаций, когда одно из определений подходит, а другое — нет.

Покажите, как построить критерий для проверки нулевых гипотез о F-независимости и квазинезависимости.

Решение

Две дискретные случайные величины Z и W независимы, если для всех z и w

$$\mathrm{pr}\,(Z=z,\ W=w)=\alpha''(z)\,\beta''(w)$$

оба определения, исследуемые в данной задаче, имеют цель обобщить понятие независимости. Квазинезависимость наиболее естественно возникает в связи с таблицами сопряженности (Гудман, 1968; Бишоп и др., 1975, § 5.2), где значение u и v — метки строк и столбцов двумерной таблицы. Частные случаи возникают в том случае, когда во всех ячейках, кроме одной или некоторой малой группы ячеек, ожидается наличие свойства независимости. Далее, если u представляет „исходное" состояние, а v — „конечное" состояние, свойство независимости может применяться условно по мере осуществления движения. Это приводит к квазинезависимости на множестве $\mathscr{S}=\{u,\ v;\ u=v\}$.

---

[1]) Улучшенный вариант.

F-независимость является более неявным понятием, вероятностная интерпретация которой менее ясна. Дэрроч (1971) подробно обсуждал возможность вывода этого понятия на основе формулировки требований к условному распределению. Заметим, что

$$\mathrm{pr}\,(X^{(n)}=x \mid Y^{(n)}=y)=\beta'(x)\,\delta'(n-x-y)\,\varepsilon'(n-y),$$

т. е. является функцией только x и $n-y$. Имеется симметричное свойство для условного распределения величины $Y^{(n)}$ при заданном значении $X^{(n)}=x$. В частности, из определения вытекает, например, что

$$\mathrm{pr}\,(X^{(10)}=6 \mid Y^{(10)}=2)=\mathrm{pr}\,(X^{(9)}=6 \mid Y^{(9)}=1)=\mathrm{pr}\,(X^{(8)}=6 \mid Y^{(8)}=0).$$

Например, $X^{(n)}$ и $Y^{(n)}$ равны числу раз среди n испытаний, когда наблюдались отклики „весьма хороший" и „хороший" соответственно. При каждом испытании возможны пять различных откликов. Условие F-независимости является одним из способов, которым может определяться взаимосвязь в семействе двумерных случайных величин.

Если $\delta'(\cdot)=1$, то F-независимость сводится к квазинезависимости о $\mathcal{S}=\{(x,\ y);\ 0\leqslant x+y\leqslant n\}$. Однако в общем случае из одного типа независимости не вытекает другой.

Чтобы проверить согласие с этими частными моделями для некоторого множества возможных значений, можно найти максимум отношения правдоподобия, сравнивая согласие для нулевой гипотезы при максимизации по мешающим параметрам, против произвольного полиномиального распределения на соответствующей решетке.

Чтобы иметь дело с простым примером, рассмотрим нулевую гипотезу квазинезависимости для множества $\{u,\ v;\ u\neq v\}$ в $t\times t$ таблице сопряженности с отдельными частостями N_{uv}. При нулевой гипотезе вероятности, отвечающие диагональным ячейкам, произвольны, поэтому о.м.п. для ожидаемых частостей этих ячеек являются наблюденные частости. Поскольку то же самое верно и в общем случае, диагональные ячейки можно не рассматривать. Для внедиагональных ячеек вклад в общий логарифм правдоподобия равен

$$\sum_{u}\sum_{v\neq u} N_{uv}\log\alpha(u)+\sum_{v}\sum_{u\neq v} N_{uv}\log\beta(v).$$

Таким образом, достаточными статистиками являются суммы по строкам и столбцам, в которых пропущены диагональные ячейки. Оценки максимума правдоподобия ожиданий частостей ячеек связаны ограничением, чтобы они давали те же суммы. Дело в том, что достаточные статистики являются о.м.п. своих собственных ожиданий для семейства экспоненциальных распределений. Сам вывод о.м.п. придется проводить на основе итеративных методов. Если окончательно принятые частости обозначить $\bar{N}_{uv}$, то отсюда будет следовать, что статистика о.п. имеет вид

$$W = 2 \sum_{u \neq v} N_{uv} \log (N_{uv}/\bar{N}_{uv}),$$

а соответствующая степень свободы равна $(t-1)(t-1)-t = t^2-3t+1$. При этом предполагается, что не существует других ограничений на таблицу, таких, например, как фиксированные нули в таблице.

Для получения о.м.п. обычно необходимы, как и в рассмотренном примере, итеративные методы. Необходима также осторожность при подсчете числа степеней свободы. В случае F-независимости число степеней свободы равно $\frac{1}{2}(n-1)(n-2)$ (Дерроч и Ретклифф, 1975).

[Теоретическая статистика, § 9.3, (iii)]

9.18. Пусть н.о.р. случайные величины $Z_1, \ldots, Z_n$ нормально $N(\mu, \sigma^2)$ распределены, и пусть $Y_j = |Z_j|$. Покажите, что при использовании Y для проверки H_0: $\mu = 0$ против H_A: $\mu \neq 0$ с неизвестной в обоих случаях дисперсией σ^2 критерий м.о.и. асимптотически эквивалентен критерию, основанному на больших значениях статистики $T = \Sigma Y_j^4/(\Sigma Y_j^2)^2$. Каково предельное распределение статистики T?

Решение

Пусть Y_j — н.о.р. случайные величины с п.р.в., пропорциональной

$$\sigma^{-1} \exp\left\{-\frac{1}{2}(y^2+\mu^2)/\sigma^2\right\} \operatorname{ch}(y\mu/\sigma^2).$$

Отсюда логарифмическая функция правдоподобия с точностью до постоянной имеет вид

$$l.(\mu, \sigma^2) = -n \log \sigma - \frac{n\mu^2}{2\sigma^2} - \frac{\Sigma y_j^2}{2\sigma^2} + \Sigma \log \operatorname{ch}(\mu y_j/\sigma^2).$$

Легко проверить, что оценки максимального правдоподобия для μ и σ являются состоятельными оценками. Необходимо учесть, конечно, что значения $\pm|\mu|$ приводят к равным значениям функции правдоподобия. В обычном смысле функция правдоподобия не является регулярной. Дело в том, что при $\mu = 0$ информационная матрица не является положительно определенной. В принятой параметризации $i_{\mu\mu}(0, \sigma)=0$. Чтобы получить асимптотические результаты, необходимо раскладывать логарифм функции правдоподобия вплоть до членов порядка μ^4, $\mu^2\sigma^2$ и σ^4. При переходе к новой параметризации $\theta=(\xi, \tau)=(\mu^2, \sigma^2)$ указанного выше затруднения можно избежать. В новой параметризации эффективный вклад имеет вид

$$U_{.\xi}(\theta)=-\frac{n}{2\tau}+\frac{1}{2\tau\sqrt{\xi}}\Sigma y_j \tanh\left(y_j\sqrt{\xi/\tau}\right),$$

$$U_{.\tau}(\theta)=\frac{n}{2\tau}+\frac{n\xi}{2\tau^2}+\frac{\Sigma y_j^2}{2\tau^2}-\frac{\sqrt{\xi}}{\tau^2}\Sigma y_j \tanh\left(y_j\sqrt{\xi/\tau}\right).$$

Еще одно затруднение состоит в том, что при нулевой гипотезе H_0: $\xi=0$ не существует открытого множества в параметрическом пространстве, окружающего истинное значение параметра. Информационная матрица положительно определена почти всюду.

Заметим, что при $\theta=(0, \tau)$

$$U_{.\xi}=U_{.\tau}=(\Sigma y_j^2-n\tau)/(2\tau^2).$$

Получаем отсюда, что $(0, n^{-1}\Sigma y_j^2)$ всегда является решением уравнения правдоподобия $U_.=0$. Однако, если матрица вторых производных для $l_.$ отрицательно определена, то указанное решение является локальным максимумом. Обозначим частные производные второго порядка $U_{.\xi\xi}$ и т. д. Получим, что при $\theta=(0, \tau)$

$$U_{.\xi\xi}=-\sum y_j^4/(3\tau^4),$$
$$U_{.\xi\tau}=(n\tau-2\Sigma y_j^2)/(2\tau^3),$$
$$U_{.\tau\tau}=(n\tau-2\Sigma y_j^2)/(2\tau^3).$$

Ясно, что матрица вторых производных для $l_.$ в точке $(0, \Sigma y_j^2/n)$ отрицательно определена в том и только том случае, если $3(\Sigma y_j^2)^2<n\Sigma y_j^4$. В противном случае в точке $(0, \Sigma y_j^2/n)$ находится седловая точка функции правдоподобия. Поэтому для получения о.м.п. в этом случае необходимо искать другие локально стационарные точки.

При малых ξ дальнейшее разложение уравнения правдоподобия приводит к следующей системе:

$$0=U_{\xi}(\theta)=(\Sigma y_j^2-n\tau)/(2\tau^2)-\xi\Sigma y_j^4/(6\tau^4)+O(\xi^2),$$
$$0=U_{\tau}(\theta)=-n/(2\tau)+n\xi/(2\tau^2)+\Sigma y_j^2/(2\tau^2)-\xi\Sigma y_j^2/\tau^3+$$
$$+\xi^2\Sigma y_j^4/(3\tau^5)+o(\xi^2).$$

Новое решение имеет вид

$$\bar{\xi}=(\Sigma y_j^2-n\bar{\tau})/n,\quad \bar{\tau}=\Sigma y_j^4/(3n).$$

Логарифм статистики отношения правдоподобий для проверки нулевой гипотезы H_0: $\xi = 0$ записывается как

$$W = 2\{l.(\hat{\xi},\ \hat{\tau}) - l.(0.\tilde{\tau}_0)\}.$$

Если $3(\Sigma y_j^2)^2 \leqslant n\Sigma y_j^4$, то он равен нулю. В противном случае разложение этой статистики приводит к следующему выражению:

$$W \simeq \hat{\xi}^2 U_{.\xi\xi}(0,\ \tilde{\tau}_0) + 2\hat{\xi}(\hat{\tau} - \tilde{\tau}_0)\, U_{.\xi\tau}(0,\ \tilde{\tau}_0) + (\hat{\tau} - \tilde{\tau}_0)^2\, U_{\tau\tau}(0,\ \tilde{\tau}_0).$$

После некоторых упрощений получим, что большие значения принимаются статистикой при малых значениях величины $\hat{\tau}/\tilde{\tau}_0 - 1 = [n\Sigma y_j^4/\{3(\Sigma y_j^2)^2\}]^{1/2} - 1$. Следовательно, критерий отношения правдоподобия асимптотически эквивалентен критерию с критической областью

$$w_\alpha = \left\{y : \frac{n\Sigma y_j^4}{3(\Sigma y_j^2)^2} < 1 - c_\alpha\right\}.$$

Отметим, что статистика критерия есть не что иное, как выборочный эксцесс, измеряющий степень пологости плотности распределения.

[Теоретическая статистика, § 9.3; Рао, § 6е; Силвей, глава 7; Хинкли, 1973]

9.20 [1]**.** В случае нормального $N(\mu,\ \sigma^2)$ распределения с неизвестной дисперсией σ^2 для задачи проверки гипотезы $H_0 : \mu = 0$ против альтернативы $H_A : \mu \neq 0$ сравните численно точные распределения W_u и W_e-статистик с предельным хи-квадрат распределением. Получите для W_u- и W_e-статистик поправки второго порядка, основанные на первых моментах. Определите численными методами, получено ли какое-нибудь улучшение по сравнению с использованием приближений первого порядка.

Решение

Логарифм функции правдоподобия для $\theta = (\mu,\ \sigma^2)$ имеет вид

$$l(\theta) = \mathrm{const} - \frac{1}{2} n \log \sigma^2 - \frac{1}{2\sigma^2} \Sigma (Y_j - \mu)^2.$$

---

[1]) Задача 9.19 опущена.

Отсюда на основе простых вычислений получаем

$$U_{\cdot} = \begin{bmatrix} \frac{1}{\sigma^2} \Sigma (Y_j - \mu) \\ -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4} \Sigma (Y_j - \mu)^2 \end{bmatrix},$$

$$\mathbf{i} = \operatorname{diag}\left(n/\sigma^2, \ \frac{1}{2} n/\sigma^4 \right).$$

Соответствующие оценки максимального правдоподобия равны $\hat{\mu} = \bar{Y}_{.}$, $\hat{\sigma}^2 = \Sigma (Y_j - \bar{Y})^2/n = \mathrm{SS}/n$. При нулевой гипотезе $H_0: \mu = 0$, $\hat{\sigma}_0^2 = (\mathrm{SS} + n\bar{Y}^2)/n$.

В последнем случае статистики W_e и W_u принимают следующий вид:

$$W_e = n\hat{\mu}^2/\hat{\sigma}^2, \quad W_u = (n\bar{Y}_{.}/\hat{\sigma}_0^2)^2 (\hat{\sigma}_0^2/n).$$

Отсюда если положить $T = \sqrt{n(n-1)/\mathrm{SS}}\,\bar{Y}_{.}$, то через эту t-статистику Стьюдента статистики W_e и W_u выразятся так:

$$W_e = nT^2/(n-1), \quad W_u = nT^2 \{1 + T^2/(n-1)\}^{-1} (n-1)^{-1}.$$

Чтобы провести точное численное сравнение допредельного с предельным нормальным распределением статистик $\sqrt{W_e}$ и $\sqrt{W_u}$, возьмем, для примера, $n = 10$ и $t = 2.262$, двустороннюю 5% точку t-распределения Стьюдента с 9 степенями свободы. Вычисление значений статистик W_e и W_u дает соответственно 5.685 и 3.625, что отвечает хи-квадрат вероятностям величины 0.017 и 0.057. Ясно, что асимптотические результаты не очень точны, особенно для W_e. Заметим к тому же, что W_u не является возрастающей функцией от T^2.

Чтобы получить при H_0 поправки к W_e и W_u, основанные на приравнивании первых моментов к среднему предельного хи-квадрат распределения, воспользуемся тем, что

$$E(T^2; H_0) = (n-1)/(n-3) \text{ и } E(T^4; H_0) = 3(n-1)^2/\{(n-3)(n-5)\}.$$

Отсюда просто получаем, что $E(W_e; H_0) = n/(n-3)$, так что вариант статистики W_e, подправленный на основе приравнивания средних, будет выглядеть как $W_e^* = (n-3)W_e/n = (n-3)T^2/(n-1)$. При рассмотрении статистики W_u разложим ее в ряд по степеням от T^2. Получим

$$E(W_u; H_0) = \frac{n}{n-1} E(T^2; H_0) - \frac{n}{(n-1)^2} E(T^4; H_0) + O(n^{-2}) = \\ = 1 + O(n^{-2}),$$

т. е. $W_u^* = W_u$. Этот же самый результат можно было бы получить непосредственно, без использования свойств t-распределения Стьюдента. Для единственного рассмотренного численного примера вычисленное значение статистики W_e^* равно 3.98, что соответствует уровню 0.046 правого хвоста хи-квадрат распределения. Итак, с практической точки зрения там, где „точное" решение получить нелегко и не требуется использование чрезвычайно высоких уровней, разумно применять статистики $W_u^* = W_u$ и W_e^*. Было бы поучительно провести более развернутое численное исследование этого и других примеров.

В принципе, всегда желательно исследовать достаточность аппроксимаций, основанных на асимптотической теории. Привлечение результатов асимптотической теории вторых порядков — один из полезных способов проведения этого.

[Теоретическая статистика, § 9.3]
Рао, § 6; Силвей, гл. 7]

9.21. Используйте асимптотические методы для нахождения доверительных пределов для обоих параметров в модели задачи 4.13 и проверки истинности модели. Считайте оба параметра неизвестными.

