

Глава 6

Достаточность и полнота

6.1 Базовая часть

6.1.1 Достаточность. Определение

Достаточность в схеме Бернулли

Давайте начнем с одного примера

Пример 1. Пусть $T(X_1, \dots, X_n) = X_1 + \dots + X_n$, X_i — н.о.р. величины в схеме Бернулли. Тогда можно представить себе эксперимент по розыгрышу X_1, \dots, X_n в два этапа:

а) Розыгрыш $T(X_1, \dots, X_n)$ по закону $\mathbf{P}_\theta(T(X_1, \dots, X_n) = k) = C_n^k \theta^k (1 - \theta)^{n-k}$

б) При известном $T(X_1, \dots, X_n)$ розыгрыш X_1, \dots, X_n в соответствии с законом $\mathbf{P}_\theta(X_1 = x_1, \dots, X_n = x_n | T(X_1, \dots, X_n) = k) = I_{x_1 + \dots + x_n = k} / C_n^k$.

Таким образом, мы будем получать $\mathbf{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = \theta^k (1 - \theta)^{n-k}$, что нам и требуется. Значит, прямой розыгрыш X_1, \dots, X_n эквивалентен такому двухэтапному розыгрышу. При этом второй этап совершенно не помогает нам в оценке θ , ведь при всех θ распределение X_1, \dots, X_n при условии $T(X_1, \dots, X_n) = k$ одно и то же.

Значит для оценки θ нам важно знать лишь $X_1 + \dots + X_n$, после этого знание самих X_1, \dots, X_n никак не помогает в определении θ . Такие статистики называются достаточными, поскольку их знания достаточно для оценки θ .

Общее определение

Более общее определение звучит так. Рассмотрим статистику $T(X_1, \dots, X_n)$.

Определение 1. Если для каждого $A \in \mathcal{B}(R^n)$ и θ

$$\mathbf{P}_\theta((X_1, \dots, X_n) \in A | T(X_1, \dots, X_n) = t) = \mathbf{E}_\theta(I_{(X_1, \dots, X_n) \in A} | T(X_1, \dots, X_n) = t)$$

не зависит от θ , то статистика T называется достаточной. Здесь под $\mathbf{E}(Y | Z = z)$ подразумевается $g(z)$, где g — такая функция, что $\mathbf{E}(Y | Z) = g(Z)$.

Строго говоря, функция g определена неоднозначно (а с точностью до множества меры ноль), поэтому более правильно сказать "существует версия функции g , не зависящая от θ ".

В частности, для дискретного случая нам необходимо проверить то, что условное распределение

$$\mathbf{P}_\theta(X_1 = x_1, \dots, X_n = x_n | T(X_1, \dots, X_n) = k)$$

не зависит от θ , для абсолютно-непрерывного — что условная плотность

$$f_{X_1, \dots, X_n | T(X_1, \dots, X_n)}(x_1, \dots, x_n | t) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{T(X_1, \dots, X_n)}(t)},$$

где $t = T(x_1, \dots, x_n)$, не зависит от θ .

Пример

Пример 2. Рассмотрим нормальное распределение со средним θ и дисперсией 1. Тогда плотность X_1, \dots, X_n будет равна

$$(2\pi)^{-n/2} \exp \left(- \sum_{i=1}^n (x_i - \theta)^2 / 2 \right),$$

величина $X_1 + \dots + X_n$ распределена $\mathcal{N}(n\theta, n)$, откуда

$$f_{X_1 + \dots + X_n}(t) = (2\pi n)^{-1/2} \exp(-(t - n\theta)^2 / (2n)),$$

а, следовательно,

$$\begin{aligned} f_{X_1, \dots, X_n | T(X_1, \dots, X_n)}(x_1, \dots, x_n | t) &= \frac{(2\pi)^{-n/2} \exp \left(- \sum_{i=1}^n \frac{(x_i - \theta)^2}{2} \right)}{(2\pi n)^{-1/2} \exp \left(- \frac{(x_1 + \dots + x_n - n\theta)^2}{2n} \right)} = \\ &= (2\pi)^{-(n-1)/2} n^{1/2} \exp \left(- \sum_{i=1}^n x_i^2 / 2 + (x_1 + \dots + x_n)^2 / (2n) \right), \end{aligned}$$

т.е. не зависит от θ . Значит $X_1 + \dots + X_n$ — достаточна.

