§2. Выборочные характеристики. Достаточные статистики

Лекция 3

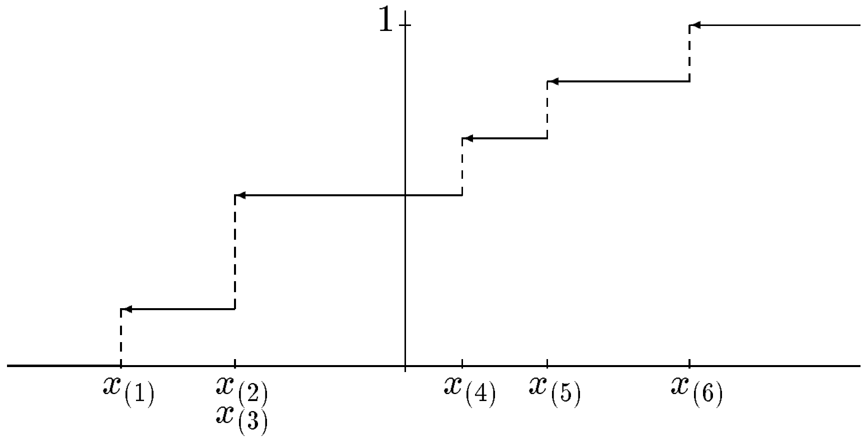
Построение вероятностных моделей в курсе теории вероятностей осуществлялось посредством спецификации функции распределения или функции плотности наблюдаемой случайной величины X. Любая из этих функций однозначно определяет распределение X на сигма-алгебре $\mathcal{A}$ борелевских множеств, порожденной интервалами в пространстве $\mathcal{X} = \mathrm{R}$ возможных значений X, и с их помощью вычислялись такие характеристики распределения, как среднее, дисперсия, коэффициенты асимметрии и эксцесса, квантили, мода и пр. В прикладной статистике существует традиция, или, можно сказать, обязательное правило, представлять полученные экспериментальные данные с помощью статистик – выборочных аналогов этих функций и характеристик распределения X. Выборочные характеристики являются оценками истинных значений своих прообразов и позволяют судить в общих чертах о характере распределения наблюдаемой случайной величины.

Такая "описательная" статистика обычно начинается с построения *вариационного ряда*: выборочные данные $x_1, \ldots, x_n$ упорядочиваются по возрастанию их значений $x_{(1)} \leq \ldots \leq x_{(n)}$, и полученный таким образом вектор с неубывающими компонентами служит реализацией случайного вектора $X_{(1)}, \ldots, X_{(n)}$, который, собственно, и следует называть вариационным рядом. Компоненты вариационного ряда называются *порядковыми статистиками*, а $X_{(1)}$ и $X_{(n)}$ – *крайними членами* вариационного ряда. Мы уже сталкивались с порядковыми статистиками, когда изучали структуру пуассоновского процесса и строили вероятностную модель "слабого звена" (распределение Вейбулла).

Упорядоченные данные наносятся на ось абсцисс, и строится ступенчатая функция, возрастающая скачками величины $1/n$ в каждой точке $x_{(1)}, \ldots, x_{(n)}$. Построенная таким образом дискретная функция распределения является реализацией случайной функции

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}(X_k < x)$$

($\mathrm{I}(A)$, как обычно, индикатор события A) и называется *эмпирической функцией распределения.*



Таким образом, дискретное эмпирическое распределение приписывает равные вероятности $1/n$ каждой из n компонент выборочного вектора, и при каждом фиксированном $x \in \mathrm{R}$ случайная величина $nF_n(x)$ подчиняется биномиальному распределению $\mathrm{B}(n,\, F(x))$:

$$P(F_n(x) = k/n) = \mathrm{C}_n^k\, F^k(x)(1 - F(x))^{n-k},\ k = 0, 1, \ldots, n.$$

В силу закона больших чисел Бернулли $F_n \underset{P}{\to} F(x)$ при любом $x \in \mathrm{R}$. Более того, теорема Гливенко–Кантелли, утверждение которой

$$D_n = \sup_{x \in \mathrm{R}} \mid F_n(x) - F(x) \mid \underset{P}{\to} 0$$

мы приводим без доказательства, указывает на равномерность этой сходимости на всей числовой оси R.