Решение

С точностью до постоянной логарифм правдоподобия имеет вид

$$l(\mu, \rho) = T_1 \log \mu + T_2 \log \rho - \mu (1 - \rho^n)/(1 - \rho),$$

где $T_1 = \Sigma Y_j$, а $T_2 = \Sigma (j-1) Y_j$. В принципе, точные доверительные интервалы для μ и ρ можно получить соответственно из условных распределений статистики T_1 при заданном значении $T_2 = t_2$ и условных распределений статистики при заданном значении $T_1 = t_1$ (см. гл. 5 и 7). Так, например, доверительные границы для ρ будут получаться на основе значений тех экстремальных значений ρ_0, для которых t_2 не лежит в критической области, построенной для проверки нулевой гипотезы H_0: $\rho = \rho_0$ против соответствующих альтернатив.

Исходя из этого, привлечение асимптотических соображений можно обосновать стремлением упростить выкладки, и, во всяком случае, при этом будут проиллюстрированы некоторые общие факты. Информационная матрица, отвечающая полной совокупности данных, будет диагональна с ведущим элементом

$$E(-\partial^2 l/\partial\mu^2) = E(T_1)/\mu^2 = \sum_{r=0}^{n-1} \rho^r/\mu. \qquad (1)$$

Другие диагональные элементы получаются аналогично, однако соответствующие выражения несколько сложнее.

Отметим следующие моменты:

(a) Исходя из общих соображений, а также и из равенства (1), ясно, что имеется разница между случаем $\rho \geqslant 1$ и случаем $\rho < 1$, только в первом случае общая информация неограниченна при $n \to \infty$.

(b) Уравнения для $\tilde{\mu}$ и $\tilde{\rho}$ найти легко. Нелинейное уравнение для $\tilde{\rho}$ получается при исключении $\tilde{\mu}$. Это уравнение значительно упрощается, если можно считать, что $\rho^n \ll 1$ или $\rho^n \gg 1$.

(c) Если $\rho \geqslant 1$, то T_1—асимптотически нормальна при $n \to \infty$. Если $\rho < 1$, то непосредственно из центральной предельной теоремы, примененной к неодинаково распределенным случайным величинам Y_j, это не следует. Дело в том, что отношение $\operatorname{var}(Y_n)/\operatorname{var}(T_1)$ к нулю не стремится. Однако этот факт из пуассоновости распределения статистики T_1. Таким образом, если состоятельная оценка для ρ может быть найдена, то доверительные интервалы для μ можно получить на основе использования равенства (1).

(d) Если $\rho < 1$, то общая информация при $n \to \infty$ ограничена, данные оканчиваются длинной серией нулей. Следовательно, при $n \to \infty$ статистика T_1 не является асимптотически нормальной. Однако, можно обосновать возможность использования нормальной аппроксимации. Это делается посредством обращения к воображаемой последовательности задач, в которых $\mu \to \infty$ при фиксированных n. Решающий момент связан с тем, будет ли статистика T_1 принимать большие значения. Аналогичные соображения применимы и к T_2. Подробное обсуждение аналогичных моделей процессов чистой гибели и чистого рождения проводится в работе Бейера и др. (1976).

Чтобы построить критерий согласия, используется, в принципе, условное распределение данных при заданном значении статистик $T_1 = t_1$ и $T_2 = t_2$. Один из способов построения соответствующей статистики и получения приближенного анализа, применимого лишь в том случае, когда всего лишь несколько наблюдений равны нулю, состоит в том, чтобы заметить, что

$$E(\log Y_j) \sim \log\mu - (j-1)\log\rho, \qquad \operatorname{var}(\log Y_j) \sim 1/E(Y_j) \sim 1/Y_j,$$

и далее применить метод взвешенного анализа наименьших квадратов для линейной регрессии величин $\log Y_j$ на j. Линейность можно проверить обычным образом, включая и квадратичные члены. Предположение о пуассоновости можно проверить на основе взвешенной суммы остаточных квадратов.

[Теоретическая статистика, § 9.3; Рао, § 6; Силвей, гл. 7]

10. Байесовские методы

Необходимые сведения

В предыдущих главах неизвестный параметр θ считался неизвестной постоянной. При этом исследовались методы, относительно которых предполагалось, что они действуют успешно в повторных выборках вне зависимости от истинного значения параметра θ. В отличие от этого при байесовском подходе θ рассматривается как значение случайной величины Θ, имеющей и в отсутствие данных $f_\Theta(\theta)$ — априорную плотность. Обычно априорную плотность считают известной. В модели также задается $f_{Y|\Theta}(y|\theta)$ — плотность величины Y при заданном значении $\Theta=\theta$.

Байесовский подход имеет два преимущества. Во-первых, априорная плотность предоставляет способ введения анализируемой в дальнейшем информации. И, во-вторых, в рамках данной выше постановки окончательная информация о θ подытоживается в виде наблюдаемой условной плотности $f_{\Theta|Y}(\theta|y)$, называемой апостериорной плотностью. Поскольку при этом имеют дело со случайными величинами, то можно делать вероятностные утверждения и нет необходимости вводить специальные доводы и соображения. Затруднения, возникающие при использовании байесовского подхода, всецело касаются исходной постановки задачи: когда θ можно считать значением случайной величины с известной априорной плотностью?

Математическая задача вычисления условной плотности $f_{\Theta|Y}(\theta|y)$ решается на основе теоремы Байеса. Для случайных величин она формулируется следующим образом:

$$f_{\Theta|Y}(\theta|y)=\frac{f_{Y|\Theta}(y|\theta)\,f_\Theta(\theta)}{f_Y(y)}=\frac{f_{Y|\Theta}(y|\theta)\,f_\Theta(\theta)}{\int f_{Y|\Theta}(y|t)\,f_\Theta(t)\,dt},$$

или просто

$$f_{\Theta|Y}(\theta|y)\propto f_{Y|\Theta}(y|\theta)\,f_\Theta(\theta).$$

Заметим, что апостериорная плотность зависит от данных только через наблюденную функцию правдоподобия. Отсюда вытекает, что выводы зависят от данных только в той мере, в какой эти данные необходимы для определения соответствующих плотностей.

Когда модель принадлежит экспоненциальному семейству, то можно найти семейство априорных плотностей, для которых апостериорные плотности принадлежат тому же самому семейству; лишь определяющие их константы изменяют свои значения в зависимости от данных. Такое семейство априорных плотностей называют сопряженным или замкнутым относительно выбора для рассматриваемой задачи. Итак, если $Y_1, \ldots, Y_n$ — н. о. р. случайные величины с плотностью $\exp\{a(\theta)b(y)+c(\theta)+d(y)\}$, то априорную плотность можно взять пропорциональной $\exp\{k_1a(\theta)+k_2c(\theta)\}$. Такую конструкцию имеют многие широко распространенные задачи.

Имеются три главные интерпретации априорной плотности:

(a) как частотная плотность распределения, когда существует естественный механизм рандомизации, порождающий значение θ;

(b) как объективная мера рациональной степени уверенности, обычно на начальной стадии „незнания";

(c) как субъективная мера того, что на самом деле полагает конкретный индивидуум, т. е. „вы", о параметре θ.

Возможность (a) является бесспорной, однако относительно редко применяемой. Двум другим подходам посвящена огромная литература. Подход (b) привлекателен, однако при воплощении идеи о формальном представлении „незнания" возникают поистине большие затруднения. Можно показать, что при внутренне согласованных операциях над субъективными вероятностями выполняются обычные законы теории вероятностей, включая и теорему Байеса. Затруднения, возникающие при этом подходе, связаны частично с трудностью приписывания численных вероятностей, особенно в сложных ситуациях, а частично и с тем, что приходится на равных основаниях иметь дело с нечеткими субъективными впечатлениями и „твердой" информацией, содержащейся в данных.

Байесовский аналог интервального оценивания связан с определением апостериорной плотности неизвестного параметра с последующим интегрированием этой плотности по всем мешающим параметрам. Часто апостериорная плотность будет приближенно нормальной.

Один из байесовских аналогов критериев значимости требует приписывания ненулевой априорной вероятности тому, что нулевая гипотеза в точности верна. После этого вычисляются апостериорные шансы справедливости гипотезы H_0. Этот подход по своим результатам сильно отличается от соответствующих критериев значимости в выборочной теории.

В эмпирических байесовских задачах априорная плотность распределения считается существующей, но неизвестной. Предполагается, что ее можно оценить на основе соответствующих данных.

Задачи

10.1. Рассмотрим (n_1+n_2) экспериментальных единиц; j-я единица характеризуется неизвестным контрольным значением y_j и неизвестным значением, отвечающим воздействию $z_j = y_j + \theta_j$. Для каждой конкретной экспериментальной единицы одновременно можно наблюдать или контрольное значение, или значение, отвечающее воздействию, но не оба. Предположим, что случайно n_1 единиц выбирается для контрольных наблюдений. Выведите функцию правдоподобия и опишите ее информационное содержание относительно θ. Далее, предположим, что y_j — случайная величина с п. р. в. $f(y)$, а Θ имеет априорную плотность $g(\theta)$. Покажите, что апостериорное распределение величины Θ не совпадает с априорным распределением, несмотря на то что только из правдоподобия, без обращения к принципу повтора выбора нельзя получить информацию о параметре θ.

Решение

Пусть $A_j = 1$ или θ в соответствии с тем, находится ли j-й индивидуум в экспериментальной или контрольной группе, а X_j — наблюдаемый отклик. Тогда существуют $\binom{n_1+n_2}{n_1}$ возможных векторов $(A_1, \ldots, A_{n_1+n_2})$, которые, согласно предположению рандомизации, равновероятны. Вероятностная модель имеет вид

$$\operatorname{pr}(A=a,\ X=x \mid Y=y,\ \theta) = \begin{cases} \binom{n_1+n_2}{n_1}^{-1} & \text{при } x_j = y_j + a_j\theta,\ j=1,\ \ldots,\ n_1+n_2, \\ 0 & \text{в противном случае.} \end{cases}$$

Поскольку y_j неизвестны, за исключением тех, которые наблюдаются непосредственно, то на этой стадии их можно считать параметрами. Функция правдоподобия постоянна для всех θ, равных допустимым значениям разности $x-y$.

Отметим, что сделанное выше предположение о существовании априорной плотности на Θ не влияет на правдоподобие, если только не предполагать также существования априорного распределения и для Y. Если значения случайной величины Y предполагать независимыми с п. р. в. $f(y)$, что с практической точки зрения часто является неестественным предположением, и если Θ не зависит от Y и имеет п. р. в. $g(\theta)$, то совместная п. р. в. вектора $(A,\ X,\ Y)$ имеет вид

$$\binom{n_1+n_2}{n_1}^{-1} \left\{ \prod_{j=1}^{n_1+n_2} f(x_j - a_j\theta) \right\} g(\theta) \prod_{j:a_j=1} \delta(y_j,\ x_j-\theta).$$

Отсюда апостериорная плотность пары (Θ, Y) при заданном значении $A=a$, $X=x$ пропорциональна

$$g(\theta) \prod_{j:a_j=1} f(x_j-\theta)\,\delta(y_j,\ x_j-\theta).$$

Интегрирование по y_j приводит к апостериорной плотности случайной величины Θ, пропорциональной

$$g(\theta) \prod_{j:a_j=1} f(x_j-\theta).$$

Это можно было бы получить непосредственно из предположения, что X — отклики при применении воздействия — суть н. о. р. случайные величины с п. р. в. $f(x-\theta)$. Апостериорное и априорное распределения на Θ различны. Помимо простой рандомизационной модели, было сделано дополнительное предложение относительно наблюдений. Отметим, что для апостериорного распределения рандомизация не играет никакой роли. Как только будет предположено существование для Y п. р. в., то „маргинальное" правдоподобие для Θ получается следующим образом:

$$\int \mathrm{pr}\,(A=a,\ X=x\,|\,Y=y;\theta) \prod_{j=1}^{n} f(y_j)\,dy_j = \binom{n_1+n_2}{n_1}^{-1} \prod_{j=1}^{n} f(x_j-a_j\theta).$$

Оно пропорционально $\prod_{j:a_j=1} f(x_j-\theta)$ — традиционной форме правдоподобия.

При более реалистическом анализе данной задачи потребовалось бы, чтобы Y_j имели п. р. в. $f(y-\gamma)$ с неизвестным параметром сдвига γ или пришлось бы считать, что неизвестна плотность распределения Y_j.

Рассмотренный пример иллюстрирует некоторые затруднения, возникающие при принятии практически важного способа рандомизации в рамках определенной формальной статистической схемы. Частично вывод заключается в том, что при подходах, подобных байесовскому и требующих довольно подробной постановки задач, наблюдается тенденция к отказу от использования рандомизации, при которой оказываются разумными те предположения, что автоматически выдвигаются при байесовской постановке. Относительно обсуждения свойств рандомизации в рамках байесовского подхода см. также работу Стоуна (1969).

Представляется, что роль рандомизации более четко определяется в рамках теории выборочного подхода, и, в частности, в связи с критериями значимости. Необходимо отметить, что при выборочном подходе в действительности используемый план неотличим в одном важном отношении от исходной совокупности планов, используемых при вычислении вероятностей.

[Теоретическая статистика, § 10.2; Рао, § 5б.1; Силвей, § 9.6; Линдли, стр. 21,35 и т. д.]

10.2[1]). Пусть Y — число „единиц" среди n наблюдений, соответствующих случайному выбору без возвращения из конечной популяции объема m, содержащей θ „единиц". Таким образом, Y имеет гипергеометрическое распределение. Предположим далее, что априорное распределение величины Θ само является гипергеометрическим. Найдите апостериорное распределение величины $\Theta - y$ числа единиц, оставшихся в популяции, при заданном значении $Y = y$. Покажите, что в частном предельном случае полученное апостериорное распределение не зависит от y. Объясните вероятностный смысл последнего результата и обсудите критически возможность его использования в приемочном контроле.

Решение

Гипергеометрические вероятности, отвечающие случайной величине Y при заданном значении $\Theta = \theta$, имеет вид

$$\mathrm{pr}\,(Y = y \mid \Theta = \theta) = \binom{\theta}{y}\binom{m-\theta}{n-y}\Big/\binom{m}{n},$$
$$(0 \leqslant y \leqslant \theta;\ 0 \leqslant n - y \leqslant m - \theta).$$

Пусть гипергеометрическое априорное распределение определяется вероятностями

$$\mathrm{pr}\,(\Theta = \theta) = \binom{m}{\theta}\binom{A-m}{B-\theta}\Big/\binom{A}{B},\ (0 \leqslant \theta \leqslant m).$$

Тогда по теореме Байеса

$$\mathrm{pr}\,(\Theta = \alpha + y \mid Y = y) \propto \binom{\alpha+y}{y}\binom{m-\alpha-y}{n-y}\binom{m}{\alpha+y}\binom{A-m}{B-\alpha-y} =$$
$$= \frac{m!\,(A-m)!}{\alpha!\,y!\,(n-y)!\,(m-n-\alpha)!\,(B-\alpha-y)!\,(A-m-B+\alpha+y)!}.$$

Апостериорное распределение не будет зависеть от y только в том случае, когда $A \to \infty$, $B \to \infty$ и $A/B = \gamma$, т. е. когда Θ имеет биномиальное априорное распределение.

Если в рамках задач контроля качества из популяции, по существу бесконечной и с известной долей „единиц", берется случайная выборка объема m, то выборка без возвращения из

[1]) Улучшенный вариант.

первоначальной выборки не дает добавочной информации о элементах первоначальной выборки, не попавших во вторичную выборку. „Единицы" появляются полностью случайно.