Замечание 1. На самом деле, это не совсем плотность, ведь величины X_1, \dots, X_n сосредоточены на поверхности $T(X_1, \dots, X_n) = t$, имеющей меньшую чем n размерность. Для проверки достаточности хватает и такой функции, а, скажем, при подсчете условных математических ожиданий, это не слишком удобно. Поэтому вместо плотности $f_{X_1, \dots, X_n | T}(x_1, \dots, x_n | t)$ надо будет найти плотность $f_{Y_1, \dots, Y_{n-1}, T}(y_1, \dots, y_{n-1}, t)$, где Y_1, \dots, Y_{n-1} какие-нибудь величины, т.ч. Y_1, \dots, Y_{n-1}, T — однозначно определяют X_1, \dots, X_n , а затем рассматривать новую плотность как функцию $n - 1$ переменных. Скажем, для подсчета математического ожидания $g(X_1, \dots, X_n)$ при условии $X_1 + \dots + X_n = t$, мы заменим плотность $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ на $f_{X_1, \dots, X_{n-1}, T}(x_1, \dots, x_{n-1}, t)$, равную $f_{X_1, \dots, X_n}(x_1, \dots, x_{n-1}, t - x_1 - \dots - x_{n-1})$ и будем считать интеграл

$$\int_{\mathbb{R}^{n-1}} g(x_1, \dots, x_{n-1}, t - x_1 - \dots - x_{n-1}) f_{X_1, \dots, X_n | T}(x_1, \dots, x_{n-1}, t - x_1 - \dots - x_{n-1} | t) dx_1 \dots dx_{n-1}.$$

Некоторые замечания

Сделаем несколько замечаний.

- Статистика T может быть векторной, например $T(X_1, \dots, X_n) = (X_1, \dots, X_n)$ — всегда достаточная статистика, поскольку условная вероятность $\mathbf{P}_\theta((X_1, \dots, X_n) \in A | (X_1, \dots, X_n) = (x_1, \dots, x_n))$ есть $I_{x_1, \dots, x_n \in A}$, т.е. не зависит от θ . В рамках нашей модели — н.о.р. X_i , такой статистикой также является $X_{(1)}, \dots, X_{(n)}$.
- Заметим, что достаточность это свойство статистики, а не *оценки*. Если статистика T достаточна, то любая взаимно однозначная функция от нее также будет достаточной, поскольку сигма-алгебры, порожденные этими величинами одинаковы.

6.1.2 Критерий факторизации

6.1.3 Критерий и его обоснование

Определение достаточной статистики не очень удобно для проверки. Поэтому зачастую используют следующий критерий, называемый критерием факторизации:

Теорема 1. Пусть T — достаточная статистика, величины X_i дискретны или абсолютно-непрерывны. Тогда статистика T достаточна тогда и только тогда, когда правдоподобие L представимо в виде

$$L(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n), \theta)h(x_1, \dots, x_n).$$

g, h — некоторые борелевские функции.

Иначе говоря, правдоподобие должно быть представимо в виде произведения двух функций — одна не зависит от θ , а другая зависит от x только через T .

Поймем почему это так для дискретного случая. Достаточность статистики означает, что при $t = T(x_1, \dots, x_n)$

$$\mathbf{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = \mathbf{P}_\theta(T(X_1, \dots, X_n) = t)h(x_1, \dots, x_n),$$

где $h(x_1, \dots, x_n)$ — условная вероятность $\mathbf{P}_\theta(X_1 = x_1, \dots, X_n = x_n | T(X_1, \dots, X_n) = t)$. Аналогично доказывается обратное утверждение.

