

Конспект лекций по теории вероятностей и
математической статистике

А.А. Соловьев

Оглавление

1. Вероятностное пространство	3
1.1. Предмет теории вероятностей	3
1.2. Вероятностные пространства	4
1.3. Дискретные вероятностные пространства	5
1.4. Условные вероятности, независимость	7
1.5. Независимые испытания	10
2. Случайные величины	12
2.1. Случайные величины (конечная схема)	12
2.2. Случайные величины (общий случай)	26
2.3. Характеристические функции	38
2.4. Центральная предельная теорема	46
2.5. Многомерные характеристические функции	48
2.6. Многомерное нормальное распределение	50
2.7. Распределения, связанные с многомерным нормальным распределением	56
2.8. Закон больших чисел	58
3. Математическая статистика	63
3.1. Предмет математической статистики	63
3.2. Эмпирическая функция распределения	63
3.3. Выборочный метод	64
3.4. Понятие оценки	66
3.5. Асимптотическая нормальность выборочных моментов	72
3.6. Методы нахождения оценок	74
3.7. Доверительные интервалы	78
3.8. Непараметрические критерии проверки гипотез. Критерий значимости	82
3.9. Статистические гипотезы. Критерий Неймана – Пирсона	88

Глава 1

Вероятностное пространство

1.1. Предмет теории вероятностей

Закономерные события – это события, которые всегда происходят как только создаются определенные условия. Закономерные же явления – это система закономерных событий.

Математика, как и любая другая наука, изучает математические модели закономерных явлений окружающего нас мира.

Случайные же события – это события, которые при одних и тех же условиях происходят или нет. Массовые случайные события – это события, для которых можно создать одни и те же условия, при которых случайное событие может произойти или нет.

Однако и случайные события подчиняются закономерностям, которые называют вероятностными закономерностями.

Пусть при проведении эксперимента N раз событие A осуществляется $N(A)$ раз. Число $N(A)$ называется частотой события, а отношение $N(A)/N$ – относительной частотой события A . Оказывается, что при больших N для массовых случайных событий относительная частота $N(A)/N$ обладает свойством устойчивости, которое состоит в следующем: в одних и тех же условиях в нескольких сериях достаточно большого числа N_1, \dots, N_s наблюдений события A как правило имеет место приближенное равенство

$$\frac{N_1(A)}{N_1} \approx \frac{N_2(A)}{N_2} \approx \dots \approx \frac{N_s(A)}{N_s}.$$

Таким образом, относительная частота события A колеблется около числа, которое характеризует данное случайное событие A . Это число $\mathbf{P}(A)$, являющееся объективной мерой возможности наступления события A , называют вероятностью события A .

Например, пусть событие A – это выпадение герба при одном бросании монеты. Если монета симметричная и однородная, то $\mathbf{P}(A) = 1/2$.

Устойчивость частот – это объективное свойство массовых случайных явлений. Это свойство лежит в основе теории вероятностей как математической науки.

Теория вероятностей изучает математические модели случайных явлений и устанавливает связи между вероятностями случайных событий в математических моделях, которые позволяют вычислять вероятности сложных событий по вероятностям более простых событий.

Основным в теории вероятностей является понятие случайного события. Вот несколько примеры случайных событий:

- а) При бросании монеты n раз герб выпадает не менее k раз.
- б) При подбрасывании игральной кости выпадает четное число очков.
- в) Монета бросается до первого появления герба.

Достоверным событием называется событие, которое всегда происходит (будем обозначать достоверное событие, как правило, символом Ω).

Невозможное событие \emptyset – событие, которое никогда не происходит.

Противоположное событие \bar{A} – событие, которое происходит тогда и только тогда, когда не происходит событие A .

Суммой или объединением событий A и B называется событие, обозначаемое $A \cup B$, которое происходит тогда и только тогда, когда происходит или A , или B (или оба вместе).

Произведением или пересечением событий A и B назовем событие, обозначаемое $A \cap B$ или AB , которое происходит тогда и только тогда, когда происходят события A и B вместе.

Разностью $A \setminus B$ событий A и B называется событие, которое происходит тогда и только тогда, когда происходит A и не происходит B .