[Теоретическая статистика, § 10.2; Рао, § 56.1; Силвей, гл. 10; Линдли, стр. 24; Муд, 1943]

10.4[1]. Пусть $Y_{jk}(j=1, \ldots, m;\ k=1, \ldots, r)$ подчиняются модели дисперсионного анализа с нормальными компонентами, т. е. $Y_{jk}=\mu+\eta_j+\varepsilon_{jk}$, где η_j и ε_{jk} все независимы и распределены нормально с нулевыми средними и дисперсиями σ_b^2 и σ_w^2 соответственно. Положим $\sigma^2=r\sigma_b^2+\sigma_w^2$. При использовании обычных обозначений минимальная достаточная статистика имеет вид $(\bar{Y}_{..}, SS_b, SS_w)$. В задаче оценивания μ сопоставляются два байесовских анализа. В первом из них используется лишь $(\bar{Y}_{..}, SS_b)$. Объясните, почему исключение статистики SS_w может казаться разумным и почему это неверно.

Для первого способа анализа, использующего лишь $(\bar{Y}_{..}, SS_b)$, априорная плотность берется пропорциональной $d\mu\, d\sigma/\sigma$, т. е. от σ_w не зависит. Получите, что апостериорное распределение параметра μ пропорционально t-распределению Стьюдента с $m-1$ степенью свободы. При втором способе анализа используется вся минимальная достаточная статистика с априорной плотностью, пропорциональной $d\mu\, d\sigma_w\, d\sigma/(\sigma_w\sigma)$ на области $\sigma \geqslant \sigma_w$. Докажите, что это приводит к более рассеянному, чем в первом случае, апостериорному распределению для μ.

Обсудите критически неожиданные свойства этого последнего результата и вытекающее из него следствие для обоснования выбора несобственных априорных распределений.

Решение

Имеем $Y_{jk}=\mu+\eta_j+\varepsilon_{jk}\,(j=1, \ldots, m; k=1, \ldots, r)$, где η_j независимы и нормально $N(0, \sigma_b^2)$ распределены, а ε_{jk} независимы и нормально $N(0, \sigma_w^2)$ распределены. Это модель дисперсионного анализа со случайными факторами. Минимальная достаточная статистика для нее получена в задаче 2.4.

Если положить $\sigma^2=r\sigma^2+\sigma_w^2$, то получим

$$f_{\bar{Y}_{..}, SS_b, SS_w}(\bar{y}_{..}, SS_b, SS_w; \mu, \sigma_b^2, \sigma_w^2)= \\ =f_{\bar{Y}_{..}, SS_b}(\bar{y}_{..}, SS_b; \mu, \sigma^2)\, f_{SS_w}(SS_w; \sigma_w^2). \quad (1)$$

В действительности случайные величины $\bar{Y}_{..}$, SS_b и SS_w независимы, случайная величина $\bar{Y}_{..}$ нормально $N\{\mu, \sigma^2/(rm)\}$ рас-

---

[1] Задача 10.3 опущена.

пределена, $SS_b = \sigma^2\chi^2_{m-1}$ и $SS_w = \sigma^2_w\chi^2_{mr-m}$. Эти результаты образуют основу классического дисперсионного анализа, в соответствии с которым, в частности, доверительные интервалы для μ основываются на t-распределении Стьюдента статистики

$$T = \sqrt{rm}\,(\bar{Y}_{..} - \mu)/\{SS_b/(m-1)\}^{1/2}. \qquad (2)$$

Далее из факторизационного равенства (1) должно следовать, что статистика SS_w является подчиненной в расширенном смысле для μ, если на параметрическое пространство не наложено ограничений. Однако $\sigma^2 \geqslant \sigma^2_w$. Поэтому весьма вероятно, что информацию о σ^2_w, содержащуюся в SS_w, можно использовать при построении выводов о σ^2.

При первом способе анализа инвариантная априорная плотность $d\mu\, d\sigma/\sigma$ является традиционной и несобственной. Стандартные выкладки показывают, что

$$f_{M\,|\,\bar{Y}_{..},\, SS_b}(\mu\,|\,\bar{y}_{..}, SS_b) \propto g_{m-1}(t), \qquad (3)$$

где $g_d(\cdot)$ — плотность t-распределения Стьюдента с d степенями свободы, а значение t определяется равенством (2).

При втором способе анализа с априорной плотностью $d\mu\, d\sigma_w\, d\sigma/(\sigma_w\sigma)$ и ограничении на область значений параметра $\sigma \geqslant \sigma_w$ информация из SS_w входит через это неравенство. Полная функция правдоподобия приводит к апостериорной плотности, пропорциональной величине

$$\sigma^{-m-1}\sigma_w^{-(r-1)m-1}\exp\left\{-\frac{rm(\bar{y}_{..}-\mu)^2}{2\sigma^2} - \frac{SS_w}{2\sigma^2_w} - \frac{SS_b}{2\sigma^2}\right\}.$$

Теперь необходимо проинтегрировать по σ и σ_w. При этом удобно сделать замену переменных:

$$u = \frac{rm(\bar{y}_{..}-\mu)^2 + SS_b}{2\sigma^2},\quad v = \frac{SS_w}{2\sigma^2_w}.$$

Якобиан равен

$$\left|\frac{\partial(u, v)}{\partial(\sigma, \sigma_w)}\right| = \frac{\partial u}{\partial\sigma}\frac{\partial v}{\partial\sigma_w} = \frac{4uv}{\sigma\sigma_w}.$$

Итак, для $v/u \geqslant SS_w\{rm(\bar{y}_{..}-\mu)^2 + SS_b\}^{-1}$ плотность $f_{M,U,V} \times$ $\times (\mu, u, v\,|\,\bar{y}_{..}, SS_b, SS_w)$ пропорциональна величине

$$\{rm(\bar{y}_{..}-\mu)^2 + SS_b\}^{-\frac{1}{2}m} = SS_w^{-\frac{1}{2}m(r-1)}u^{-\frac{1}{2}m-1}v^{-\frac{1}{2}m(r-1)-1}\times$$
$$\times e^{-u-v} = SS_b^{-\frac{1}{2}m}SS_w^{-\frac{1}{2}m(r-1)}\{1+t^2/(m-1)\}^{-\frac{1}{2}m}u^{-\frac{1}{2}m-1}\times$$
$$\times v^{-\frac{1}{2}m(r-1)}e^{-u-v}.$$

Интеграл по u и v пропорционален

$$h[(SS_b/SS_w)\{1+t^2/(m-1)\}],$$

где

$$h(x)=\int_0^x z^{\frac{1}{2}(m-1)}(1+z)^{-\frac{1}{2}m}\,dz$$

является возрастающей функцией по x. Итак,

$$f_{M\mid \bar{Y}_{..},SS_b,SS_w}(\mu\mid \bar{y}_{..},SS_b,SS_w)\propto g_{m-1}(t)\,h[(SS_b|SS_w)\times \times\{1+t^2/(m-1)\}].$$

Последний результат удивителен в том смысле, что полученной плотности соответствует бо́льшая дисперсия, чем для плотности (3), причем обе они центрированы в $\bar{y}_{..}$. Чувствуется, что привлечение к анализу случайной величины SS_w должно вести к более тонким выводам относительно μ, например более коротким доверительным интервалам. Однако произошло обратное. Здесь целесообразно отметить два момента (Стоун и Спрингер (1965)). Во-первых, выбор априорного распределения при втором способе анализа фактически обращает отношение SS_b/SS_w в нуль с безусловной вероятностью, равной единице. И, во-вторых, более „естественным“ априорным распределением для σ_w при заданных σ и μ является равномерное распределение на $[0,\sigma]$. И этим устраняются все парадоксы. Основной вывод состоит в том, что если стремиться избежать противоречий, то шаг, связанный с выбором объективного априорного распределения, окажется нетривиальным. С дальнейшими важными результатами на эту тему можно ознакомиться в работах Стоуна и Дэвида (1967) и Дэвида и др. (1975).

[Теоретическая статистика, §§ 10.3, 10.4; Рао, § 5.6; 1; Силвей, гл. 10; Линдли, стр. 55]

10.5. Пусть Y имеет нормальное $N(\mu,\sigma_0^2)$ распределение. Сопоставьте апостериорные ожидаемые значения параметра μ при заданном $Y=y$ для больших значений y, когда в качестве априорных распределений рассматривают стандартное нормальное и стандартное t-распределение Стьюдента.

Решение

Если априорное распределение параметра M нормально $N(\xi_0,v_0)$, то апостериорное распределение величины M при заданном y имеет среднее $(y/\sigma_0^2+\xi_0/v_0)(1/\sigma_0^2+1/v_0)$. Следовательно,

$$E(M \mid Y = y) - y = -\sigma_0^2 (y - \xi_0)/(\sigma_0^2 + \nu_0),$$

а это неограниченная функция от y. Этот факт иногда рассматривают в качестве аргумента против нормального априорного распределения: при нем конкретное ошибочное задание величины ξ_0 может привести к весьма существенным уклонениям для μ в апостериорном среднем.

Если $f(\mu)$ — априорная плотность для M, соответствующее среднее равно нулю, а Y — нормальна $N(\mu, 1)$ с плотностью $\phi(y - \mu)$, то получим

$$E(M \mid Y = y) - y = -\int w f(y - w)\,\phi(w)\,dw \Big/ \int f(y - w)\,\phi(w)\,dw. \quad (1)$$

Можно показать, что при соответствующих условиях на априорную плотность $f(\cdot)$ правая сторона равенства (1) стремится к нулю, когда $|y|$ возрастает.

Ниже приводится набросок доказательства для частного случая $f(w) \propto (1 + w^2)^{-a}$. Чтобы вычислить числитель правой части равенства (1), рассмотрим следующее соотношение:

$$\int w f(y - w)\,\phi(w)\,dw - f(y) \int w\phi(w)\,dw =$$

$$= f(y) \int w\phi(w) \left[\frac{(1 + y^2)^a}{(1 + (y - w)^2)^a} - 1 \right] dw. \quad (2)$$

Не так уж трудно показать, что интеграл в правой части равенства (2) стремится к нулю при $y \to \infty$. С другой целью предположим, что $y > 0$, и заметим, что:

(i) для $|w| \leqslant B$ при соответствующем выборе y выражение в квадратных скобках становится как угодно малым;

(ii) для $w < -B$ это выражение лежит между -1 и 0, а следовательно, при соответствующем выборе B интеграл мал;

(iii) для $w > B$ рассматриваемое выражение имеет максимум порядка w^{2a}, а потому при соответствующем выборе интеграл становится малым.

Аналогичные выводы можно сделать и для знаменателя выражения (i). Отсюда получим, что $E(M \mid Y = y)$ стремится к нулю при $y \to \infty$.

Результаты, полученные Дэвидом, значительно более общи и включают в себя следующие интересные выводы:

(a) для ситуации, рассмотренной выше, полное априорное распределение величины M стремится к $f(\mu - y)$, где f — плотность t-распределения Стьюдента;

(b) если $Y - \mu$ имеет t-распределение Стьюдента, а M имеет априорное распределение с п.р.в. $\phi(\mu - \xi_0)$, то для больших y получим $E(M \mid Y = y) \sim E(M) = \xi_0$.

[Теоретическая статистика, § 10.4; Рао, § 5b.1]

10.6. Прокомментируйте критически следующее суждение о связи между субъективной и частотной интерпретациями вероятности. Пусть $A_1, \ldots, A_n$ — события, которые могут произойти или не произойти, причем все события отвечают порознь большому числу совершенно разнородных ситуаций, например отдельным областям исследований. Предположим, что все они имеют для вас приближенно одну и ту же субъективную вероятность p. Далее пусть F будет событием, состоящим в том, что доля осуществления событий $A_1, \ldots, A_n$ заключена между $p-\varepsilon$ и $p+\varepsilon$ для некоторого фиксированного малого ε, например $\varepsilon = 10^{-2}$. Тогда:

(i) поскольку события A_j связаны с разнородными ситуациями, то разумно для вас считать их независимыми;

(ii) поскольку субъективные вероятности подчиняются обычным законам вероятности, то из закона больших чисел вытекает, что для достаточно больших n субъективная вероятность F близка к единице;

(iii) чтобы быть последовательным, вы должны быть готовы принять крупное пари, что F истинно, т. е. что ваши субъективные вероятности имеют гипотетическую частотную интерпретацию.

Решение

(i) То, что имеется связь между субъективной вероятностью и частотой, является составной частью теории субъективной вероятности в том виде, как она сформулирована, например, де Финетти (1972). Доводы, выдвинутые им, до некоторой степени являются вариантом сформулированных выше.

(ii) В принятой трактовке субъективных вероятностей ставится акцент на измерение того, что вы думаете о шансах на осуществление события A и не спрашивается, почему. В более формализованной теории субъективных вероятностей считается, что необходимо выделить некоторое число событий, о которых по некоторым разумным соображениям можно думать, что они имеют ту же вероятность, что и событие A. После этого для оценивания субъективной вероятности события A следует найти осуществившуюся на самом деле долю упомянутых выше событий. Такой подход противоположен подходу де Финетти.

(iii) Остается что-то парадоксальное в только что изложенных соображениях. Действительно, имеется ряд несвязанных событий, у которых, по вашему, существует одна и та же субъективная вероятность. Отсюда вытекает сильная уверенность в осуществлении в реальном мире довольно странного эмпирического события, а именно F.

[Теоретическая статистика,
§ 10.4; Линдли, § 2]

10.7[1]. Наблюдение y связано с такой постоянной θ, что представляется случайной величиной Y с плотностью $f(y;\ \theta)$. В повторных экспериментах последовательные значения параметра θ имеют известное частотное распределение. Таким образом, частные значения θ можно представить случайной величиной Θ с упомянутым распределением. Доверительная область $\mathscr{R}_\alpha(y)$ уровня $1-\alpha$ для значения θ, отвечающего наблюдению y, такова, что двойная безусловная вероятность $\text{pr}\{\Theta \in \mathscr{R}_\alpha(Y)\}$ равна $(1-\alpha)$. Доверительную область называют оптимальной в том случае, если она имеет наименьшую ожидаемую площадь. Примените общую лемму Неймана—Пирсона из задачи 4.2, чтобы показать, что оптимальная доверительная область имеет вид

$$\{\theta;\ f_{\Theta\mid Y}(\theta\mid y) \geqslant k_\alpha\},$$

где k_α — постоянная, соответствующая доверительному уровню $1-\alpha$.

Решение

Обозначим $g(y)$ маргинальную п.р.в. случайной величины Y. Отсюда совместная п.р.в. пары (Y,θ) представляется как $g(y)f(\theta\mid y)$. Пусть w — область в $\Omega\times\mathscr{Y}$, которая получается объединением соответствующих доверительных областей $\mathscr{R}_\alpha(y)$ для всех возможных значений y. При этом цель состоит в том, чтобы минимизировать $\int\limits_w g(y)\,d\theta\,dy$ при условии что $\int\limits_w g(y)f(\theta\mid y)\,d\theta\,dy=1-\alpha$. В терминах обобщенной леммы Неймана — Пирсона (см. задачу 4.2) речь идет о том, чтобы максимизировать $\int\limits_{\bar w} g(y)\,d\theta\,dy$ при условии, что $\int\limits_{\bar w} g(y)f(\theta\mid y)\,d\theta\,dy=\alpha$. Решением будет:

$$\bar w=\{(y,\theta): g(y)\geqslant k_\alpha^{-1}g(y)f(\theta\mid y)\},$$
$$w=\{(y,\theta): f(\theta\mid y)\geqslant k_\alpha\}.$$

Важно отметить, что постоянная k_α от y не зависит. Полученную доверительную область называют областью наивысшей апостериорной плотности. Однако ее размер является функцией от y

[1]) Улучшенный вариант.