Примеры

Пример 3. Для распределения Бернулли

$$L(x_1, \dots, x_n; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} = f\left(\sum x_i, \theta\right),$$

т.е. $X_1 + \dots + X_n$ — достаточная статистика. Аналогично для нормального распределения со средним θ и дисперсией 1

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= (2\pi)^{-n/2} \exp\left(-\sum (x_i - \theta)^2/2\right) = \\ &= (2\pi)^{-n/2} \exp\left(-\sum x_i^2/2 + \theta \sum x_i - n\theta^2/2\right) = f\left(\sum x_i, \theta\right) h(x_1, \dots, x_n), \end{aligned}$$

т.е. $X_1 + \dots + X_n$ — достаточная статистика.

Для равномерного распределения на отрезке $[0, \theta]$

$$L(x_1, \dots, x_n; \theta) = \theta^{-n} I_{x_1, \dots, x_n \in (0, \theta)} = \theta^{-n} I_{\max(x_i) < \theta} I_{\min(x_i) > 0} = f(\max(x_i), \theta) h(x_1, \dots, x_n),$$

т.е. $\max(X_1, \dots, X_n)$ — достаточная статистика.

6.1.4 Рао-Блэкуэллизация

Дисперсия УМО

Перейдем к еще одному важному свойству достаточных статистик. Для условного математического ожидания справедливо следующее соотношение:

$$\mathbf{D}\mathbf{E}(X|Y) = \mathbf{E}(\mathbf{E}^2(X|Y)) - (\mathbf{E}(\mathbf{E}(X|Y)))^2 \leq \mathbf{E}(\mathbf{E}(X^2|Y)) - (\mathbf{E}(\mathbf{E}(X|Y)))^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 = DX,$$

где неравенство вытекает из того, что $\mathbf{E}(X^2|Y) \geq (\mathbf{E}(X|Y))^2$ (неравенство Иенсена), а следующее за ним равенство — из свойства $\mathbf{E}(\mathbf{E}(X|Y)) = \mathbf{E}X$. При этом $\mathbf{E}(\mathbf{E}(X|Y)) = \mathbf{E}(X)$, поэтому заменяя оценку $\hat{\theta}(X_1, \dots, X_n)$ на ее условное матожидание при условии какой-либо статистики T , я получу случайную

величину с тем средним и с той же или меньшей дисперсией. Казалось бы, отличный способ улучшать оценки, но, к сожалению, новая случайная величина может и не быть оценкой, т.к. будет зависеть от θ :

Пример 4. Рассмотрим схему Бернулли с параметром θ . Возьмем статистику $T(X_1, \dots, X_n) = X_1$ и несмещенную оценку $\hat{\theta}(X_1, \dots, X_n) = X_2$. Тогда

$$\mathbf{E}_{\theta} \left(\hat{\theta}(X_1, \dots, X_n) | T(X_1, \dots, X_n) \right) = \mathbf{E}_{\theta}(X_1 | X_2) = \mathbf{E}_{\theta} X_1 = \theta.$$

Полученная случайная величина не является оценкой, так как зависит от параметра.

Теорема Колмогорова-Блэкуэлла-Рао

Иначе дело обстоит, если статистика T достаточна. Тогда для любой оценки $\hat{\theta}$ с конечным математическим ожиданием ее условное распределение $\mathbf{P}_{\theta}(\hat{\theta} \in A | T(X_1, \dots, X_n) = t)$ не будет зависеть от θ , а, следовательно, $\mathbf{E}_{\theta}(\hat{\theta} | T(X_1, \dots, X_n))$ будет статистикой, т.е. измеримой функцией от T , не зависящей от θ . При этом у нее будет то же математическое ожидание (в силу свойства $\mathbf{E}(\mathbf{E}(X|Y)) = \mathbf{E}X$) и, как мы доказали, не большая дисперсия. Таким образом, мы доказали следующую теорему.