Мы закончим обсуждение свойств эмпирической функции распределения формулировкой широко известного результата А.Н. Колмогорова:

$$\lim_{n \to \infty} P(\sqrt{n} D_n < x) = \sum_{k=-\infty}^{+\infty} (-1)^k \mathrm{e}^{-k^2 x^2}.$$

Полученная им формула для асимптотического ($n \to \infty$) распределения статистики $\sqrt{n} D_n$, характеризующей величину расхождения между теоретическим F и эмпирическим F_n распределениями, используется для построения *критерия согласия* выборочных данных с предположением, что F является истинной функцией распределения, из которого извлекается выборка (гипотезой о том, что F есть функция распределения наблюдаемой случайной величины X).

Итак, мы установили, что эмпирическое распределение сходится по вероятности к истинному (или, как обычно говорят прикладники, теоретическому) распределению, и теперь можем обратиться к вычислению моментных и квантильных характеристик распределения F_n. Его нецентральные

$$a_k = \int_{\mathrm{R}} x^k dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

и центральные

$$m_k = \int_{\mathrm{R}} (x - a_1)^k dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} (X_i - a_1)^k$$

моменты служат выборочными аналогами соответствующих теоретических моментов $\alpha_k,\ k = 1, 2 \ldots$, и $\mu_k,\ k = 2, 3, \ldots$, и называются *выборочными моментами.*

Если теоретические моменты существуют, то в силу закона больших чисел выборочные моменты сходятся по вероятности к своим теоретическим прообразам. Среди выборочных моментов особое место занимают моменты первого и второго порядков. Выборочный момент a_1 называется *выборочным средним* и имеет специальное обозначение $\overline{X}$; *выборочная дисперсия* $m_2 = a_2 - \overline{X}^2$ обычно обозначается S^2. Соответствующим образом определяются *выборочный коэффициент асимметрии* $g_1 = m_3/S^3$ и *выборочный коэффициент эксцесса* $g_2 = m_4/S^4 - 3$.

При выборе из m-мерного, $m > 1$, распределения эмпирическое распределение также приписывает массу n^{-1} каждому выборочному (векторному) значению $X_k = (X_{k1}, \ldots, X_{km}),\ k = 1, \ldots, n$. В соответствии с этим мы можем определить вектор выборочных средних $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_m)$ с компонентами

$$\overline{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ki},\ k = 1, \ldots, m,$$

выборочную ковариационную матрицу $S = \| S_{kj} \|$ с элементами

$$S_{kj} = \frac{1}{n}\sum_{i=1}^{n}(X_{ki} - \overline{X}_k)(X_{ji} - \overline{X}_j) = \frac{1}{n}\sum_{i=1}^{n} X_{ki}X_{ji} - \overline{X}_k\overline{X}_j,\ k, j = 1, \ldots, m,$$

и матрицу *выборочных коэффициентов корреляции* $R = \| r_{kj} \|$ с элементами $r_{kj} = S_{kj}/\sqrt{S_{kk}S_{jj}},\ k, j = 1, \ldots, m$. Смешанные моменты более высоких порядков в многомерном случае обычно не вычисляются.

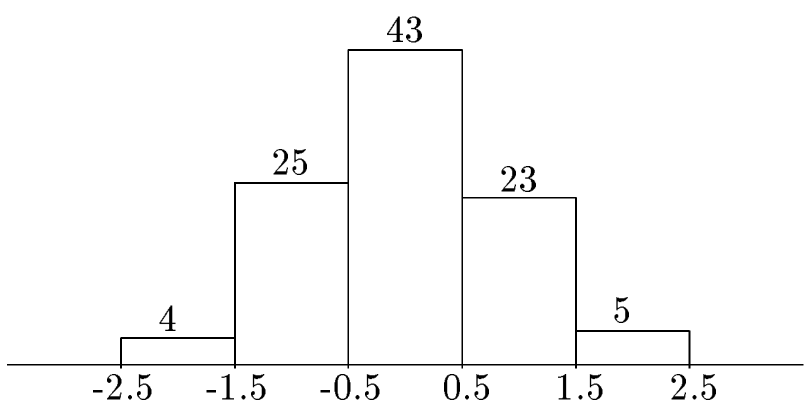
Если выбор происходит из распределения, для которого справедлива теорема сложения (предложение 12.2 курса ТВ), то распределение выборочного среднего устанавливается достаточно просто. В общем же случае можно только утверждать об асимптотической ($n \to \infty$) нормальности этой статистики при условии существования второго момента у теоретического распределения. Аналогичное утверждение справедливо и для моментов любого k-го порядка, если у $F(x)$ существует момент порядка $2k$.