и равен $1-\alpha(y)$, где $\int \alpha(y)\, g(y)\, dy = \alpha$. Для наблюдений y, которые сравнительно мало информативны относительно θ, доверительные области могут быть весьма обширными, иногда совпадающими даже со всем параметрическим пространством. Приписывание безусловного уровня α таким процедурам противоречит принципу правдоподобия.

Дальнейшие подробности вместе с примерами можно найти в работе Неймана (1952).

[Теоретическая статистика, § 10.5; Силвей, § 10.3; Линдли, стр. 18]

10.8. Пуассоновский процесс с неизвестной интенсивностью наблюдается непрерывно во времени. Априорная п.р.в. величины P имеет вид

$$\frac{t_0 (t_0\rho)^{n_0-1} e^{-t_0\rho}}{\Gamma(n_0)}.$$

При наблюдении n-е событие осуществилось в момент t. Получите апостериорную п.р.в. для P. Докажите, что вероятность отсутствия события в последующем интервале $(t, t+h)$ равна $(t_0+t)^{n_0+n}/(t_0+t+h)^{n_0+n}$. Для каких значений пары (t_0, n_0) апостериорное распределение величины P дает ответы, согласующиеся формально с ответами, получаемыми на основе доверительных интервалов?

Решение

Из теоремы Байеса следует

$$f_{P\mid Y}(\rho\mid y) \propto f_P(\rho)\, f_{Y\mid P}(y\mid\rho) = \frac{t_0 (t_0\rho)^{n_0-1} e^{-\rho t_0}}{\Gamma(n_0)} \cdot \frac{\rho(\rho t)^{n-1} e^{-\rho t}}{\Gamma(n)}.$$

После нормировки получим

$$f_{P\mid Y}(\rho\mid y) = (t_0+t)^{n+n_0} \rho^{n+n_0-1} e^{-\rho(t+t_0)}/\Gamma(n_0+n),$$

т. е. апостериорное распределение также является гамма распределением. Для события $\mathcal{E}_0$ — отсутствия событий в интервале $(t, t+h)$ — получим отсюда

$$\mathrm{pr}(\mathcal{E}_0\mid y) = \int_0^\infty \mathrm{pr}(\mathcal{E}_0\mid\rho, y)\, f_{P\mid Y}(\rho\mid y)\, d\rho =$$

$$= \int_0^\infty e^{-\rho h} f_{P\mid Y}(\rho\mid y)\, d\rho = \{(t+t_0)/(t+t_0+h)\}^{n+n_0},$$

Последнее равенство вытекает, например, из формулы для производящей функции моментов гамма распределения.

Поскольку наблюдения задаются посредством случайного интервала, определяемого T_n-моментом осуществления n-го события, то подход к задаче предсказания на основе доверительного оценивания состоит в построении интервала $X^\dagger$ от t до следующего события. Итак, $T_n = X_1 + \ldots + X_n$ — достаточная статистика. Случайные величины $X_1, \ldots, X_n, X^\dagger$ — н.о.р. с экспоненциальным распределением с параметром ρ. Наиболее простой подход к задаче заключается в том, чтобы заметить, что $R = X^\dagger/T_n$ имеет распределение, которое не зависит от ρ. А далее следует показать непосредственно или на основе распределения отношения дисперсий, что

$$\mathrm{pr}(R > r) = \int_0^\infty e^{-ru} \frac{u^{n-1}e^{-u}}{\Gamma(n)}\, du = (1+r)^{-n}.$$

Итак, нижняя граница для $X^\dagger$ на уровне α равна $h = r_\alpha^* t$, где $(1 + r_\alpha^*)^{-n} = 1 - \alpha$, т. е. $\alpha = 1 - (1 + h/t)^{-n}$. Формально это согласуется с байесовским решением при $t_0 = n_0 = 0$, т. е. для несобственной априорной плотности $f_P(\rho) \propto \rho^{-1}$. Такое соответствие результатов является общим фактом для задач о параметрах масштаба.

[Теоретическая статистика, § 10.5; 7.5; Рао, § 5в.1; Линдли, § 9]

10.10 [1]. Пусть y_1 и y_2 — числа успехов в последовательности из n_1 и n_2 испытаний Бернулли соответственно, а соответствующие вероятности успехов равны θ_1 и θ_2. Предположим, что гипотеза строгого равенства параметров $H_0: \theta_1 = \theta_2$ имеет положительную априорную вероятность и что если H_0 истинна, то Θ_1 и Θ_2 независимы и имеют априорные бета распределения. Используя подходящее априорное распределение для мешающего параметра, когда H_0 истинна, вычислите отношение априорных и апостериорных шансов для H_0.

Решение

Общие теоретические факты, связанные с рассматриваемой задачей, можно подытожить следующим образом. Предположим, что θ представляется как (ψ, λ), а гипотезой H_0 задается, что $\psi = \psi_0$. Пусть мешающий параметр Λ при гипотезе H_0 имеет в качестве априорной плотности $p_0(\lambda)$. Предположим далее,

[1] Задача 10.9 опущена

что при альтернативной гипотезе H_A пара (Ψ, Λ) имеет априорную плотность $p_A(\psi, \lambda)$. Применяя теорему Байеса, получим

$$\frac{\mathrm{pr}(H_0 \mid Y=y)}{\mathrm{pr}(H_A \mid Y=y)} = \frac{\mathrm{pr}(H_0)}{\mathrm{pr}(H_A)} \times \frac{\int p_0(\lambda) f_{Y \mid \Psi, \Lambda}(y \mid \psi_0, \lambda)\, d\lambda}{\iint\limits_{\psi \neq \psi_0} p_A(\psi, \lambda) f_{Y \mid \Psi, \Lambda}(y \mid \psi, \lambda)\, d\psi\, d\lambda}. \quad (1)$$

При разумных предположениях относительно априорной плотности случайной величины Λ, например, что она непрерывна при $\psi = \psi_0$ и что, в частности, $p_0(\lambda) \propto p_A(\psi_0, \lambda)$, отношение (1) упрощается и принимает вид:

апостериорные шансы = априорные шансы ×

$$\times \frac{f_{\Psi \mid Y, H_A}(\psi_0 \mid y, H_A)}{f_{\Psi \mid H_A}(\psi_0 \mid H_A)}, \quad (2)$$

где в качестве апостериорных плотностей взяты пределы при $\psi \to \psi_0$.

В этом конкретном случае при альтернативной гипотезе возьмем в качестве сопряженной априорной плотности на (Θ_1, Θ_2) плотность вида

$$p_A(\theta_1, \theta_2) \propto \theta_1^{a_{11}-1} (1-\theta_1)^{a_{12}-1} \theta_2^{a_{21}-1} (1-\theta_2)^{a_{22}-1}.$$

Конкретный выбор параметров a_{ij} зависит от приложений. Для n_{i1} „успехов" и n_{i2} „неудач" в выборке с номером i $(i=1, 2)$ анализ основан на апостериорной п.р.в.

$$f_{\Theta \mid Y, H_A}(\theta \mid n, H_A) = \frac{\prod\limits_{i=1}^{2} \theta_i^{m_{i1}-1} (1-\theta_i)^{m_{i2}-1}}{B(m_{11}, m_{12})\, B(m_{21}, m_{22})},$$

где $m_{ij} = a_{ij} + n_{ij}$, а $B(.,.)$ — бета функция.

Оказывается, что точная форма критерия зависит от выбора параметризации пары (ψ, λ). Соответствующий эффект проиллюстрируем на основе двух возможных параметризаций, а именно

(i) $\psi = \theta_1 - \theta_2, \quad \lambda = \frac{1}{2}(\theta_1 + \theta_2)$,

(ii) $\theta_1 = (1 + \psi\lambda)^{-1}, \quad \theta_2 = (1 + \lambda)^{-1}$.

Последний случай соответствует анализу в шкале логарифмов шансов (естественной шкале параметров).

Для параметризации типа (i) якобиан перехода равен единице. Отсюда совместное апостериорное распределение пары

(Ψ, Λ) при H_A в точке $\psi=0$ имеет вид

$$f_{\Psi\Lambda\mid Y,\,H_A}(0, \lambda \mid y, H_A)=\frac{\lambda^{m_{\cdot 1}-2}(1-\lambda)^{m_{\cdot 2}-2}}{B(m_{11}, m_{12})\,B(m_{21}, m_{22})},$$

где $m_{\cdot i}=m_{1i}+m_{2i}$.

Далее, интегрируя по λ по отрезку $0\leqslant\lambda\leqslant 1$, получим

$$f_{\Psi\mid Y,\,H_A}(0\mid y, H_A)=\frac{B(m_{\cdot 1}-1, m_{\cdot 2}-1)}{B(m_{11}, m_{12})\,B(m_{21}, m_{22})}.$$

Из соотношения (2) следует, что апостериорные шансы гипотезы H_0 против гипотезы H_A равны

$$\frac{p_0}{1-p_0}\times\frac{B(m_{\cdot 1}-1, m_{\cdot 2}-1)}{B(m_{11}, m_{22})\,B(m_{21}, m_{22})}\times\frac{B(a_{11}, a_{12})\,B(a_{21}, a_{22})}{B(a_{\cdot 1}-1, a_{\cdot 2}-1)}.$$

При параметризации типа (ii) якобиан перехода равен

$$|J|=\left|\frac{\partial(\theta_1, \theta_2)}{\partial(\psi, \lambda)}\right|=\lambda(1+\psi\lambda)^{-2}(1+\lambda)^{-2},$$

поэтому

$$f_{\Psi,\Lambda\mid Y,\,H_A}(0, \lambda\mid y, H_A)=\frac{(1+\lambda)^{-m_{\cdot\cdot}}\lambda^{m_{\cdot 2}-1}}{B(m_{11}, m_{12})\,B(m_{21}, m_{22})}.$$

Интегрирование по λ по интервалу $(0, \infty)$ дает следующее выражение:

$$f_{\Psi\mid Y,\,H_A}(0\mid y, H_A)=\frac{B(m_{\cdot 1}, m_{\cdot 2})}{B(m_{11}, m_{12})\,B(m_{21}, m_{22})}.$$

Оно отличается от полученного в случае (i), а следовательно дает другие апостериорные шансы для H_0.

С некоторых точек зрения второй путь анализа предпочтителен по крайней мере в том смысле, что логарифм отношения шансов является естественной параметризацией. Однако вызывает некоторое беспокойство с теоретической точки зрения, что ответ неединствен и что в рассматриваемом случае возможны существенные численные расхождения в решениях. Аналогичное явление происходит в задаче Стьюдента, где критерии проверки гипотез $H_0:\mu=0$ и $H_0':\mu/\sigma=0$ приводят к разным решениям. Этот факт обсуждался в работе Дики (1971).

[Теоретическая статистика, § 10.5; Линдли, стр. 30 и т. д.]

10.11. Предположим, что имеются две альтернативные модели для случайного вектора Y и обе модели зависят от мешающих параметров. Обозначим соответствующие п. р. в. как $g(y;\,\phi)$ и $h(y;\,\chi)$. Пусть априорные вероятности моделей равны соответственно π_g и π_h. Далее, при условии справедливости соответствующих моделей пусть априорными п. р. в. величин Φ и X будут $p_g(\phi)$ и $p_h(\chi)$. Воспользуйтесь асимптотическими байесовскими соображениями, чтобы показать, что отношение апостериорных вероятностей этих двух моделей приближенно равно

$$\frac{\pi_g}{\pi_h}\times\frac{g(y;\ \hat{\phi})}{h(y;\ \hat{\chi})}\times\frac{(2\pi)^{\frac{1}{2}q_g}}{(2\pi)^{\frac{1}{2}q_h}}\times\frac{p_g(\hat{\phi})}{p_h(\hat{\chi})}\times\frac{\Delta_{\phi}^{-\frac{1}{2}}}{\Delta_{\chi}^{-\frac{1}{2}}},$$

где Δ_ϕ и Δ_χ — информационные определители, связанные с оцениванием параметров ϕ и χ, а q_g и q_h — размерности параметров.

Исследуйте, в частности, случай, когда обе модели являются нормальными линейными моделями с известной дисперсией.

Прокомментируйте затруднения, возникающие при приписывании численного значения отношению $p_g(\hat{\phi})/p_h(\hat{\chi})$. Приведите обоснования того, что это отношение приближенно пропорционально величине $(\Delta_\phi/\Delta_\chi)^{\frac{1}{2}}$ по крайней мере в том случае, когда степени свободы для двух моделей равны.

Решение

Отношение апостериорных вероятностей двух моделей равно

$$\frac{\pi_g\int\limits_{-\infty}^{\infty} g(y;\ \phi)\,p_g(\phi)\,d\phi}{\pi_h\int\limits_{-\infty}^{\infty} h(y;\ \chi)\,p_h(\chi)\,d\chi}. \tag{1}$$

Центральный результат асимптотической теории состоит в том, что, когда n — число компонент наблюдений велико, логарифм правдоподобия представляется квадратической функцией в окрестности максимума:

$$\log g(y;\ \phi)=\log g(y;\ \hat{\phi})-\frac{1}{2}(\phi-\hat{\phi})^{\mathrm{T}}\,\mathcal{I}_g(\phi-\hat{\phi})+o(\|\phi-\hat{\phi}\|^2),$$

$$\log h(y;\ \chi)=\log h(y;\ \hat{\chi})-\frac{1}{2}(\chi-\hat{\chi})^{\mathrm{T}}\,\mathcal{I}_h(\chi-\hat{\chi})+o(\|\chi-\hat{\chi}\|^2),$$

где $\mathcal{I}_g$ и $\mathcal{I}_h$ — матрицы, умноженные на минус единицу вторых производных, вычисленных в точке максимума. Они имеют, как правило, порядок n. Функции $p_g(\phi)$ и $p_h(\chi)$ можно считать приближенно постоянными, если:

(i) они непрерывны и отличны от нуля для истинного значения параметра;

(ii) они удовлетворяют слабым условиям регулярности, когда второй сомножитель в подынтегральных выражениях отношения (1) мал. В этом случае, например, числитель приближенно равен

$$\pi_g p_g(\hat{\phi}) \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(\phi-\hat{\phi})^{\tau} \mathcal{I}_g (\phi-\hat{\phi})\right\} d\phi =$$

$$= \pi_g p_g(\hat{\phi}) (2\pi)^{\frac{1}{2} q_g} \Delta_{\phi}^{-\frac{1}{2}}.$$

Итак, приближенное отношение апостериорных вероятностей вычисляется.

В частности, если обе модели линейны, имеют известные дисперсии σ_g^2 и σ_h^2 и матрицы планов x_g и x_h,

$$g(y;\ \hat{\phi}) = \frac{1}{(2\pi)^{\frac{1}{2} n} \sigma_g^n} \exp\left\{-\frac{SS_g}{2\sigma_g^2}\right\},\quad \Delta_{\phi} = |\,x_g^{\tau} x_g|/\sigma_g^{2q_g},$$

где SS_g — остаточная сумма квадратов для первой модели. Таким образом, приближенное отношение апостериорных вероятностей равно

$$\frac{\pi_g}{\pi_h} \times \left(\frac{\sigma_h}{\sigma_g}\right)^n \exp\left(-\frac{SS_g}{2\sigma_g^2} + \frac{SS_h}{2\sigma_h^2}\right) \times \frac{(2\pi)^{\frac{1}{2} q_g}}{(2\pi)^{\frac{1}{2} q_h}} \times \frac{p_g(\hat{\phi})}{p_h(\hat{\chi})} \times \frac{|\,x_g^{\tau} x_g|^{-\frac{1}{2}} \sigma_g^{q_g}}{|\,x^{\tau} x_h|^{-\frac{1}{2}} \sigma_h^{q_h}}. \quad (2)$$

Если обе дисперсии неизвестны, то во второй и последний сомножители, содержащие σ_g и σ_h, подставляют оценки максимума правдоподобия. Условные априорные плотности в предпоследнем сомножителе содержат вклад для дисперсий.

