Д. КОКС
Д. ХИНКЛИ

Задачи по теоретической статистике с решениями

ИЗДАТЕЛЬСТВО «МИР»

Издательство
·МИР·

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. Бряндинская. 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 $60\times90^1/_{16}$. 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 $\Phi(x)=\frac{1}{2\pi}\int\limits_{-\infty}^{x} e^{-u^2/2}\,du$.—*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 $\mathrm{lik}(\theta;\ y)$ полагается равным $f_Y(y;\ \theta)$ и рассматривается как функция параметра θ. Если Y — вектор с независимыми компонентами, логарифм правдоподобия $l(\theta;\ y) = \log \mathrm{lik}(\theta;\ y)$ 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

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

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 $\operatorname{lik}(\theta;\ y)$;

(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:

$$-m_1m_2\log\sigma-\Sigma\Sigma(y_{jk}-\mu-\alpha_j-\beta_k)^2/(2\sigma^2)=$$
$$=-m_1m_2\log\sigma-\Sigma\Sigma\{y_{jk}-\bar{y}_{j.}-\bar{y}_{.k}+\bar{y}_{..})+$$
$$+(\bar{y}_{j.}-\bar{y}_{..}-\alpha_j)+(\bar{y}_{.k}-\bar{y}_{.}-\beta_k)+(\bar{y}_{..}-\mu)\}^2/(2\sigma^2).$$

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 $\Sigma\Sigma(\bar{y}_{jk}-\bar{y}_{j.}-\bar{y}_{.k}+y_{..})^2$, 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 временно́го ряда и случайного процесса обычно проще всего получить, вычисляя последовательно условные вероятности, т. е. представляя плотность вероятности в следующем виде:

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

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_0. 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

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

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

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

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; Билинсгли, 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 $Y_j=(Y_{j1}, \ldots, Y_{jr})^{\mathrm{T}}$ независимы и имеют многомерное нормальное распределение со средним $(\mu, \ldots, \mu)^{\mathrm{T}}=\mu 1$ и ковариационной матрицей

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

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

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

Логарифм правдоподобия принимает вид

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

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

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

где $\lambda=\sigma_b^2/(\sigma_W^2+r\sigma_b^2)$. Поскольку $\mathbf{J}=11^{\mathrm{T}}$, 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

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

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