Обратимся теперь к выборочным аналогам квантилей распределения F наблюдаемой случайной величины X. Напомним, что для непрерывного распределения квантиль порядка p определялась как решение x_p уравнения $F(x) = p$, а в случае дискретного распределения – как наибольшее $x = x_p$ из носителя распределения, при котором $F(x_p) \leq p$. Поскольку эмпирическое распределение дискретно, и его функция распределения $F_n(\cdot)$ возрастает скачками в точках, соответствующих компонентам вариационного ряда, то *выборочная квантиль* порядка p полагается равной порядковой статистике $X_{([\,np\,])}$, где $[\,x\,]$, как обычно, означает целую часть x. Естественно, для повышения точности оценки истинной квантили x_p можно проводить интерполяцию между статистиками $X_{([\,np\,])}$ и $X_{([\,np\,]+1)}$. Так, *выборочная медиана*, будучи квантилью порядка $p = 0.5$, обычно определяется как $\left(X_{([\,n/2\,])} + X_{([\,n/2\,]+1)}\right)/2$. Что же касается оценки моды распределения – точки наибольшего сгущения выборочных данных, то здесь нам придется обратиться к выборочным аналогам функции плотности.

При больших объемах наблюдений выборочные данные обычно подвергаются группировке, при этом индивидуальные выборочные значения не приводятся, а указываются лишь количества наблюдений, попавших в интервалы некоторого разбиения множества $\mathcal{X}$ значений наблюдаемой случайной величины. Поясним процедуру группировки на примере выборки из непрерывного одномерного распределения, когда $\mathcal{X} = \mathrm{R}$.

В декартовой системе координат ось абсцисс разбивается на $r \geq 2$ интервалов $(-\infty,\, a_1],\, (a_1,\, a_2],\, \ldots,\, (a_{r-2},\, a_{r-1}],\, (a_{r-1},\, +\infty)$, причем внутренние интервалы выбираются, как правило, одинаковой длины: $a_i - a_{i-1} = \Delta,\; i = 2, \ldots, r-1$. Выборочные данные сортируются по интервалам разбиения и подсчитываются частоты $n_i,\; i = 1, \ldots, r$

попадания данных в каждый интервал. Над каждым внутренним интервалом рисуется прямоугольник высоты $n_i/n\Delta$, так что площадь n_i/n каждого прямоугольника с номером $i = 2, \ldots, r-1$ служит реализацией частотной оценки ν_i/n вероятности попадания наблюдаемой случайной величины X в соответствующий интервал. Здесь ν_i – статистика, которую можно записать с помощью индикаторов событий $A_{ij} = \{X_j \in (a_{i-1}, a_i]\}$, $i = 1, \ldots, r$, $a_0 = -\infty$, $a_r = +\infty$, $j = 1, \ldots, n$, а именно $\nu_i = \sum_1^n \mathrm{I}(A_i)$. Полученная таким образом случайная ступенчатая функция, принимающая нулевые значения на крайних интервалах $(-\infty, a_1]$, $(a_{r-1}, +\infty)$ и равная $\nu_i/n\Delta$ на внутренних интервалах с номерами $i = 2, \ldots, r-1$, называется *гистограммной оценкой* f_n функции плотности $f(x)$, $x \in \mathrm{R}$ распределения X, а ее реализация (ν_i заменяются на наблюдаемые частоты n_i, $i = 1, \ldots, r$) – *гистограммой* выборки $x^{(n)}$.



В математической статистике существует ряд теорем, устанавливающих, что при определенных условиях на плотность f, гистограммная оценка $f_n(x) \underset{P}{\to} f(x)$ при любом $x \in \mathrm{R}$, если $n \to \infty$ и одновременно $r \to \infty$, а $\Delta \to 0$ со скоростью, зависящей определенным образом от n и r.

В случае гистограммной оценки функции плотности естественно считать выборочным аналогом (оценкой) моды распределения X середину интервала разбиения, в котором гистограмма принимает наибольшее значение.