В принципе, нет затруднений при вычислении отношения (2). В то же время отметим, что при приближенно равномерных априорных плотностях $p_g(\phi)$ и $p_h(\chi)$ не сокращаются, как это было бы в выкладках „в пределах“ индивидуальной модели. Обычно нет причин ожидать, что отношение $p_g(\hat{\phi})\ p_h(\hat{\chi})$ будет близким к единице. В действительности параметрические пространства, на которых определены эти плотности, в общем случае совершенно различны. С субъективной точки зрения возможно, в принципе, для любого конкретного приложения определить ваше численное значение для этого отношения. При выводе приближенных численных значений иногда полезны следующие соображения.

Хотя это и не обязательно, но часто две модели дают в грубом приближении похожие результаты. В противном случае тщательное сравнение их вряд ли необходимо. Итак, рассмотрим случай, когда $q_g = q_h$ и линейные модели в некотором приближенном смысле почти эквивалентны. Далее, если априорная плотность, например, для параметра ϕ нормальна и имеет большую дисперсию, то $p_g(\tilde{\phi}) \simeq (2\pi)^{\frac{1}{2} q_g} |\Sigma_\phi|^{-1/2}$, где Σ_ϕ — априорная ковариационная матрица. Тогда числитель в последних трех сомножителях отношения (2) равен по существу отношению квадратных корней априорной обобщенной дисперсии параметра ϕ к апостериорной обобщенной дисперсии. Можно ожидать, что это отношение будет приближенно одним и тем же для обеих моделей. Если модели в точности эквивалентны, то отношение будет в точности равным единице. Поскольку любая общая модель локально линейна, то общий вывод состоит в том, что когда $q_g = q_h$, то

$$p_g(\tilde{\phi})/p_h(\tilde{\chi}) \simeq (\Delta_\phi/\Delta_\chi)^{1/2}.$$

Если $q_g \neq q_h$ и, например, $q_g > q_h$, то приведенные доводы принимают характер рабочей гипотезы. Предположим, что первая модель репараметризована так, что первые q_h параметров приближенно эквивалентны аналогичным параметрам второй модели, а оставшиеся $q_g - q_h$ параметров ортогональны первой совокупности и придают дополнительное разнообразие первой модели. Тогда предыдущие доводы применимы к первой группе параметров. Относительно оставшихся $q_g - q_h$ параметров предположим, что априорная информация эквивалентна информации из данных рассматриваемой структуры объемом n_0. Тогда вклад в итоговый сомножитель равен $(n_0/n)^{q_g - q_h}$. Подытоживая все это для конкретного и приближенного выбора априорных плотностей, получим, что апостериорные вероятности равны

$$\frac{\pi_g}{\pi_h} \times \frac{g(y;\ \tilde{\phi})}{h(y;\ \tilde{\chi})} \times \left(\frac{n_0}{n}\right)^{q_g - q_h}.$$

Часто разумно предположить, что n_0 принимает значения от $\frac{1}{2}$ до 2. Отметим, что, когда n возрастает, возникают соображения против использования модели с большим числом параметров.

Относительно дальнейших подробностей по этой трудной и важной задаче см., например, работу Кокса (1961) и Бокса и Хилла (1967).

[Теоретическая статистика. § 10.5, 10.6]

10.12. Постройте эмпирическое байесовское решение для задачи оценивания среднего нормального распределения по ряду m совокупностей данных с различными дисперсиями. Как и в задаче 9.2, предположите, что число совокупностей данных m велико, а дисперсии $\sigma_1^2, \ldots, \sigma_m^2$ имеют в качестве априорной плотность сопряженного обратного гамма-распределения. Покажите, что мода о. м. п. апостериорного распределения среднего μ задается посредством статистики той же самой общей формы, как и $\bar{\mu}_a$ в задаче 9.2.

Решение

Предположим сначала, что дисперсия τ имеет известную априорную плотность

$$p_0(\tau) = \left(\frac{1}{2} f_0 \tau_0'\right)^{\frac{1}{2} f_0} \tau^{-\frac{1}{2} f_0 - 1} \exp\left(-\frac{1}{2} f_0 \tau_0'/\tau\right) \Gamma\left(\frac{1}{2} f_0\right),$$

где f_0 — эффективная степень свободы, а среднее $\tau_0 = \tau_0' f_0/(f_0 - 2)$. Рассмотрим далее одну совокупность случайных величин $Y_1, \ldots, Y_n$, которые для заданного τ н. о. р. случайные величины нормально $N(\mu, \tau)$ распределены. Маргинальной п. р. в. совокупности $Y_1, \ldots, Y_n$ будет

$$\int_0^\infty (2\pi)^{-\frac{1}{2} n} \tau^{-\frac{1}{2} n} \exp\left\{-\frac{\Sigma (y_j - \mu)^2}{2\tau}\right\} p_0(\tau)\, d\tau.$$

Отсюда вытекает, что $(\bar{Y}, \mathrm{MS})$ образует минимальную достаточную статистику. Если задано несколько таких независимых выборок, правдоподобие получается в виде произведения.

Опуская сомножители, не содержащие μ, полный логарифм правдоподобия получим в следующем виде:

$$-\frac{1}{2} \Sigma (f_j + f_0 + 1) \log\left\{1 + \frac{n_j (\bar{y}_j - \mu)^2}{f_j \mathrm{MS}_j + f_0 \tau_0'}\right\}, \tag{1}$$

где для j-й выборки $f_j = n_j - 1$. Один из способов понять форму логарифма правдоподобия заключается в том, чтобы заметить, что для отдельной совокупности данных статистика MS является подчиненной, имеющей не зависящее от μ распределение. Далее, $\bar{Y}$ условно относительно MS имеет вид

$$\mu + \left\{\frac{(f\,\mathrm{MS} + f_0 \tau_0')}{n(f + f_0)}\right\} T_{f+f_0},$$

где T_a обозначает случайную величину, имеющую t-распределение Стьюдента с a степенями свободы. Выражение (1) получается, по существу, из произведения плотностей t-распределений Стьюдента.

При максимизации суммы (1) получим для оценивания следующее уравнение:

$$\sum \frac{n_j (f_j + f_0 + 1)(\bar{y}_j - \hat{\mu})}{f_j \, \mathrm{MS}_j + f_0 \tau_0^2 + n_j (\bar{y}_j - \hat{\mu})^2} = 0.$$

Если f_0 и τ_0^2 неизвестны, что и бывает, как правило. Их можно оценить, например, методом максимума правдоподобия на основе совместной плотности средних квадратов внутри выборок [1].

[Теоретическая статистика, § 10.7; Рао, § 5b.4; Кокс, 1975]

---

[1] Задача 10.13 опущена.

11. Теория решений

Необходимые сведения

При формулировке проблематики гл. 10 в предложение о виде вероятностной модели для наблюдаемой случайной величины Y включалось существование априорной плотности на множестве вероятностных значений параметра. В теории статистических решений задается также множество возможных решений $\mathscr{D}$. Одно из этих решений d необходимо выбрать на основе учета данных. Кроме того, известно, что функция полезности $u(\theta, d)$ определяет выигрыш, который возникает при использовании решения d, когда θ — истинное значение параметра. Цель состоит в том, чтобы максимизировать ожидаемую полезность. Решающая функция $\delta(y)$ — это правило, определяющее конкретное решение, которое должно быть принято для любого заданного $Y=y$.

В принципе, решение поставленной задачи состоит в том, что вычисляется апостериорное распределение случайной величины Θ при заданном значении $Y=y$, находится ожидаемая условная полезность, отвечающая некоторому решению d

$$E\{u(\Theta, d) \mid Y=y\},$$

и после этого выбирается решение, максимизирующее эту ожидаемую условную полезность. Соответствующее решающее правило называют байесовским.

Для некоторых задач полезность можно выразить в денежном эквиваленте. При некоторых разумных предположениях внутренней согласованности полезность существует для любого индивидума. При согласовывании конфликтующих функций полезности возникают сложные проблемы.

Хотя вычисление байесовского решающего правила требует наличия всех определенных выше элементов, тем не менее можно рассматривать и неполностью заданные задачи решения. Функцию полезности преобразуют в функцию ущерба или функцию потерь:

$$w(\theta, d) = \sup_{d^* \in \mathscr{D}} u(\theta, d^*) - u(\theta, d).$$

Функция риска для любого решающего правила определяется как ожидаемое значение ущерба и зависит от θ:

$$r(\theta, \delta) = \Sigma_Y[w\{\theta, \delta(Y)\}; \theta].$$

Итак, решающие правила сравниваются на основе функций риска как функции θ. В общем случае простого выбора здесь не существует, однако если решающее правило имеет функцию риска, мажорирующую функцию риска некоторого другого правила, то первое правило называют недопустимым. В противном случае правило называют допустимым. Байесовское правило по отношению к собственному априорному распределению является допустимым.

Один довольно искусственный способ выбора наилучшей среди совокупности функций риска связан с характеризацией каждой функции риска ее максимумом

$$m(\delta) = \sup_{\theta \in \Omega} r(\theta, \delta).$$

После этого берётся правило, минимаксное правило, с минимальным значением $m(\delta)$. Функция риска манимаксного правила (обычно) постоянна.

В проведенном выше обсуждении речь шла об отдельной задаче решения. В приложениях часто возникает последовательность взаимосвязанных задач решения. Обычно соответствующее решение проблемы проводится рекуррентно при помощи методов динамического программирования.

Задачи

11.1. По данным y находится апостериорная п.р.в. скалярного параметра θ. Возможные решения d_t снабжены в качестве индекса вещественным числом t. Рассматриваются три возможные функции полезности, а именно:

(i) $a - b_1 (t - \theta)^2$,

(ii) $a - b_2 |t - \theta|$,

(iii) $a - b_3 (t - \theta)^3$,

где $b_k > 0$ $(k = 1, 2, 3)$.

Покажите, что байесовским решающим правилом для (i) и (ii) являются соответственно среднее и медиана апостериорного распределения. Для функции полезности (iii) им является $\mu_\theta + \varkappa_0 \sigma_\theta$, где $\varkappa_\theta^2 + 3\varkappa_0 - \gamma_{1\theta} = 0$, а μ_θ, σ_θ и $\gamma_{1\theta}$ — среднее, стандартное уклонение и асимметрия апостериорного распределения соответственно. Объясните качественно смысл различия между ответами для функций полезности (i) и (ii).

Решение

При заданном значении y параметр θ является значением случайной величины Θ, имеющий известное распределение — апостериорное распределение. Для функции полезности типа (i) и $\bar{\pi}$ равна $O(1/\sqrt{\pi})$ по вероятности, а разность ожидаемых полезностей решающих правил, использующих π и $\bar{\pi}$, стремится к нулю при $n \to \infty$. Более простой оценкой в сравнении с $\bar{\pi}$ является несмещенная оценка $\bar{\pi} = \frac{1}{2}(1 + \overline{Y}.)$, принимающая значения из интервала $[0, 1]$.

Если ставить задачу получить решение по одному значению y в отдельности, то будет разумно принимать решение d_1 тогда и только тогда, когда $y > 0$. Очевидно, что существенное улучшение возможно при рассмотрении задачи на основе всех данных, по крайней мере тогда, когда доля „единичек" не слишком близка к $\frac{1}{2}$.

В другом варианте задачи существуют уже наблюденные данные $y_1, \ldots, y_n$, а решение относительно соответствующего популяционного среднего μ_{n+1} необходимо принять на основе будущего наблюдения y_{n+1}. В качестве крайнего примера правила, использующего оценку $\bar{\pi}$, приведем следующий. Если $y_1 = 0$, то при учете лишь этого наблюдения, неразумно предполагать, что $\mu_1 = -1$. Однако, в сочетании с $y_2 > 0$ это указывает на то, что оба средних положительны.

Подобные эмпирические байесовские задачи впервые изучались Роббинсом [1951]; см. также недавние работы Мэритца [1970] и Копаса [1969].

[Теоретическая статистика, § 11.3, 10.6; Рао, § 5b; Линдли, стр. 64]

11.3. Чтобы представить задачу интервального оценивания на языке теории принятия решения, предполагается, что для скалярного параметра θ функцией ущерба, соответствующей интервалу $[t_1, t_2]$, значение t выбирается так, чтобы минимизировать $E\{(\Theta - t)^2\} = E(\Theta^2) - 2tE(\Theta) + t^2$. Дифференцирование показывает, что максимум ожидаемой полезности достигается при $t = E(\Theta) = \mu_0$, т. е. при среднем значении апостериорного распределения.

Аналогичные вычисления для функции полезности типа (ii) приводят к медиане апостериорного распределения случайной величины Θ. Минимизация проводится на основе простых вычислений. Для функции полезности типа (iii) необходимо минимизировать

$$E\{(\Theta - t)^4\} = E[\{(\Theta - t) + (t - \mu_\theta)\}^4] =$$
$$= E\{(\Theta - \mu_0)^4\} + 4\gamma_{1\theta}\sigma_\theta^3(\mu_0 - t) + 6\sigma_\theta^2(\mu_0 - t)^2 + (\mu_\theta - t)^4.$$

Соответствующее минимизируемое значение t удовлетворяет уравнению

$$-4\gamma_{1\theta}\sigma_\theta^3-12\sigma_\theta^2(\mu_\theta-t)-4(\mu_\theta-t)^3=0.$$

Если положить $t=\mu_\theta+\varkappa_\theta\sigma_\theta$, то получим $\varkappa_\theta^3+3\varkappa_\theta-\gamma_{1\theta}=0$. Необходима осторожность при выборе правильного корня. Заметим, что если $\gamma_{1\theta}=0$, $\varkappa_\theta=0$, и вновь имеют дело с апостериорным средним.

Различие в ответах на качественном уровне объясняется относительной зависимостью решения задачи минимизации от поведения хвостов распределения случайной величины Θ.

[Теоретическая статистика, § 11.3; Рао, § 5b; Силвей, § 11.6]

11.2. Предположим, что Y имеет нормальное $N(\mu, 1)$ распределение, а μ с априорной вероятностью π равно 1, а с априорной вероятностью $1-\pi$ равно -1. Найдите байесовское правило для различения двух возможных значений μ, когда используется функция ущерба со значениями нуль — один.

Предположим теперь, что условно $Y_1, \ldots, Y_n$ независимы и нормально $N(\mu_1, 1), \ldots, N(\mu_n, 1)$ распределены, а μ_j независимо распределены в соответствии с определенным выше двухточечным априорным распределением. Постройте эмпирическое байесовское дискриминационное правило оценивания π из данных. Исследуйте критически качественные количественные свойства этого правила.

Решение

При заданном значении $Y=y$ случайная величина M, представляющая μ, такова, что

$$\frac{\mathrm{pr}(M=1\mid Y=y)}{\mathrm{pr}(M=-1\mid Y=y)}=\left(\frac{\pi}{1-\pi}\right)\exp\left\{-\frac{1}{2}(y-1)^2+\frac{1}{2}(y+1)^2\right\}=\frac{\pi e^{2y}}{1-\pi}.$$

Отсюда вытекает, что решение d_1, соответствующее $M=1$, берется в том и только в том случае, если $y>\frac{1}{2}\log\{(1-\pi)/\pi\}$.