Теорема 2 ((Колмогорова-Блэкуэлла-Рао)). *Если X — несмещенная оценка $f(\theta)$, T — достаточная статистика, то $\mathbf{E}_{\theta}(X|T)$ также будет несмещенной оценкой T , причем будет иметь меньшую или равную дисперсию.*

Блэкуэллизация

Получаем процедуру улучшения оценок, которую иногда называют Рао-блэкуэллизацией (Rao-Blackwellization):

- Строим достаточную статистику.
- Строим несмещенную оценку параметра.
- Улучшаем несмещенную оценку, взяв математическое ожидание относительно достаточной статистики.

Пример

Пример 5. В примере 4 возьмем в качестве T достаточную статистику $X_1 + \dots + X_n$. Тогда, как мы подсчитали на прошлом семинаре,

$$\mathbf{E}_{\theta}(X_1 | X_1 + \dots + X_n) = \bar{X}.$$

Полученная случайная величина будет являться несмещенной оценкой с дисперсией $\theta(1 - \theta)/n$, значительно лучшей, чем $D_{\theta}X_1 = \theta(1 - \theta)$.

6.1.5 Полные достаточные статистики

Определение и примеры

Еще большего эффекта можно достигать при помощи так называемых полных достаточных статистик.

Определение 2. Статистика T называется полной, если для любой борелевской функции f , такой, что $\mathbf{E}_{\theta}f(T) = 0$ при всех θ , справедливо соотношение $f(T) = 0$ п.н.

Пример 6. В схеме Бернулли с $\theta \in (0, 1)$ статистика (X_1, \dots, X_n) , $n > 1$ не является полной статистикой, поскольку

$$\mathbf{E}_\theta(X_1 - X_2) = 0,$$

хотя $X_1 = X_2$ выполнено только с вероятностью $1 - 2\theta(1 - \theta) < 1$. При этом статистика X_1 полна. Действительно,

$$\mathbf{E}_\theta f(X_1) = \theta f(1) + (1 - \theta)f(0) = f(0) + \theta(f(1) - f(0)),$$

т.е. равенство 0 при всех θ возможно лишь при $f(1) = f(0) = 0$, что равносильно тому что $f(T) = 0$ п.н. Статистика $X_1 + \dots + X_n$ также полна. Действительно, соотношение

$$\mathbf{E}_\theta f(X_1 + \dots + X_n) = \sum_{k=0}^n f(k) C_n^k \theta^k (1 - \theta)^{n-k} = 0$$

при всех $\theta \in (0, 1)$ влечет за собой равенство нулю многочлена на отрезке, откуда все его коэффициенты равны 0. Свободный член этого многочлена равен $C_n^0 f(0)$, коэффициент при θ — $C_n^1 f(1)$ и т.д., откуда $f(i) = 0$ для всех неотрицательных целых i . Значит статистика $X_1 + \dots + X_n$ полна. При этом статистика X_1 тоже полна по тем же причинам.

Пример 7. Для равномерного распределения X_1, \dots, X_n на $[0, \theta]$ из

$$\mathbf{E}_\theta f(\max X_i) = \theta^{-n} \int_0^\theta f(t) t^{n-1} dt = 0$$

следует, что $\int_a^b f(t) t^{n-1} dt = 0$ для любых $a, b > 0$, следовательно, $\int_A f(t) t^{n-1} dt = 0$ для любого борелевского A , откуда $f(t) t^{n-1} = 0$ при п.в. $t > 0$. Значит, статистика $\max X_i$ полна.

6.1.6 Рао-Блэкуэллизация относительно полной статистики

Если статистика T полна, то для каждой измеримой $f(\theta)$ найдется не более одной измеримой функции от T , являющейся несмещенной оценкой $f(\theta)$. Действительно, если их нашлось бы две функции $f_1(T)$ и $f_2(T)$, то величина $f_1(T) - f_2(T)$ имела бы мат. ожидание 0 при каждом θ , а значит $f_1(T) = f_2(T)$ п.н.