События A и B называются несовместными, если $AB = \emptyset$. Сумму попарно несовместных событий A и B будем обозначать через $A + B$.

Будем писать $A \subseteq B$ и говорить, что событие A влечет за собой событие B , если из наступления события A следует наступление события B .

Если $A \subseteq B$ и $B \subseteq A$, то события A и B будем называть равносильными и писать $A = B$.

Возможные взаимно исключающие исходы эксперимента называются *элементарными событиями*.

Общепризнанным считается подход, когда события отождествляются с подмножествами множества всевозможных элементарных событий. Тогда объединение, пересечение и противоположное событие совпадают с соответствующими теоретико-множественными операциями, а класс всех событий, связанных с экспериментом и включающих интересующее нас событие, должен быть замкнут относительно этих операций. Вероятность же есть числовая характеристика класса событий, свойства которой должны быть аналогичны свойствам относительной частоты осуществления события.

1.2. Вероятностные пространства

Начнем с определений.

Определение 1. Назовем класс \mathcal{A} подмножеств пространства Ω *алгеброй множеств*, если

- 1) $\emptyset \in \mathcal{A}$, $\Omega \in \mathcal{A}$;
- 2) из $A \in \mathcal{A}$ следует $\bar{A} \in \mathcal{A}$;
- 3) из $A, B \in \mathcal{A}$ следует $A \cup B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$.

Определение 2. Алгебра множеств \mathcal{A} называется σ -алгеброй, если из $A_n \in \mathcal{A}$, $n = 1, 2, \dots$, следует

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}, \quad \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}.$$

Определение 3. Тройку $(\Omega, \mathcal{U}, \mathbf{P})$, где Ω – пространство элементарных событий, \mathcal{U} – σ -алгебра подмножеств Ω , называемых случайными событиями, \mathbf{P} – числовая функция, определенная на \mathcal{U} и называемая вероятностью, будем называть вероятностным пространством, если \mathbf{P} удовлетворяет следующим аксиомам:

- 1° $\mathbf{P}(A) \geq 0$ для всех $A \in \mathcal{U}$ (неотрицательность \mathbf{P});
- 2° $\mathbf{P}(\Omega) = 1$ (нормированность \mathbf{P});
- 3° $\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$, если $A \cap B = \emptyset$ (аддитивность \mathbf{P});
- 4° Если события A_n в последовательности $\{A_n\}$ попарно несовместны, то

$$\mathbf{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

Свойства вероятностной функции на σ -алгебре \mathcal{U} :

- 1) $\mathbf{P}(\emptyset) = 0$;
- 2) Если $A \subseteq B$, то $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(B)$ и, в частности, $\mathbf{P}(A) \leq \mathbf{P}(B)$;
- 3) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$;
- 4) $\mathbf{P}(\cup_k A_k) \leq \sum_k \mathbf{P}(A_k)$;
- 5) Если $A_k \uparrow A$, т.е. $A_1 \subseteq A_2 \subseteq \dots$ и $\bigcup_{k=1}^{\infty} A_k = A$, то $\mathbf{P}(A) = \lim \mathbf{P}(A_k)$;
- 6) Если $A_k \downarrow A$, т.е. $A_1 \supseteq A_2 \supseteq \dots$ и $\bigcap_{k=1}^{\infty} A_k = A$, то $\mathbf{P}(A) = \lim \mathbf{P}(A_k)$.

Определение 4. Если некоторое свойство Q выполняется на вероятностном пространстве $(\Omega, \mathcal{U}, \mathbf{P})$ с вероятностью 1, то говорят, что свойство Q имеет место почти наверное (п.н.).