Маргинальная п. р. в. случайной величины Y имеет вид

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2-\frac{1}{2}\right)\{\pi e^y+(1-\pi)e^{-y}\}.$$

Параметр π на основе наблюдений $y_1, \ldots, y_n$ можно оценить методом максимума правдоподобия как $\tilde\pi$. Разность между π

$$w(\theta;\ t_1, t_2)=\begin{cases} a(t_1-\theta)+(t_2-t_1) & (\theta<t_1),\\ (t_2-t_1) & (t_1\leqslant\theta\leqslant t_2),\\ b(\theta-t_2)+(t_2-t_1) & (t_2<\theta),\end{cases}$$

где $a, b > 0$. Рассматривая для простоты ситуацию, где апостериорное распределение величины Θ является нормальным $N(\bar{y}., \sigma_0^2/n)$, исследуйте связь между эффективным доверительным интервалом a и b.

Обобщите рассуждения так, чтобы можно было построить интервальную оценку для среднего нормального распределения с неизвестной дисперсией. Используйте сопряженные априорные распределения и масштабно-инвариантную функцию ущерба.

Решение

Предложенная функция ущерба выражает одновременно два момента:

(i) чем длиннее интервал, тем менее он полезен;

(ii) возникают некоторые потери, если истинное значение параметра лежит вне интервала, и чем оно дальше от этого интервала, тем больше потери.

Ожидаемый ущерб равен

$$(t_2-t_1)+a\int_{-\infty}^{t_1}(t_1-\theta)f_\Theta(\theta\mid y)\,d\theta+b\int_{t_1}^{\infty}(\theta-t_2)f_\Theta(\theta\mid y)\,d\theta,$$

где $f_\Theta(\theta\mid y)$ — апостериорная плотность случайной величины Θ. Необходимо выбрать t_1 и t_2, минимизирующие ожидаемый ущерб. Дифференцирование приводит к уравнениям для стационарного значения

$$-1+a\int_{-\infty}^{t_1}f_\Theta(\theta\mid y)\,d\theta=0,\quad 1-b\int_{t_2}^{\infty}f_\Theta(\theta\mid y)\,d\theta=0.$$

Отсюда вытекает, что t_1 и t_2 являются квантилями апостериорного распределения уровня $1/a$ и $(b-1)/b$ соответственно. Таким образом, требуется выполнение условий $a, b > 1$ и $1/a+1/b \leqslant 1$. Действительно, если указанные условия не выполнены, то будет лучше взять интервал нулевой длины в точке $t_1=t_2$, соответствующей квантили уровня $b/(a+b)$, и примириться с возникающими здесь потерями.

Отметим, что рассмотренные выше соображения применимы и к нормальным моделям вне зависимости от того, известно или не известно σ, лишь бы функция ущерба задавалась независимо от σ. Часто, однако, наилучшая аппроксимация состоит в том, чтобы за ущерб взять заданную функцию, умноженную на σ^{-1}. Действительно, тогда наказание за принятие длинного доверительного интервала будет зависеть от его длины, поделенной на σ. В таком случае существует чрезвычайно простое обобщение полученных ранее результатов. Апостериорное распределение среднего случайной величины Θ и стандартное уклонение σ теперь уже используются. Ожидаемый риск, который необходимо минимизировать, имеет вид

$$(t_2-t_1)E(\Sigma^{-1}|Y=y)+a\int_0^\infty d\sigma\int_{-\infty}^{t_1}\left(\frac{t_1-\theta}{\sigma}\right)f(\theta,\sigma|y)\,d\theta+$$
$$+b\int_0^\infty d\sigma\int_{t_2}^\infty\left(\frac{\theta-t_2}{\sigma}\right)f(\theta,\sigma|y)\,d\theta.$$

Уравнение для t_1, например, принимает вид

$$-E(\Sigma^{-1}|Y=y)+a\int_{-\infty}^{t_1}d\theta\int_0^\infty\frac{d\sigma}{\sigma}f(\theta,\sigma|y)=0.$$

Существует много частных случаев. Например, когда Σ не зависит от Θ, когда апостериорная плотность имеет сдвигово-масштабный тип. Явные формулы для нормальной модели получены Эйтчисоном и Дансмором [1968].

[Теоретическая статистика, § 11.3, 11.7; Рао, § 5b; Силвей, гл. 10 и 11; Линдли, стр. 57]

11.4. Имеются m нормальных распределений с дисперсией σ^2, средние которых $\mu_1, \ldots, \mu_m$ — независимые случайные величины, имеющие в качестве априорного нормальное $N(\xi_0, v_0)$ распределение. Для каждого из распределений проводится r независимых наблюдений. Распределение с наибольшим наблюдением средним выбирается как "лучшее". Полезность, получающаяся при выборе популяции со средним μ, равна $a\mu$. Найдите сначала, например, условную ожидаемую полезность при заданном наибольшем наблюденном среднем. Покажите, что ожидаемая полезность процедуры равна

$$a\xi_0+\frac{ag_{mm}v_0}{(v_0+\sigma_0^2/r)^{1/2}},$$

где g_{mm} — ожидаемое значение наибольшего среди m наблюдений из стандартного нормального распределения.

Решение

Если выборка имеет среднее $\bar{y}$, то соответствующее популяционное среднее имеет апостериорное распределение со средним

$$\frac{\xi/v_0+\bar{y}/(\sigma_0^2/r)}{1/v_0+1/(\sigma_0^2/r)}.$$

Таким образом, если выбирается именно это среднее, то ожидаемая полезность равна полученной величине, умноженной на a. Далее, выборочные средние $\overline{Y}_1, \ldots, \overline{Y}_m$ в качестве маргинальных имеют нормальные распределения со средним ξ_0 и дисперсией $\nu_0+\sigma_0^2/r$. Таким образом, ожидаемое значение выбираемого выборочного среднего равно $\xi_0+g_{mm}(\nu_0+\sigma_0^2/r)^{1/2}$. Ожидаемая полезность рассматриваемой процедуры равна, следовательно,

$$a\,\frac{\xi_0/\nu_0+\{\xi_0+g_{mm}(\nu_0+\sigma_0^2/r)^{1/2}\}/(\sigma_0^2/r)}{1/\nu_0+1/(\sigma_0^2/r)},$$

что сводится к требуемой форме.

[Теоретическая статистика, § 11.3; Рао, § 5b; Силвей, гл. 10 и 11]

11.5. Пусть $S=\Sigma b(Y_j)/n$ — достаточная статистика для однопараметрического экспоненциального семейства с $\mu(\theta)=E(S;\ \theta)$. Используя нижнюю границу Крамера — Рао из гл. 8, покажите, что любая оценка $T=t(Y)$, имеющая равномерно меньший риск, чем статистика S при оценивании $\mu(\theta)$ с функцией ущерба вида $\{t-\mu(\theta)\}^2$, должна быть несмещенной. Получите, таким образом, что S допустима.

Решение

Статистика S достигает нижней границы Крамера — Рао для несмещенных оценок. Поэтому если T имеет смещение $b(\theta)$, то

$$\operatorname{var}(T;\ \theta)\geqslant\{1+b'(\theta)\}^2\operatorname{var}(S;\ \theta).$$

Следовательно, поскольку среднеквадратическая ошибка статистики T равна

$$\operatorname{var}(T;\ \theta)+\{b(\theta)\}^2,$$

то из того, что T имеет равномерно меньшую среднеквадратическую ошибку, чем S, вытекает, что по крайней мере

$$\{1+b'(\theta)\}^2\operatorname{var}(S;\ \theta)+\{b(\theta)\}^2\leqslant\operatorname{var}(S;\ \theta). \qquad (1)$$

Причем для некоторых θ достигается строгое неравенство. Следовательно, $-1\leqslant 1+b'(\theta)\leqslant 1$ со строгим неравенством для некоторых θ, т. е. $-2\leqslant b'(\theta)\leqslant 0$. Однако тот же самый результат должен иметь место для любого преобразования параметра $\psi(\theta)$. Это означает, что $-2\leqslant b'(\theta)\,d\theta/d\psi\leqslant 0$ для всех $\psi(\theta)$. Очевидно, что это невозможно, за исключением случая, когда $b'(\theta)=0$ всюду. В последнем случае в неравенстве (1) имеет место строгое равенство и $b'(\theta)=0$ всюду.

[Теоретическая статистика, § 11.6; Рао, § 5b; Силвей, § 2.10 Гиршик и Сэвидж, 1951]

11.6. Предположим, что Y нормально $N(\mu, \sigma_0^2)$ распределена и что μ имеет нормальное $N(0, v_0)$ априорное распределение. Здесь и σ_0^2, и v_0 известны. Если риск пропорционален среднеквадратической ошибке, то байесовская оценка δ_B для μ имеет минимальный байесовский риск, однако ее функция риска неограниченна, тогда как о. м. п. $\delta_{ML}(Y)$ имеет минимаксный риск. Компромиссная оценка имеет вид

$$\delta_C(Y) = \begin{cases} \delta_{ML}(Y) + a & (Y < -k), \\ \delta_B(Y) & (-k \leqslant Y \leqslant k), \\ \delta_{ML}(Y) - a & (Y > k), \end{cases}$$

где a и k выбраны так, чтобы сделать $\delta_C(\cdot)$ непрерывной функцией. В остальном они произвольны. Сравните байесовский риск и функцию риска оценки $\delta_C(Y)$ с функциями риска о. м. п. и байесовской оценки. Прокомментируйте возможность применения такой компромиссной оценки.

Опишите и исследуйте соответствующую компромиссную оценку между о. м. п. и оценкой Стейна для вектора нормальных средних.

Решение

Оценка максимума правдоподобия и байесовская равны соответственно

$$\delta_{ML} = Y, \quad \delta_B = Y\{1 - (1 + \gamma_0)^{-1}\}, \quad \gamma_0 = v_0/\sigma_0^2.$$

В предположении, что потери квадратические, функции риска примут вид

$$r(\mu;\ \delta_{ML}) = \sigma_0^2,$$
$$r(\mu;\ \delta_B) = \mu^2 (1 + \gamma_0)^{-2} + \gamma_0^2 (1 + \gamma_0)^{-2} \sigma_0^2.$$

Заметим, что

$$r(\mu;\ \delta_B) \leqslant r(\mu;\ \delta_{ML}) \text{ для } \mu^2 \leqslant \sigma_0^2 + 2v_0,$$

откуда вытекает, что функция риска для правила $\delta_C(Y)$ может иметь хорошие свойства. Соответствующие байесовские риски оценок δ_{ML} и δ_B имеют вид

$$r_B(\delta_{ML}) = \sigma_0^2,$$
$$r_B(\delta_B) = \sigma_0^2 (1 + \gamma_0)^{-1} \gamma_0.$$

Рассматривая компромиссную оценку $\delta_C(Y)$ отметим, что из соотношения $\delta_{ML}(k)-a=\delta_B(k)$ вытекает, что $k=a(\gamma_0+1)$. Положим $a=c\sigma_0$ и будем считать при проведении вычислений, что $\sigma_0=1$. Тем самым у функций риска отбрасывается сомножитель σ_0^2. Вычислим сначала байесовский риск оценки $\delta_C(Y)$, которую удобно представить в следующем виде:

$$\delta_C(Y)=\left\{1-\psi\left(\frac{Y^2}{1+\gamma_0}\right)\right\}Y, \quad \psi(u)=$$

$$=\begin{cases}(1+\gamma_0)^{-1} & (0\leqslant u\leqslant c^2(1+\gamma_0)),\\ (c\sqrt{y}+1)/\sqrt{u} & (c^2(1+\gamma_0)<u).\end{cases}$$

Итак,

$$(r_B(\delta_C)=E\{M-\delta_C(Y)\}^2=E\{M-\delta_B(Y)+\delta_B(Y)-\delta_C(Y)\}^2=$$

$$=\frac{\gamma_0}{\gamma_0+1}+\frac{1}{\gamma_0+1}E\left[\left(\frac{Y^2}{\gamma_0+1}\right)\left\{1-(\gamma_0+1)^2\psi\left(\frac{Y^2}{\gamma_0+1}\right)\right\}^2\right].$$

Далее, величина $Y^2(\gamma_0+1)$ имеет хи-квадрат распределение с одной степенью свободы. Для любой функции $h(\cdot)$ имеет место тождество $E\{\chi_1^2h(\chi_1^2)\}=E\{h(\chi_3^2)\}$. Поэтому

$$r_B(\delta_C)=\frac{\gamma_0}{\gamma_0+1}+\frac{1}{\gamma_0+1}E[\{1-(\gamma_0+1)^2\psi(\chi_3^2)\}^2].$$

Ожидание вычисляется довольно просто, и в результате получаем

$$r_B(\delta_C)=\frac{\gamma_0}{\gamma_0+1}+\frac{2}{\gamma_0+1}\{(b^2+1)\Phi(-b)-b\varphi(b)\},\ b=c(\gamma_0+1)^{1/2}.$$

Когда $b\simeq 0{,}4$, то $r_B(\delta_C)$ лежит посредине между $r_B(\delta_{ML})$ и $r_B(\delta_B)$.

Функцию риска $r(\mu;\ \delta_C)$ получают простым интегрированием величины $\{\delta_C(Y)-\mu\}^2$ по мере нормального $N(\mu,\ 1)$ распределения. После простых выкладок получим

$$r(\mu,\ \delta_C)=1+c^2+g(\mu^2),$$

где функция $g(\mu^2)$ отрицательна. Она стремится при $\mu^2\to\infty$ к своему максимальному значению, равному нулю. Таким образом, риск ограничен и, естественно, максимален, когда $\delta_C(Y)=\delta_{ML}(Y)-a$.

Все сказанное можно обобщить на случай, когда вектор нормально $MN_p(\mu,\ I)$ распределен, а следовательно, с помощью канонических преобразований и на случай общей линейной модели с известной дисперсией. Когда параметры $\mu_1,\ \ldots,\ \mu_p$ нормально $N(0,\ v_0)$ распределены, то байесовское правило вновь имеет вид

$$\delta_B=\{1-(\gamma_0+1)^{-1}\}Y,$$

а эмпирический байесовский аналог — это оценка Джеймса — Стейна

$$\delta_{ST} = \{1 - (p-2)/\Sigma Y_j^2\} Y,$$

функция риска которой при потерях, равных сумме квадратов ошибок, всегда меньше функции риска оценки δ_{ML} при $p \geqslant 3$. По аналогии с $\delta_C(Y)$ строятся оценки отдельно для каждой координаты. Цель состоит в том, чтобы ограничить покоординатные среднеквадратические ошибки, которые могут быть столь же велики, как $\Sigma\mu_j^2$. Одна из таких "компромиссных" оценок

$$\delta_{CST}(Y) = \left[1 - \frac{p-2}{\Sigma Y_j^2}\psi\left\{\frac{(p-2)Y_i^2}{\Sigma Y_j^2}\right\}\right] Y_i \quad (i = 1, \ldots, p),$$

где $\psi(u) = \min(1, d/\sqrt{u})$. Оценка $\delta_{CST}(Y)$ подробно исследовалась Эфроном и Моррисом [1972]. Они обсуждали также случай с неизвестной дисперсией σ^2. Из численных расчетов, проведенных ими, можно заключить на качественном уровне, что довольно эффективной границы можно добиться для функции риска одной компоненты, в то время как для многих из оставшихся компонент будут обладать преимуществами оценки Джеймса — Стейна.