Для полных достаточных статистик справедлива следующая теорема:

Теорема 3 (О полной достаточной статистике). Пусть g, f — борелевские функции, T — полная достаточная статистика, $\mathbf{E}_\theta g(T) = f(\theta)$. Тогда статистика $g(T)$ имеет наименьшую возможную дисперсию среди несмещенных оценок $f(\theta)$.

Доказательство. Пусть некоторая оценка $\hat{\theta}$ — несмещенная оценка $f(\theta)$ с меньшей дисперсией. Тогда $\mathbf{E}(\hat{\theta}|T)$ будет несмещенной оценкой с дисперсией меньше чем $g(T)$. Но $\mathbf{E}(\hat{\theta}|T)$ — функция от T , имеющая то же математическое ожидание, что и T при всех θ . Т.к. T — полная, то эта функция равна $g(T)$, что противоречит тому, что у нее дисперсия меньше. \square

Таким образом, имея любую несмещенную оценку, мы можем взять от нее условное математическое ожидание при условии полной достаточной статистики (если такая найдется) и это будет несмещенной оценкой с наименьшей дисперсией. В частности, в примерах 4 и 5 мы убедились, что \bar{X} — оптимальная (т.е. несмещенная с наименьшей дисперсией) оценка θ в схеме Бернулли, а $(n+1)\max X_i/n$ — для равномерного распределения на $[0, \theta]$.

Пример

Пример 8. Пусть X_1, \dots, X_n — н.о.р. нормальные величины со средним θ и дисперсией 1. Можно доказать, что $X_1 + \dots + X_n$ — полная достаточная статистика (см. факультатив). Построим несмещенную

оценку с минимальной дисперсией для θ^2 . Мы знаем, что функция $X_1^2 - 1$ является несмещенной оценкой θ^2 ,

$$\mathbf{E}_\theta X_1^2 = D_\theta X_1 + (\mathbf{E}_\theta X_1)^2 = 1 + \theta^2.$$

т.к. математическое ожидание X_1^2 есть сумма дисперсии X_1 и квадрата математического ожидания X_1 . Следовательно, несмещенной оценкой с минимальной дисперсией будет $\mathbf{E}_\theta(X_1^2 - 1 | X_1 + \dots + X_n)$. Плотность X_1 при условии $X_1 + \dots + X_n$ описывается соотношением

$$f_{X_1, X_1 + \dots + X_n}(x, y) = f_{X_1, X_2 + \dots + X_n}(x, y - x) = (2\pi)^{-1} (n-1)^{-1/2} e^{-x^2/2} e^{-(y-x)^2/(2(n-1))},$$

откуда условная плотность —

$$\begin{aligned} f_{X_1 | X_1 + \dots + X_n}(x | y) &= f_{X_1, X_2 + \dots + X_n}(x, y - x) = \frac{\sqrt{n} \exp\left(-\frac{x^2}{2} - \frac{(y-x)^2}{2(n-1)}\right)}{\sqrt{2\pi(n-1)} \exp\left(-\frac{y^2}{2n}\right)} = \\ &= \frac{1}{\sqrt{2\pi(n-1)/n}} \exp\left(-\frac{n}{n-1} \left(x - \frac{y}{n}\right)^2 / 2\right), \end{aligned}$$

т.е. нормальная плотность со средним y/n и дисперсией $(n-1)/n$. Отсюда

$$\mathbf{E}(X_1^2 | X_1 + \dots + X_n = y) = (y/n)^2 + (n-1)/n,$$

откуда $\hat{\theta} = (\bar{X})^2 - 1/n$ будет оптимальной оценкой θ^2 .

6.2 Факультатив

6.2.1 Минимальные достаточные статистики

Можно указать универсальную достаточную статистику (X_1, \dots, X_n) , правда, от ее использования мало выгоды. Чуть более удачной является статистика $X_{(1)}, \dots, X_{(n)}$, являющаяся достаточной для выборки из н.о.р. случайных величин.