Заметим также, что вектор частот $(\nu_1, \ldots, \nu_r)$ имеет мультиномиальное распределение $\mathcal{M}(r, n, \mathbf{p})$ с вероятностями исходов $p_i = F(a_i) - F(a_{i-1})$, $i = 1, \ldots, r$, что позволяет найти распределение оценки $f_n(x)$ при любом $x \in \mathrm{R}$ и построить критерий согласия выборочных данных с гипотезой о виде распределения наблюдаемой случайной величины. Это широко используемый на практике *критерий хи-квадрат*, основанный на статистике (сравните с критерием Колмогорова D_n)

$$X^2 = \sum_1^r \frac{(\nu_i - np_i)^2}{np_i}\,.$$

Асимптотическое распределение этой статистики мы изучим в параграфе, посвященном статистической проверке гипотез.

Итак, мы рассмотрели основные выборочные аналоги распределения наблюдаемой случайной величины и его основных характеристик. Мы высказали также ряд утверждений о распределении этих статистик, что позволит нам в последующем вычислять последствия от их использования в качестве решающих функций. Для того, чтобы уяснить, насколько важно знать хотя бы среднее значение статистики, претендующей на роль решающей функции, обратимся снова к примеру 1.1 по аттестации партии дизельного топлива, где обсуждалась сопутствующая проблема оценки дисперсии σ^2 наблюдаемой случайной величины $X \sim \mathcal{N}(\mu,\, \sigma^2)$.

Предлагалось оценивать σ^2 по накопленному в лаборатории архиву испытаний аттестуемых партий дизельного топлива, то есть по данным большого числа N выборок $X_1^{(n)}, \ldots, X_N^{(n)}$ малого объема n. Каждая i-ая выборка извлекается из нормального $(\mu_i,\, \sigma^2)$ распределения, причем средние μ_i могут быть различными для разных выборок, $i = 1, \ldots, N$, но дисперсия σ^2 у всех выборок одна и та же. Предлагается следующая оценка σ^2. В каждой выборке вычисляется выборочная дисперсия S_i^2, $i = 1, \ldots, N$, и затем берется их арифметическое среднее: $\hat{\sigma}_N^2 = (1/N)\sum_1^N S_i^2$. Распределение каждой S_i^2 не зависит от μ_i, $i = 1, \ldots, N$, поскольку выборочная дисперсия инвариантна относительно сдвигов $X_k \to X_k + a$. Следовательно, предлагаемая оценка есть нормированная на N сумма независимых, одинаково распределенных случайных величин – копий статистики $S^2 = (1/n)\sum_1^n (X_k - \overline{X})^2$, и в силу закона больших чисел $\hat{\sigma}_N^2 \underset{P}{\to} \mathbf{E}S^2$ при неограниченном возрастании объема N архивных данных. Вычислим это математическое ожидание:

$$\mathbf{E}S^2 = \mathbf{E}\left(\frac{1}{n}\sum_1^n X_k^2 - \overline{X}^2\right) = \mathbf{E}X^2 - \mathbf{E}\overline{X}^2 = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2,$$

поскольку $\alpha_2 = \mathbf{E}X^2 = \mathbf{D}X + \mathbf{E}^2X$, а $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

Таким образом, предлагаемая оценка обладает значительным смещением при малом объеме n испытаний каждой партии дизельного топлива. Например, в случае $n = 2$ мы занижаем дисперсию в два раза, поскольку $\hat{\sigma}^2_N \underset{P}{\to} \sigma^2/2$. Естественно, этот дефект легко устраним – достаточно использовать исправленную на смещение оценку $\tilde{\sigma}^2_N = (n/(n-1))\hat{\sigma}^2_N$.

<u>*Лекция 4*</u>

В завершении этого параграфа мы изучим еще один класс замечательных статистик, используя которые можно редуцировать выборочные данные только к их значениям без потери информации. К сожалению, не все статистические структуры обладают такими статистиками, но, по существу, только в тех структурах, где имеются достаточные статистики, возможно построение оптимального статистического правила, на котором достигается минимум риска.

Идея, состоящая в том, что в определенных случаях для принятия решения без увеличения риска *достаточно* знать только значения некоторых статистик, а не все выборочные данные, не требует введения специальных мер информации, содержащейся в выборочных данных и статистиках, – все становится ясным при рассмотрении следующей простейшей задачи, с которой мы имели дело в самом начале курса теории вероятностей.