[Теоретическая статистика, § 11.6; Рао, § 5b; Линдли, стр. 49]

11.7. Пусть н. о. р. случайные величины $Y_1, \ldots, Y_n$ нормально $N(\mu, \sigma^2)$ распределены; μ и σ^2 неизвестны. Сдвигово-инвариантные оценки параметра σ^2 являются с необходимостью функциями максимального инварианта $SS = \Sigma(Y_j - \overline{Y}.)^2$. Докажите, что любая такая инвариантная оценка недопустима для функции ущерба $w(\sigma^2, d) \propto (d - \sigma^2)^2$. Для этого показывается, что оценка

$$T = \min\left(\frac{SS}{n+1}, \frac{SS + n\overline{Y}^2}{n+2}\right)$$

имеет равномерно меньший риск, чем оценки $SS/(n+1)$. Объясните на качественном уровне, почему оценка T может быть хорошей оценкой.

Решение

Можно показать, что оценка $SS/(n+1)$ минимизирует ожидаемый ущерб среди всех оценок, зависящих только от SS.

Величины SS и $n\overline{Y}^2$ распределены соответственно как $\sigma^2\chi_{n-1}^2(0)$ и $\sigma^2\chi_1^2(\lambda)$ случайные величины, где $\lambda = n\mu^2/\sigma^2$, а через

$\chi^2_m(\lambda)$ обозначена хи-квадрат случайная величина с m степенями свободы и параметром нецентральности λ, т. е. ожиданием $m+\lambda$. Стейн (1964) рассмотрел более общую ситуацию, включив оценки вида

$$T^* = \min\left\{\frac{\sigma^2\chi^2_{\nu_1}(0)}{\nu_1+2}, \frac{\sigma^2\chi^2_{\nu_1}(0)+\sigma^2\chi^2_{\nu_2}(\lambda)}{\nu_1+\nu_2+2}\right\},$$

которые равномерно лучше оценок $T^\dagger = \sigma^2\chi_{\nu_1}(0)/(\nu_1+2)$. Решение, приведенное ниже, распространяется непосредственно и на этот случай.

Обозначим SS и nY^2 через W и Z и, не теряя общности, положим $\sigma^2=1$. Тем самым у функции риска будет отброшен сомножитель σ^4. Далее, при $\nu=n-1$ можно написать

$$T=\min\left(\frac{W}{W+Z}, \frac{\nu+2}{\nu+3}\right)\left(\frac{W+Z}{\nu+2}\right).$$

Используем теперь представление $Z=E\{\chi^2_{1+2J}(0)\}=E(Z_J)\}$, где J — случайная величина, имеющая пуассоновское распределение со средним $\frac{1}{2}\lambda$. Функцию риска обычной оценки можно выразить следующим образом:

$$E\left\{\left(\frac{W}{W+Z}\right)\left(\frac{W+Z}{\nu+2}\right)-1\right\}^2 = \\ = EE\left[\left\{\left(\frac{W}{W+Z}\right)\left(\frac{W+Z}{\nu+2}\right)-1\right\}^2\middle| J, \frac{W}{W+Z}\right].$$

Но поскольку $W+Z_J$ и $W/(W+Z_J)$ независимы, то условное ожидание равно

$$\left(\frac{W}{W+Z_J}\right)^2\frac{(\nu+3+2J)(\nu+5+2J)}{(\nu+2)^2}-2\left(\frac{W}{W+Z_J}\right)\frac{(\nu+3+2J)}{(\nu+2)}+1= \\ =\frac{(\nu+3+2J)(\nu+5+2J)}{(\nu+2)^2}\left\{\left(\frac{W}{W+Z_J}\right)-\left(\frac{\nu+2}{\nu+5+2J}\right)\right\}^2+\frac{2}{\nu+5+2J}.$$

В то же время

$$\left|\min\left(\frac{W}{W+Z_J}, \frac{\nu+2}{\nu+3}\right)-\left(\frac{\nu+2}{\nu+5+2J}\right)\right|\leqslant\left|\frac{W}{W+Z_J}-\frac{\nu+2}{\nu+5+2J}\right|$$

для всех значений величин $W/(W+Z_J)$ и J. Отсюда следует, что соответствующие условные, а следовательно и безусловные риски оценки T, по крайней мере столь же малы.

Точное выражение для функции риска можно выписать через нецентральное F-распределение. Заметим, что если n и (или) μ/σ весьма велики, так что λ весьма велико, то с большой вероятностью случайная величина J принимает большие значения, а следовательно T сведется к обычной оценке SS/$(n+1)$.

Можно непосредственно показать, что функция риска оценки T является возрастающей функцией от λ с максимумом, равным

$$\left\{1-\left(\frac{n-1}{n+1}\right)^2\right\}\sigma^4.$$

Необходимо сделать несколько замечаний относительно недопустимости инвариантной оценки и превосходства оценки T. Во-первых, напомним, что SS не является достаточной статистикой для σ^2. Поэтому использование в явном виде информации из Y не должно быть удивительным. Качественная форма статистики T представляется довольно разумной. Дело в том, что если при t-критерии Стьюдента гипотеза $\mu=0$ представляется вероятной, то можно обосновать, что $n\overline{Y}^2$ является почти несмещенной для σ^2. Далее, если известно, что $\mu=0$, то $(SS+n\overline{Y}^2)/(n+2)$ — оценка для σ^2 с минимальной среднеквадратической ошибкой. Если μ/σ сильно отличается от нуля, то T почти всегда совпадает с инвариантной оценкой. Отметим, что если известно, что μ близко к m, то для того, чтобы получить преимущество от использования оценки T, необходимо $\overline{Y}.$ заменить на $\overline{Y}.-m$. Другими словами, при полном отсутствии сведений о μ/σ использование оценок, подобных T, не приводит к большому уменьшению среднеквадратической ошибки. В этом случае инвариантность может оказаться разумным требованием. Было бы неблагоразумным считать, что теоретические результаты не имеют практической значимости. Аналоги статистики T часто используются в связи с задачами дисперсионного и множественного регрессионного анализа.

[Теоретическая статистика, § III.6, 11.8; Рао, § 5.b]

11.8. Для сравнения двух способов воздействия имеется n_0 парных наблюдений, ведущих к нормальному апостериорному распределению для разности средних откликов двух способов воздействия. При этом делаются обычные предположения нормальной теории. Имеется возможность брать дальнейшие наблюдения, и число таких наблюдений необходимо определить. Цена n_1 дальнейших наблюдений равна $k_0+k_1n_1$ ($n_1=1, 2, \ldots$). В задаче выбора лучшего среди двух способов воздействий найдите выражение для оптимального n_1 в предположении, что полезности линейны относительно средних откликов способов воздействия.

Решение

Если средние откликов двух способов воздействия равны μ_1 и μ_2 соответственно, если априорные распределения для M_1+M_2 и $\Delta=M_2-M_1$ независимы и, наконец, если разность полезностей решений d_2 и d_1 равна $a+b\delta$, то анализ можно основывать на разностях откликов двух способов воздействия. Если наблюденное среднее разности откликов воздействий равно x и построено оно по n парам наблюдений, то апостериорное распределение величины Δ является нормальным со средним и дисперсией соответственно

$$\left(\frac{x}{\sigma_0^2/n}+\frac{\xi_0}{\nu_0}\right)\Big/\left(\frac{1}{\sigma_0^2/n}+\frac{1}{\nu_0}\right) \text{ и } \left(\frac{1}{\sigma_0^2/n}+\frac{1}{\nu_0}\right)^{-1}.$$

Здесь используются обычные обозначения и предположение об априорной нормальности разности Δ.

Чтобы определить оптимальное n_1, начнем с конца. Пусть n фиксировано. Тогда оптимальное решение зависит от $a+$ $+bE(\Delta\,|\,x)$. Условная ожидаемая полезность при заданном $X=x$ равна

$$\max\left\{a+b\left(\frac{x}{\sigma_0^2/n}+\frac{\xi_0}{\nu_0}\right)\Big/\left(\frac{1}{\sigma_0^2/n}+\frac{1}{\nu_0}\right),\ 0\right\}.$$

Маргинальное распределение величины X нормально со средним ξ_0 и дисперсией $\nu_0+\sigma_0^2/n$. Поэтому безусловная ожидаемая полезность получается, если взять соответствующие ожидания. После упрощений получим

$$(a+b\xi_0)\Phi\left\{\frac{a+b\xi_0}{b\nu_0(\nu_0+\sigma_0^2/n)^{1/2}}\right\}+\left(a+\frac{b\xi_0\sigma_0^2/n}{\nu_0+\sigma_0^2/n}\right)\phi\times$$
$$\times\left\{\frac{a+b\xi_0}{b\nu_0(\nu_0+\sigma_0^2/n)^{1/2}}\right\}.$$

Если $n=n_0$, то придем к "общей" ожидаемой полезности. Если $n=n_0+n_1$, то стоимость $k_0+k_1n_1$ необходимо вычесть. Оптимум можно найти численно. Более детальный разбор задачи содержится в работе Гранди и др. (1956).

[Теоретическая статистика, § 11.9; Рао, § 5b; Силвей, гл. 11; Линдли, стр. 29]

Список литературы

В качестве источников сведений общего характера по проблемам, затронутым в каждой из задач, там, где это необходимо, отсылаем к пяти перечисленным ниже книгам. Остальные литературные ссылки касаются первоисточников и конкретных деталей отдельных задач.

Кокс и Хинкли (Cox D. R., Hinkley D. V.)
(1974) Theoretical Statistics. London, Chapman and Hall. (Ссылки переводятся по титульному листу. Русский перевод: Кокс Д., Хинкли Д. Теоретическая статистика.— М.: Мир, 1978.)
Леман (Lehmann E. L.)
(1959) Testing Statistical Hypothesis. New York, Wiley (Русский перевод: Леман Э. Проверка статистических гипотез.— М.: Наука, 1966.)
Линдли (Lindley D. V.)
(1971) Bayesian Statistics, a Review. Philadelphia, S.I.A.M.
Рао (Rao C. R.)
(1973) Linear Statistical Interference and its Applications, 2nd edition, New York, Willey. (Русский перевод 1-го издания: Рао С. Р. Линейные статистические методы и их применения.— М.: Наука, 1968.)
Силвей (Silvey S. D.)
(1975) Statistical Inference. London, Chapmann and Hall.
Абрамовиц и Стиган (Abramowitz M., Stegan I. A.)
(1965) Handbook of mathematical functions. New York, Dover. (Русский перевод: Справочник по специальным функциям. Под ред. Абрамовиц М., И. Стиган. Пер. с англ.— М.: Наука, 1979.)
Андерсен (Andersen E. B.)
(1970) Asymptotic properties of conditional maximum-likelihood estimators, J. R. Statist. Soc., B, 32, 283—301.
Андерсон (Anderson T. W.)
(1958) An introduction to multivariate statistical analysis, New York, Wiley. (Русский перевод: Андерсон Т. Введение в многомерный статистический анализ.— М.: Физматгиз, 1963.)
Андерсон и Гудмен (Anderson T. W., Goodman L. A.)
(1957) Statistical inference about Markov chains, Ann. Math. Statist., 28, 89—110.
Атикулла (Atiqullah M.)
(1962) The estimation of residual variance in quadratically balanced leastsquares problems and the robustness of the F-test, Biometrica, 49, 83—91.
(1970) A method for discriminating between models (with discussion), J. R. Statist. Soc., B, 32, 323—353.
Барнард (Barnard G. A.)
(1947) Review of book Sequential analysis by Wald A., J. Amer. Statist. Assoc., 42, 658—664.
(1963) The logic of least squares, J. R. Statist. Soc., B, 25, 124—127.
Барнард и Спротт (Barnard G. A., Sprott D. A.)
(1971) A note on Basu's examples of anomalous ancillary statistics (with discussion). In Foudations of statistical inference, pp. 163—176, eds. Godambe V. P. and Sprott D. A., Toronto, Holt, Rinehart and Winston.
Бартлетт (Bartlett M. S.)
(1936a) The information available in small samples, Proc. Camb. Phil. Soc., 32, 560—566.
(1966) An introduction to stochastic processes, 2nd edition, Cambridge University Press. (Русский перевод: Бартлетт М. Введение в теорию случайных процессов.— М.: ИЛ, 1958.)
Басу (Basu D.)
(1955) On statistics independent of sufficient statistics, Sankhya, 20, 223—226.
(1964) Recovery of ancillary information, Sankhya, A, 26, 3—16.
Бахадур (Bahadur R. R.)
(1954) Sufficiency and statistical decision functions, Ann. Math. Statist., 25, 423—462.
Бейер, Кейдинг и Симонсен (Beyer J. E., Keiding N., Simonsen W.)
(1976) The exact behaviour of maximum likelihood estimator in pure birth process and the pure death process, Scand. J. Statist. 3, 61—72.
Бёлер (Buehler R. J.)
(1959) Some validity criteria for statistical inferences, Ann. Math. Statist., 30, 845—863.
(1971) Measuring information and uncertainty (with discussion). In Foundations of statistical inference, pp. 330—341, eds Godambe V. P. and Sprott D. A., Toronto, Holt, Rinehart and Winston.
Биллингсли (Billingsley P.)
(1961a) Statistical methods in Markov chains, Ann. Math. Statist., 32, 12—40.
(1961b) Statistical inference for Markov processes, University of Chicago Press.
(1968) Convergence of probability measures, New York, Wiley. (Русский перевод: Биллингсли П. Сходимость вероятностных мер.— М.: Наука, 1977.)
Бишоп, Файнберг и Холланд (Bishop Y.M.M., Feinberg S. E., Holland P. W.)
(1975) Discrete Multivariate Analysis. Cambridge, Mass. M.I.T. Press.
Бокс и Хилл (Box G. E. P., Hill W. J.)
(1967) Dicrimination among mechanistic models, Technometrics, 9, 57—71.
Вальд (Wald A.)
(1947) Sequential analysis, New York, Wiley. (Русский перевод: Вальд А. Последовательный анализ.— М.; Физматгиз, 1960.)