Среди достаточных статистик нам наиболее интересные как можно более "маленькие" в смысле порожденных ими сигма-алгебр. Попросту говоря, мы бы хотели бы выбрать из достаточных статистик такую $T(x_1, \dots, x_n)$, что любая другая достаточная статистика $S(x_1, \dots, x_n)$ такова, что $T(x_1, \dots, x_n) = g(S(x_1, \dots, x_n))$ при некоторой измеримой функции g . Такая статистика называется минимальной достаточной.

Как же построить такую статистику? Ответ прост: мы хотим склеивать статистикой $T(x_1, \dots, x_n)$ как можно больше точек x_1, \dots, x_n , не нарушая условия достаточности.

Теорема 4. *Для минимальности и достаточности статистики T необходимо и достаточно (извините за каламбур), чтобы $T(\vec{x}) = T(\vec{y})$ тогда и только тогда, когда $L(\vec{x}; \theta)/L(\vec{y}; \theta)$ не зависит от θ .*

Доказательство. Доказательство несложно. Если T — достаточна, то $L(\vec{x}; \theta) = g(T(\vec{x}); \theta)h(\vec{x})$. Следовательно, при $T(\vec{x}) = T(\vec{y})$ мы имеем $L(\vec{x}; \theta)/L(\vec{y}; \theta) = h(\vec{x})/h(\vec{y})$, т.е. не зависит от θ . С другой стороны, если T будет постоянной на каждом классе $D_{\vec{x}}$, где $\vec{y} \in D_{\vec{x}}$ тогда и только тогда, когда $L(\vec{x}; \theta)/L(\vec{y}; \theta)$ не зависит от θ и разных классах будет принимать различные значения, то она будет достаточной по критерию факторизации. Действительно, сопоставим каждому значению t статистики один $\vec{x} = u(t)$, таких что $T(\vec{x}) = t$. Тогда

$$L(\vec{x}; \theta) = L(\vec{x}; \theta)/L(u(T(\vec{x})); \theta) L(u(T(\vec{x})); \theta).$$

Первый множитель не зависит от θ , а второй зависит только от $T(\vec{x})$, откуда T достаточна. Минимальность T вытекает из первой части. \square

Пример 9. Покажем, что для распределения Коши с плотностью $f_\theta(x) = 1/(\pi * (x^2 + \theta^2))$ минимальная достаточная статистика — упорядоченная по возрастанию последовательность модулей $|x_1|, \dots, |x_n|$. Действительно, рассмотрим

$$\frac{L(\vec{x}; \theta)}{L(\vec{y}; \theta)} = \frac{(\theta^2 + y_1^2) \cdots (\theta^2 + y_n^2)}{(\theta^2 + x_1^2) \cdots (\theta^2 + x_n^2)} = u(x_1, \dots, x_n, y_1, \dots, y_n).$$

Но тогда многочлен

$$(\theta^2 + y_1^2) \cdots (\theta^2 + y_n^2) - (\theta^2 + y_1^2) \cdots (\theta^2 + y_n^2) u(x_1, \dots, x_n, y_1, \dots, y_n) = 0$$

при всех θ , откуда все его коэффициенты равны 0, значит $u = 1$, а набор y_1^2, \dots, y_n^2 совпадает с x_1^2, \dots, x_n^2 . Таким образом, T должна принимать различные значения на любых двух наборах, у которых отличаются $|x_1|, \dots, |x_n|$ и $|y_1|, \dots, |y_n|$ без учета порядка, что и требовалось.

Отметим еще одну любопытную теорему.

Теорема 5. Если функция правдоподобия непрерывна слева (справа) как функция θ при каждом \vec{x} , ОМП $\hat{\theta}$ единственна и является достаточной статистикой, то $\hat{\theta}$ — минимальная достаточная статистика.

Напоследок сделаем естественное замечание

Теорема 6. Полная достаточная статистика является минимальной достаточной.