Предположим, что мы хотим узнать вероятность наследования доминантного признака в опытах Менделя и располагаем результатами $x_1, \ldots, x_n$ скрещиваний n пар, где, как обычно, каждое x_i есть индикатор наследования признака, $i = 1, \ldots, n$, а совокупность выборочных данных представляет реализацию случайной выборки $X_1, \ldots, X_n$ из двухточечного распределения с функцией плотности $f(x \,|\, \theta) = P_\theta(X = x) = \theta^x(1-\theta)^{1-x}$, отличной от нуля только в точках $x = 0$ и 1. Частотная оценка $\hat{\theta}_n = T/n$ вероятности θ наследования признака определяется статистикой $T = \sum_1^n X_k$, выборочное значение $t = \sum_1^n x_k$ которой соответствует числу потомков в эксперименте, наследовавших доминантный признак. Естественно, возникает вопрос, а нельзя ли извлечь дополнительную информацию о величине параметра θ из номеров $k_1, \ldots, k_t$ выборочных данных, принявших значение 1? Нетрудно понять, что это возможно только в том случае, если *распределение выборочного вектора $X^{(n)}$ при условии, что статистика T приняла фиксированное значение t, зависит от параметра* θ. Действительно, если мы будем наблюдать случайную величину, которая не имеет никакого отношения к интересующему нас параметру, то откуда этой информации взяться? Итак, найдем условное распределение $X^{(n)}$ относительно T.

Используя формулу условной вероятности, получаем, что

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T = t\right) = \frac{P_\theta\left(\{X^{(n)} = x^{(n)}\} \wedge \{\sum_1^n X_k = t\}\right)}{P_\theta\left(\sum_1^n X_k = t\right)}.$$

Если значения компонент вектора $x^{(n)}$ таковы, что $\sum_1^n x_k \neq t$, то события $X^{(n)} = x^{(n)}$ и $\sum_1^n X_k = t$, очевидно, несовместны, и поэтому в этом случае условная вероятность равна нулю (не зависит от θ). Если же $\sum_1^n x_k = t$, то событие $X^{(n)} = x^{(n)}$ влечет событие $\sum_1^n X_k = t$, и формула для вычисления условной вероятности упрощается:

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T = t\right) = \frac{P_\theta\left(X^{(n)} = x^{(n)}\right)}{P_\theta\left(\sum_1^n X_k = t\right)}.$$

Так как

$$P_\theta\left(X^{(n)} = x^{(n)}\right) = f_n(X^{(n)} \,|\, \theta) = \theta^{\sum_1^n x_k}(1-\theta)^{n-\sum_1^n x_k},$$

$$P_\theta\left(\sum_1^n X_k = t\right) = \mathrm{C}_n^t\, \theta^t (1-\theta)^{n-t},$$

то в случае $\sum_1^n x_k = t$ условное распределение выборочного вектора $X^{(n)}$ относительно статистики T имеет вид

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T = t\right) = \frac{1}{\mathrm{C}_n^t},$$

и также не зависит от θ.

Итак, наши выкладки показывают, что распределение выборочного вектора на "плоскости" $\sum_1^n X_k = t$ не зависит от θ, и поэтому расположение значений $x_k = 1$ в последовательности $x_1, \ldots, x_n$ при фиксированном количестве таких значений не несет информации о параметре θ.

Определение 2.1. Статистика $T = T(X^{(n)})$ называется *достаточной* для статистической структуры $\mathcal{P}_n = \{P_{\theta,n},\, \theta \in \Theta\}$, если условное распределение выборочного вектора $X^{(n)}$ относительно статистики T не зависит от θ.

В общей теории статистического вывода в рамках более общего определения статистического правила устанавливается замечательный факт: *если статистическая структура обладает достаточной статистикой T, то, каково бы ни было статистическое правило $\delta = \delta(X^{(n)})$, всегда существует правило $\delta^* = \delta^*(T)$, основанное только на T, риск которого совпадает с риском правила δ.* Таким образом, построение оптимальных статистических правил следует начинать с поиска достаточных статистик. Следующая теорема дает критерий существования у статистических структур достаточных статистик и, одновременно, указывает простой способ их нахождения.