Ватсон (Watson G. S.)
(1958) On chi-squared goodness of-fit tests for continuous distributions (with discussion), J. R. Statist. Soc., B, 20, 44—72.
Гейвер и Хоуэл (Gaver D. P., Hoel D. G.)
(1970) Comparison of certain small-sample Poisson probability estimates, Technometrics, 12, 835—850.
Гиршик и Сэвидж (Girshick M. A., Savage L. J.)
(1951) Bayes and minimax estimates for quadratic loss functions, Proc. 2nd Berkeley Symp., 53—73.
Гнанадесикан (Gnanadesikan)
(1977) Methods for statistical Data Analysis of Multivariate Observations, New York, Wiley.
Годемб (Godambe V. P.)
(1960) An optimum property of regular maximum likelihood estimation, Ann. Math. Statist., 31, 1208—1211.
Годемб и Томпсон (Godambe V. P., Thompson M. E.)
(1971) The specification of prior knowledge by classes of prior distributions in survey sampling estimation (with discussion). In Foundations of statistical inference, pp. 243—258, eds. Gedambe V. P. and Sprott D. A., Toronto, Holt Rinehart and Winston.
Гранди и Хили (Grundy P. M., Healy M. J. R.)
(1950) Restricted randomization and quasi-Latin squares, J. R. Statist. Soc., B, 12, 286—291.
Гранди, Хили и Рис (Grundy P. M., Healy M. J. R., Ress D. H.)
(1956) Economic choive of the amount of experimentation (with discussion). J. R. Statist. Soc., B, 18, 32—55.
Гудман (Goodman L. A.)
(1953) Sequential sampling tagging form population size problems, Ann. Math. Statist., 24, 56—69.
(1968) The analysis of cross-classified data: independence, quasi-independence and interactions in contingency tables with or without missing entries, J. Amer. Statist. Assoc., 63, 1091—1131.
Дар (Dar S. N.)
(1962) On the comparison of the sensitivities of experiments, J. R. Statist. Soc., B, 24, 447—453.
Дики (Dickey J. M.)
(1971) The weighted likelihood ratio, linear hypotheses on normal location parameters, Ann. Math. Statist., 42, 204—223.
Дурбин (Durbin J.)
(1960) Estimation of parameters in time-series regression models, J. R. Statist. Soc., B, 22, 139—153.
(1973) Distribution theory for tests based on the sample distribution function, Philadelphia, S. I. A. M.
Дэвид (Dawid A. P.)
(1973) Posterior expectations for large observations, Biometrika, 60, 664—667.
Дэвид, Стоун и Зидек (Dawid A. P., Stone M., Zidek J.)
(1973) Marginalization paradoxes in Bayesian and structural inference (with discussion), J. R. Statist. Soc., B, 35, 189—233.
Дэрроч (Darroch J. N.)
(1971) A definition of independence for bounded-sum, non-negative, integervalued variables, Biometrika, 58, 357—368.
Дэрроч и Рэтклифф (Darroch J. N., Ratcliffe D.)
(1973) Tests of F-independence with reference to quasi-independence and Waite's fingerprint data, Biometrika, 60, 395—401.
Кемпторн (Kempthorne O.)
(1952) The design and analysis of experiments, New York, Wiley.
(1966) Some aspects of experimental inference. J. Amer. Statist. Assoc., 61, 11—34.
Кемпторн и Дерфлер (Kempthorne O., Doerfler T. E.)
(1969) The behaviour of some significance tests under experimental randomization, Biometrika, 56, 231—248.
Кендалл (Kendall M. G.)
(1962) Rank correlation methods, 3rd edition, London, Griffin.
(1973) Entropy, probability and information. Rev. Int. Inst. Statist., 41, 59—68.
Кендалл и Стьюарт (Kendall M. G., Stuart A.)
(1967 — 1969) Advanced theory of statistics, vols. 1—3 (3rd 2nd and 2nd editions), London, Griffin. (Русские переводы: Кендалл М. Дж., Стьюарт А. т. 1, Теория распределений. — М.: Наука, 1966; т. 2, Статистические выводы и связи. — М.: Наука, 1975; т. 3, Многомерный статистический анализ и временные ряды. — М.: Наука, 1976.)
Кенуй (Quenouille M. H.)
(1949) Approximate tests of correlation in time-series, J. R. Statist. Soc., B, 11, 68—81.
(1956) Notes on bias in estimation, Biometrika, 43, 353—360.
Клотц (Klotz J.)
(1973) Statistical inference in Bernoulli trials with dependence, Ann. Statist., 1, 373—379.
Кокс (Cox D. R.)
(1961) Tests of separate families of hypotheses, Proc. 4th Berkely Symp., 1, 105—123.
(1962) Further resnlts on tests of separate families of hypotheses, J. R. Statist. Soc., B, 24, 406—424.

(1964a) Some problems of statistical analysis connected with congestion (with discussion). In Congrestion Theory, pp. 289—316, eds. Smith W. L. and Wilkinson W. E. Univ. of North Carolina Press.
(1964b) Some applications of exponential ordered scores. J. R. Statist. Soc., B, 26, 103—110.
(1968) Notes on some aspects of regression analysis. J. R. Statist. Soc. A, 131, 265—279.
(1970) The analysis of binary data. London, Methuen.
(1971) The choice between alternative ancillary statistics. J. R. Statist. Soc., B, 33, 251—255.
(1975) A note on partially Bayes inference and the linear model. Biometrika, 62, 651—654.
(1978) Some remarks on the role in statistics of graphical methods. Appl. Statist.
Кокс и Льюис (Cox D. R., Lewis P. A. W.)
(1966) The Statistically analysis of series of events, London, Methuen.
Копас (Copas J. B.)
(1969) Compound decisions and empirical Bayes (with discussion). J. R. Statist. Soc., B. 31, 397—425.
(1972) The likelihood surface in the linear functional relationship problem. J. R. Statist. Soc., B. 34, 274—278.
Кормак (Cormack R. M.)
(1968) The statistics of capture-recapture methods, Oceanogr. Mar. Biol. Ann. Rev., 6, 455—506.
Коэн (Cohen L.)
(1958) On mixed single sample experiments, Ann. Math. Statist., 29, 947—971.
Крамер (Cramer H.)
(1946) Mathematical methods of statistics, Princeton University Press. (Русский перевод: Крамер Г. Математические методы статистики, 2-е изд.— М.: Мир, 1975.)
Кульбак (Kullback S.)
(1968) Information theory and statistics, New York, Dover. (Русский перевод: Кульбак С. Теория информации и статистика.— М.: Наука, 1967.)
Купмен (Koopman B. O.)
(1936) On distribution admitting a sufficient statistic, Trans. Amer. Math. Soc., 39, 399—409.
Кэлбфлейш (Kalbfleisch J. D.)
(1974) Sufficiency and conditionality, Unpublished paper.
Лауритцен (Lauritzen S.)
(1974) Sufficiency prediction and extreme models, Scandinavian J. Statist., to appear.
Линдли, Ист и Гамильтон (Lindley D. V., East D. A., Hamilton P. A.)
(1960) Tables for making inferences about the variance of a normal distribution. Biometrika, 47, 433—437.
Литтел и Фолкс (Littell R. C., Folks J. L.)
(1971) Asymptotic optimality, of Fisher's method of combining independent tests. J. Amer. Statist. Assoc., 66, 802—806.
Мандель (Mandel J.)
(1971) A new analysis of variance model for non-additive data, Technometrics, 13, 1—18.
Мардиа (Mardia K. V.)
(1971) The effect of non-normality on some multivariate tests and robustness to nonnormality in the linear model, Biometrika, 58, 105—121.
(1972) Statistics of directional data, London, Academic Press. (Русский перевод: Мардиа К. Статистический анализ угловых наблюдений.— М.: Наука, 1978.)
Мариц (Maritz J.)
(1970) Empirical Bayes' methods, London, Methuen.
Мартин-Леф (Martin-Löf P.)
(1974) The notion of redundancy and its use as a qualititave measure of the discrepancy between a statistical hypothesis and a set of observed data. Scand. J. Statist., 1, 3—18.
Муд (Mood A. M.)
(1943) On the dependence of sampling inspection plans upon population distributions, Ann. Math. Statist., 14, 415—425.
Нейман (Neyman J.)
(1962) Lectures and conferences on mathematical statistics, 2nd edition, Washington, U. S. Dept. Agric. Grad. School.
(1959) Optimal asymptotic tests of composite statistical hypotheses, In Probability and statistics, pp. 213—234, ed. Grenander U., Stockholm, Almqvist and Wiksell.
Нейман и Пирсон (Neyman J., Pearson E. S.)
(1928) On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika, A, 20, 175—240 and 263—294.
(1933) On the problem of the most efficient tests of statistical hypotheses, Phil. Trans. R. Soc., A, 231, 289—337.
(1936) Contributions to the theory of testing statistical hypotheses, I, Unbiased critical regions of type A and type A, Stat. Rev. Mem., 1, 1—37.
(1967) Joint statistical papers. Cambridge University Press.
Нейман и Скотт (Neyman J., Scott E. L.)
(1948) Consistent estimates based on partially consistent observations, Econometrica, 16, 1—32.
(1960) Correction for bias introduced by a transformation of variables. Ann. Math. Statist., 31, 643—661.
Нелдер и Веддерберн (Nelder J., Wedderburn R. W. M.)
(1972) Generalized linear models. J. R. Statist. Soc., A, 135, 370—384.
Пирс (Pierce D. A.)

(1973) On some difficulties in a frequency theory of inference. Ann. Statist., 1, 241—250.
Пирсон (Pearson K.)
(1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of varia bles is such that is can reasonably by supposed to have arisen from random sampling. Phil. Mag. Series. 5, 50, 157—175.
Пирсон и Хартли (Pearson E. S., Hartley H. O.)
(1970) Biometrika tables for statisticians. Vol. 1. 3rd edition. Cambridge University Press.
(1972) Biometrika tables for statisticians. Vol. 2. Cambridge University Press.
Питман (Pitman E. J. G.)
(1936) Sufficient statistics and intrinsic accuracy. Proc. Camb. Phil. Soc., 32, 567—579.
Райффа и Шлайфер (Raiffa H., Schlaifer R.)
(1961) Applied statistical decision theory. Boston, Harvard Business School.
Рао (Rao C. R.) (1961) Asymptotic efficiency and limiting information. Proc. 4th Berkeley Symp., 1, 531—545.
Роббинс (Robbins H.)
(1951) Asymptotically subminimax solutions of compound statistical decision problems. Proc. 2nd Berkeley Symp., 131—148.
Ройалл (Royалл H.)
(1976) Likelihood functions in finite population sampling theory. Biometrika, 63, 605—614.
Свердруп (Sverdrup P. V.)
(1975) Tests without power. Scand. J. Statist., 2, 158—160.
Себер (Seber G. A. F.)
(1973) The estimation of animal abundance and er related parameters. London, Griffin.
Секхэтм (Sukhatme P. V.)
(1936) On the analysis of K samples from exponentional populations with especial reference to the problem of random intervals. Statist. Res. Mem., 1, 94—112.
Стейн (Stein C.)
(1945) A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist., 16, 243—258.
(1964) Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. Ann. Inst. Statist. Math., 16, 155—160.
Стоун (Stone M.)
(1969) The role of significance testing: some data with a message. Biometrika, 56, 485—493.
(1974) Cross-validatory. choice and assessment of statistical predictions (with discussion). J. R. Statist. Soc., B, 36, 111—147.
(1977) Asymptotics for and against cross-validation. Biometrika, 64, 29—35.
Стоун и Дэвид (Stone M., and Dawid A. P.)
(1972) Un Bayesian implications of improper Bayes inference in routine statistical problems. Biometrika, 59, 369—375.
Стоун и Спрингер (Stone M., Springer B. G. F.)
(1965) A paradox involving quasi prior distributions. Biometrika, 52, 623—627.
Таллис (Tallis G. M.)
(1969) Note on a calibration problem. Biometrika, 56, 505—508.
Уайт (White L. V.)
(1973) An extension of the general equivalence theorem to nonlinear models. Biometrika, 60, 345—348.
Уилк, Гнанадесикан и Фринн (Wilk M. B., Gianadesikan R., Freeny A. E.)
(1963) Estimation of error variance from smallest ordered contrasts. J. Amer. Statist. Assoc., 58, 152—160.
Уильямс (Williams E. J.)
(1962) Exact fiducial limits in nonlinear estimation. J. R. Statist. Soc., B, 24, 125—139.
Утхофф (Uthoff V. A.)
(1970) An optimum test property of two well-known statistics. J. Amer. Statist. Assoc., 65, 1597—1600.
Уэзерилл (Wetherill G. B.)
(1975) Sequntial method in statistics. 2nd edition. London, Chapman and Hall.
Феллер (Feller W.)
(1971) An introduction to probability theory. Vol. 1, 3nd edition. New York, Wiley. (Русский перевод с 2-го изд.: Феллер В. Введение в теорию вероятностей и ее применение, т. 1. — М.: Мир, 1967.)
Фергюсон (Ferguson T. S.)
(1961) On the rejection of outliers. Proc. 4th Berkeley Symp., 1, 253—287.
де Финетти (de Finetti B.)
(1972) Probability, induction and statistics. London, Wiley.
Фишер (Fisher R. A.)
(1925) Theory of statistical estimation. Proc. Camb. Phil. Soc., 22, 700—725.
(1934) Two new properties of mathematical likelihood. Proc. R. Soc., A, 144, 285—307.
Фрэзер (Fraser D. A. S.)
(1961) The fiducial method and invariance. Biometrika, 48, 261—280
(1968) The structure of inference. New York, Wiley.
Хёфдинг (Hoeffding W.)
(1948) A class of statistics with asymptotically normal distribution. Ann. Math. Statist., 19, 293—325.
(1965) Asymptotically optimal tests for multinomial distributions (with discussion). Ann. Math. Statist., 36, 369—408.

Хили (Healy M. J. R.)
(1968) Multivariate normal plotting. Appl. Statistics, 17, 157—161.
Хинкли (Hinkley D. V.)
(1973) Two-sample tests with unordered pairs. J. R. Statist. Soc., B, 35, 337—346.
(1977) Conditional inference about normal mean with known coefficient of variation. Biometrika, 64, 105—108.
Хогг (Hogg R. V.)
(1956) On the distribution of the likelihood ratio. Ann. Math. Statist., 27, 529—532.
Чернов и Леман (Chernoff H., Lehmann E. L.)
(1954) The use of maximum likelihood estimates in χ^2 tests for goodness of fit. Ann. Math. Statist., 25, 579—586.
Шапиро и Уилк (Shapiro S. S., Wilk M. B.)
(1965) An analysis of variance test for normality (complete samples). Biometrika, 52, 591—611.
Шапиро, Уилк и Чен (Shapiro S. S., Wilk M. B., Chen H. J.)
(1968) A comparative study of various tests for normality. J. Amer. Statist. Assoc., 63, 1343—1372.
Шеффе (Scheffe H.)
(1959) The analysis of variance. (Русский перевод: Шеффе Г. Дисперсионный анализ.— М.: ГИФМЛ, 1963.)
Эйтчисон и Дансмор (Aitchison J., Dunsmore I. R.)
(1968) Linear-loss interval estimation of location and scale parameters. Biometrika, 55, 141—148.
Эйтчисон и Силвей (Aitchison J., Silvey S. D.)
(1958) Maximum-likelihood estimation of parameters subject to restrainte. Ann. Math. Statist., 29, 813—828.
Эндрюс, Гнанадесикан и Уорнер (Andrews D. F., Gnanadesikan R., Warner J. L.)
(1971) Transformations of multivariate data. Biometrika, 27, 825—840.
(1973) Methods for assesing multivariate normallity. In Multivariate Analysis, III, pp. 95—116, ed. Krishnaiah P. R. New. York, Academic Press.
Эфрон (Efron B.)
(1971) Does an observed sequence of numbers follow a simple rule? (with discussion). J. Amer. Statist. Assoc., 66, 552—568.
(1975) Defining the curvature of a statistical problem (with discussion). Ann. Statist., 3, 1189—1242.
Эфрон и Моррис (Efron B., Morris C.)
(1972) Limiting the risk of Bayes and empirical Bayes estimators Part II: the empirical Bayes case. J. Amer. Statist. Assoc., 67, 130—139.
Эфрон и Хинкли (Efron B., Hinkley D. V.)
(1978) Assessing the accurancy of the maximum likelihood estimator: observed versus expected Fisher information (to be published).

Оглавление

От редактора перевода 5
Предисловие 6

Глава 1. Введение 7
Необходимые сведения 7

Глава 2. О некоторых общих понятиях 9
Необходимые сведения 9
Задачи 11

Глава 3. Слабые критерии значимости 30
Необходимые сведения 30
Задачи 31

Глава 4. Критерии значимости: простые нулевые гипотезы 51
Необходимые сведения 51
Задачи 53

Глава 5. Критерии значимости: сложные нулевые гипотезы 72
Необходимые сведения 72
Задачи 74

Глава 6. Критерии, свободные от распределения, и критерии рандомизации 89
Необходимые сведения 89
Задачи 90

Глава 7. Интервальное оценивание 107
Необходимые сведения 107
Задачи 109

Глава 8. Точечное оценивание 123
Необходимые сведения 123
Задачи 130

Глава 9. Асимптотическая теория 149
Необходимые сведения 149
Задачи 152

Глава 10. Байесовские решения 187
Необходимые сведения 187
Задачи 18

Глава 11. Теория решений 20
Необходимые сведения 20
Задачи 20

Список литературы 22