6.2.2 Проверка полноты в более сложных семействах

Проверка полноты для абсолютно-непрерывных распределений с носителями, лежащими в множестве A приводит нас к вопросу: когда набор значений

$$I(\theta) = \int_A f_\theta(x) g(x) dx, \quad \theta \in \Theta, \quad (6.1)$$

однозначно определяет функцию $g(x)$ с точностью до п.н. Для этого достаточно, чтобы $I_{[a,b]}$ для любых $a < b$ лежал в замыкании множества линейных комбинаций $\sum_{i=1}^N f_{\theta_i}(x)$ в пространстве $L^1(A)$. В этом случае для любых $a < b$

$$\int_a^b g(x) dx = 0,$$

откуда функция $g(x)$ будет нулевой п.н. Среди таких классов, например, классы x^n на отрезках A , классы $\{e^{-tx}, t > 0\}$ или $x^n e^{-x}$ на $A = \mathbb{R}^+$, $\{\sin(tx), \cos(tx), t > 0\}$ и $x^n e^{-x^2/2}$ на \mathbb{R} и многие другие.

Для большого числа преобразований (??) известно, что они обладают теоремой единственности — однозначно задают $g(x)$, если все такие преобразования существуют: это преобразования Лапласа (все те же $\exp(-tx)$ при $x > 0$), двусторонние преобразования Лапласа ($\exp(-tx)$ на всей прямой), преобразования Фурье ($\exp(itx)$, $x \in \mathbb{R}$). Большой список таких преобразований указан здесь.

Пример 10. Для $X_i \sim \mathcal{N}(\theta, 1)$ из

$$\mathbf{E}_\theta f(X_1) = \int_{\mathbb{R}} f(x) e^{-\frac{(x-\theta)^2}{2}} dx = e^{-\frac{\theta^2}{2}} \int_{\mathbb{R}} e^{-x^2/2} f(x) e^{x\theta} dx = 0.$$

Двустороннее преобразование Лапласа единственно, откуда

$$\int_{\mathbb{R}} g_1(x) e^{\lambda x} dx = \int_{\mathbb{R}} g_2(x) e^{\lambda x} dx$$

при всех $\lambda \in \mathbb{R}$, то есть $g_1(x) = g_2(x)$ п.в. Отсюда, $f(x)e^{-x^2/2} = 0$ п.н., а значит $f(x) = 0$ п.н. Поэтому X_1 — полная статистика.

6.2.3 Теорема Базу

Теорема 7 (Теорема Базу). *Если T — полная достаточная, а распределение T_1 не зависит от θ , то T и T_1 независимы.*

Доказательство. Рассмотрим УМО $g(x, t) = \mathbf{E}_\theta(I_{T_1 \leq x} | T = t)$ при любом x . В силу достаточности $g(x, t)$ не зависит от θ . При этом

$$\mathbf{E}_\theta g(x, T) = \mathbf{P}(T_1 \leq x)$$

не зависит от θ . Следовательно,

$$\mathbf{E}_\theta(g(x, T) - \mathbf{P}(T_1 \leq x)) = 0.$$

Но при любом x выше мы имеем $\mathbf{E}_\theta h(T) = 0$, где $h(t) = g(x, t) - \mathbf{P}(T_1 \leq x)$. Значит, $h(T) = 0$ п.н., откуда

$$\mathbf{E}_\theta(I_{T_1 \leq x} | T = t) = \mathbf{P}(T_1 \leq x).$$

Следовательно,

$$\mathbf{P}_\theta(T \leq x, T_1 \leq y) = \mathbf{E}_\theta(I_{T \leq x} \mathbf{E}_\theta(I_{T_1 \leq y} | T)) = \mathbf{P}_\theta(T \leq x) \mathbf{P}(T_1 \leq y),$$

что и требовалось доказать. □

С помощью этой теоремы можно доказывать независимость некоторых пар случайных величин или пар случайных векторов.

Пример 11. Покажем, что $X - Y$ и $X + Y$, где X, Y — независимые $\mathcal{N}(a, 1)$ величины. Воспользуемся тем, что для выборки X, Y статистика $X + Y$ есть полная достаточная. При этом $X - Y \sim \mathcal{N}(0, 2)$ имеет распределение, не зависящее от a . Значит, $X - Y$ и $X + Y$ независимы.