Теорема 2.1. *Для того, чтобы $T = T(X^{(n)})$ была достаточной статистикой для статистической структуры, определяемой функцией плотности $f_n(x^{(n)} \,|\, \theta)$, необходимо и достаточно,чтобы эта функция допускала представление*

$$f_n(x^{(n)} \,|\, \theta) = g_\theta\left(T(x^{(n)})\right) h(x^{(n)}), \tag{1}$$

где функция h не зависит от параметра θ, а функция g зависит от θ и аргумента $x^{(n)}$ только через значения $T(x^{(n)})$ статистики $T = T(X^{(n)})$,

Доказательство. теоремы мы проведем только для дискретного распределения наблюдаемой случайной величины, когда функция плотности выборки $f_n(x^{(n)} \,|\, \theta) = P_\theta\left(X^{(n)} = x^{(n)}\right)$. В случае непрерывного распределения схема доказательства та же, но придется делать замену в n-кратном интеграле.

Достаточность. Пусть выполняется факторизационное представление (1); требуется показать, что условное распределение $X^{(n)}$ относительно T не зависит от θ. Как и в только что рассмотренном примере с двухточечным распределением, воспользуемся формулой условной вероятности для вычисления условной плотности $X^{(n)}$ относительно T :

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t\right) = \frac{P_\theta\left(\{X^{(n)} = x^{(n)}\} \wedge \{T(X^{(n)}) = t\}\right)}{P_\theta(T(X^{(n)}) = t)}.$$

События, стоящие в числителе, будут несовместными, если $T(x^{(n)}) \neq t$, и в этом случае условная вероятность равна нулю (не зависит от θ). Если же $T(x^{(n)}) = t$, то первое по порядку событие в числителе влечет второе, и поэтому формула для вычисления условной вероятности упрощается:

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t\right) = \frac{P_\theta\left(X^{(n)} = x^{(n)}\right)}{P_\theta(T(X^{(n)}) = t)}.$$

Так как $P_\theta(X^{(n)} = x^{(n)}) = f_n(x^{(n)} \,|\, \theta)$, то используя представление (1), получаем, что (напомним, $T(x^{(n)}) = t$)

$$P_\theta\left(X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t\right) =$$

$$\frac{g_\theta(T(x^{(n)})h(x^{(n)})}{\sum\limits_{y^{(n)}:\, T(y^{(n)})=t} g_\theta(T(y^{(n)})h(y^{(n)})} = \frac{h(x^{(n)})}{\sum\limits_{y^{(n)}:\, T(y^{(n)})=t} h(y^{(n)})}.$$

Таким образом, условное распределение не зависит от θ, и поэтому статистика T достаточна для $\mathcal{P}_n$.

Необходимость. Пусть T – достаточная статистика, так что условное распределение $P_\theta\left(X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = t\right) = K(x^{(n)}, t)$, где функция K не зависит от θ. Требуется показать, что в этом случае для функции плотности выборки справедливо представление (1).

Имеем

$$f_n(x^{(n)} \,|\, \theta) = P_\theta\left(X^{(n)} = x^{(n)}\right) =$$

$$P_\theta\left(\{X^{(n)} = x^{(n)}\} \bigwedge \{T(X^{(n)}) = T(x^{(n)})\}\right) =$$

$$P_\theta(T(X^{(n)}) = T(x^{(n)})) \cdot P_\theta\left(X^{(n)} = x^{(n)} \,|\, T(X^{(n)}) = T(x^{(n)})\right).$$

Мы получили представление (1) с $g_\theta(T(x^{(n)})) = P_\theta(T(X^{(n)}) = T(x^{(n)}))$ и $h(x^{(n)}) = K(x^{(n)}, T(x^{(n)}))$. Теорема доказана.

Рассмотрим несколько примеров на применения полученного критерия достаточности к статистическим структурам, соответствующим вероятностным моделям из нашего курса теории вероятностей. Начнем с двухточечного распределения (выбор в схеме Бернулли), где мы непосредственными вычислениями условного распределения убедились в достаточности статистики, реализующей число успешных испытаний, – посмотрим, как это делается с помощью представления (1).

1^0. *Двухточечное распределение* $\mathrm{B}(1,\theta)$ имеет функцию плотности

$$f(x\,|\,\theta)=\theta^{\,x}(1-\theta)^{1-x},$$

отличную от нуля только в точках $x=0$ и 1. Параметрическое пространство этого распределения $\Theta=(0;\,1)$, а функция плотности случайной выборки

$$f_n(x^{(n)}\,|\,\theta)=\theta^{\sum_1^n x_k}(1-\theta)^{n-\sum_1^n x_k}.$$

Представление (1) выполняется с $h(x^{(n)})\equiv 1$ и $T(x^{(n)})=\sum_1^n x_k$. Следовательно, $T=\sum_1^n X_k$ – достаточная статистика.

2^0. *Распределение Пуассона* $\mathrm{P}(\theta)$, для которого

$$f(x\,|\,\theta)=\frac{\theta^{\,x}\,\mathrm{e}^{-\theta}}{x!},\ x=0,1,\ldots,\quad \Theta=\mathrm{R}_+,$$

функция плотности выборки

$$f_n(x^{(n)}\,|\,\theta)=\theta^{\sum_1^n x_k}\,\mathrm{e}^{-n\theta}\,/\,\prod\nolimits_1^n x_k!.$$

Следовательно, в представлении (1)

$$h(x^{(n)})=\left[\,\prod\nolimits_1^n x_k!\,\right]^{-1},$$

и $T=\sum_1^n X_k$ – достаточная статистика.

3^0. *Показательное распределение* $\mathrm{E}(\theta)$ с

$$f(x\,|\,\theta)=\frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\},\ x\geq 0,\quad \Theta=\mathrm{R}_+,$$

и

$$f_n(x^{(n)}\,|\,\theta)=\frac{1}{\theta^n}\exp\left\{-\frac{1}{\theta}\sum_1^n x_k\right\}$$

Также обладает достаточной статистикой $T=\sum_1^n X_k$.

4^0. *Равномерное распределение* $\mathrm{U}(a,\,b)$, функция плотности которого

$$f(x\,|\,\theta)=\frac{\mathrm{I}_{[\,a;\,b\,]}(x)}{b-a}$$

отлична от нуля и постоянна на отрезке $[\,a;\,b\,]$, на что указывает стоящая в числителе индикаторная функция отрезка $[\,a;\,b\,]$. В этом распределении $\theta = (a,\, b)$ – двумерный параметр и параметрическое пространство $\Theta = \{(a,\, b) : (a,\, b) \in \mathrm{R}^2,\, a < b\}$. Статистическая структура определяется функцией плотности

$$f_n(x^{(n)} \,|\, \theta) = \frac{\prod_1^n \mathrm{I}_{[\,a;\, b]}(x_k)}{(b-a)^n},$$

и поскольку произведение индикаторов, стоящее в числителе этой функции, принимает значение 1 в случае

$$a \le \min_{1 \le k \le n} x_k \le \max_{1 \le k \le n} x_k \le b$$

и значение 0 при нарушении этих неравенств, то вектор $T = (\,X_{(1)},$ $X_{(n)}\,)$ крайних членов вариационного ряда является достаточной статистикой.

5^0. *Нормальное распределение* $\mathcal{N}(\mu,\, \sigma^2)$. Это распределение обладает двумерным параметром $\theta = (\mu,\, \sigma)$ с областью значений (параметрическим пространством) $\Theta = \mathrm{R} \times \mathrm{R}_+$. Функции плотности наблюдаемой случайной величины X и случайной выборки $X^{(n)}$ определяются соответственно как

$$f(x \,|\, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

и

$$f_n(x^{(n)} \,|\, \theta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_1^n (x_k - \mu)^2\right\} =$$

$$\frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_1^n x_k^2 - 2\mu \sum_1^n x_k + n\mu^2\right)\right\}.$$

Последнее выражение для плотности $X^{(n)}$ показывает, что двумерная статистика $T = (T_1,\, T_2)$ с $T_1 = \sum_1^n X_k$ $T_1 = \sum_1^n X_k^2$ достаточна для статистической структуры нормального распределения. Кроме того, поскольку $T_1 = n\overline{X}$ и $T_2 = n(S^2 + \overline{X}^2)$, то факторизационное равенство (1) указывает на достаточность статистик $\overline{X}$ и S^2, которые имеют конкретную статистическую интерпретацию и поэтому более удобны для практического использования. Понятно, что это замечание носит общий характер: *любые взаимно однозначные преобразования достаточной статистики наследуют свойство достаточности.*

Отметим также, что в случае известного (фиксированного) σ статистическая структура имеет параметрическое пространство, совпадающее с областью значений параметра μ, и достаточной статистикой будет выборочное среднее $\overline{X}$. Аналогичное утверждение имеет место для статистики S^2 при фиксированном μ.

6^0. *Гамма-распределение* $\mathrm{G}(\lambda,\, a)$ имеет функцию плотности

$$f(x \,|\, \theta) = \frac{1}{a^{\lambda}\Gamma(\lambda)} x^{\lambda-1} \exp\left\{-\frac{x}{a}\right\},\ x > 0, \quad \theta = (a,\, \lambda),\, a > 0,\, \lambda > 0,$$

так что функция плотности выборочного вектора

$$f(x \,|\, \theta) = \frac{1}{a^{n\lambda}\Gamma^n(\lambda)} \left[\prod_1^n x_k\right]^{\lambda-1} \exp\left\{-\frac{1}{a}\sum_1^n x_k\right\}.$$

Тождество (1) указывает, что достаточной является двумерная статистика $\left(\sum_1^n X_k,\, \prod_1^n X_k\right)$ или более удобная в вычислительном отношении статистика $\left(\sum_1^n X_k,\, \sum_1^n \ln X_k\right)$. Для этого распределения можно сделать то же замечание, что и для нормального: первая компонента достаточной статистики "отвечает" за масштабный параметр a, в то время как вторая соответствует параметру формы λ.

7^0. *Биномиальное распределение* $\mathrm{B}(m,\, p)$. Это дискретное распределение, сосредоточенное в точках $x = 0, 1, \ldots, m$, с функцией плотности

$$f(x \,|\, \theta) = \mathrm{C}_m^x\, p^x (1-p)^{m-x},$$

зависящей от двумерного параметра $\theta = (m,\, p)$, первая компонента m которого может принимать только значения из множества $\mathrm{N} = \{1, 2, \ldots\}$, а вторая компонента $p \in (0;\, 1)$. Функция плотности выборочного вектора

$$f_n(x^{(n)} \,|\, \theta) = \prod_{k=1}^n \mathrm{C}_m^{x_k} \cdot p^{\sum_1^n x_k} (1-p)^{nm - \sum_1^n x_k}.$$

Применение критерия (1) показывает, что для статистической структуры с параметрическим пространством $\Theta = \mathrm{N} \times (0;\, 1)$ достаточной статистикой может быть только весь выборочный вектор $X^{(n)}$, но если $\Theta = (0;\, 1)$ (значение параметра m известно), то $\sum_1^n X_k$ – достаточная статистика.

8^0. *Распределение Коши* $\mathrm{C}(a,\, b)$ имеет функцию плотности выборочного вектора

$$f_n(x^{(n)}\,|\,\theta) = \pi^{-n} b^{-n} \prod_{k=1}^{n} \left(1 + \left(\frac{x_k - a}{b}\right)^2\right)^{-1},$$

и в силу критерия (1) его статистическая структура обладает только *тривиальной* достаточной статистикой $T = X^{(n)}$.

Мы не будем выписывать статистические структуры многомерных распределений в силу их чрезвычайной громоздкости, но нетрудно установить по аналогии с рассмотренными примерами, что у структуры мультиномиального распределения $\mathcal{M}(m, 1, \mathbf{p})$ с $m \geq 2$ исходами и вектором $\mathbf{p} = (p_1, \ldots, p_m)$ вероятностей соответствующих исходов достаточным будет вектор, состоящий из частот этих исходов в мультиномиальной схеме испытаний, а у структуры многомерного нормального распределения $\mathcal{N}_m(\mu,\, \Lambda)$ достаточную статистику образуют вектор выборочных средних и выборочная ковариационная матрица.

На этом завершается вводная часть нашего курса математической статистики. Мы сделали постановку проблемы статистического вывода, провели классификацию основных статистических структур и теперь мы готовы к решению конкретных статистических проблем по оценке параметров распределения наблюдаемой случайной величины и проверке гипотез, касающихся структуры параметрического пространства этого распределения.