1.3. Дискретные вероятностные пространства

Пусть пространство элементарных событий Ω – не более чем счетное множество, σ -алгебра случайных событий \mathcal{U} – система всех подмножеств множества Ω и вероятностная функция определяется заданием вероятности $\mathbf{P}(\omega)$ каждого исхода $\omega \in \Omega$, удовлетворяя соотношению $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$. Тройка $(\Omega, \mathcal{U}, \mathbf{P})$ называется дискретным вероятностным пространством. Вероятность события $A \in \Omega$ задается равенством

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Частным случаем дискретного вероятностного пространства является так называемое классическое вероятностное пространство, когда вероятности всех исходов равны друг другу. Классическая модель вероятностного пространства используется тогда, когда элементарные события равновозможны. Обозначим символом $|A|$ число элементов множества A . Так как $1 = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = \mathbf{P}(\omega)|\Omega|$, поэтому

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|}.$$

Элементы комбинаторики

Установленный в конечном множестве порядок называется перестановкой. Множество с заданным порядком называется упорядоченным множеством.

Конечное упорядоченное множество называется размещением. Число размещений A_n^k объема k в множестве из n элементов равно

$$A_n^k = \frac{n!}{(n-k)!}.$$

Конечные подмножества конечного множества называются сочетаниями. Число сочетаний объема m в множестве из n элементов равно

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Урновые схемы

Выборка без возвращения. Пусть имеется урна с N одинаковыми неразличимыми на ощупь шарами, занумерованных числами $1, 2, \dots, N$. Предположим, что шары с номерами $1, 2, \dots, M$ белого цвета, остальные – черного. Выборка без возвращения состоит в последовательном извлечении n шаров, не возвращая их обратно. В этом случае за пространство элементарных событий Ω можно принять множество всех упорядоченных наборов $(\omega_1, \omega_2, \dots, \omega_n)$ чисел ω_k , $1 \leq \omega_k \leq N$, не равных друг другу. Исходы будем считать равновозможными. Общее число исходов в этом случае равно A_N^n – числу размещений из N элементов по n . Вычислим вероятность события B_m , состоящего в том, что среди извлеченных n шаров имеется ровно m белых. Число благоприятных исходов определяется как произведение числа способов выбора m координат из общего их количества n , на которые мы разместим номера белых шаров, A_M^m – числа различных способов разместить белые шары на m мест и A_{N-M}^{n-m} – числа различных способов разместить черные шары на оставшиеся $n-m$ мест. В итоге, получаем

$$\mathbf{P}(B_m) = \frac{C_n^m A_M^m A_{N-M}^{n-m}}{A_N^n}.$$

Пользуясь комбинаторными формулами, вероятность $\mathbf{P}(B_m)$ можно представить в виде

$$\mathbf{P}(B_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Если нас не интересует порядок извлечения шаров, а только состав выборки, то выбрать m белых шаров из n шаров можно $C_M^m C_{N-M}^{n-m}$ способами при общем количестве исходов, равном C_N^n . Поэтому и в этом случае вероятность события B_m будет равна

$$\mathbf{P}(B_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Таким образом, в приведенном примере выборки без возвращения элементарными событиями можно считать как упорядоченные, так и не упорядоченные выборки.

Выборки с возвращением. Пусть имеется та же урна, из которой последовательно извлекается по одному шару, фиксируется номер шара, а сам шар возвращается обратно в урну. В этом случае пространство элементарных событий состоит из векторов $(\omega_1, \omega_2, \dots, \omega_n)$, координаты которых $1 \leq \omega_j \leq N$. Поэтому

$$|\Omega| = N^n,$$

а вероятность события B_m равна

$$\mathbf{P}(B_m) = C_n^m \frac{M^m (N-M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m}.$$

Если же нас интересует только состав шаров, а не последовательность, в которой эти шары извлекаются, элементарными событиями будут векторы (n_1, \dots, n_N) , где $n_1 + \dots + n_N = n$. Здесь n_k число извлеченных шаров с номером k . В этом случае вероятность события B_m равна

$$\mathbf{P}(B_m) = \frac{C_{M+m-1}^m C_{N-M+n-m-1}^{n-m}}{C_{N+n-1}^n}.$$

Этот результат просто выводится из леммы

Лемма 1.1. Число возможных способов решить уравнение

$$r_1 + \dots + r_n = r, \quad r_k \geq 0,$$

равно $C_{n+r-1}^r = C_{n+r-1}^{n-1}$.

Биномиальная схема

Пусть $(\Omega, \mathcal{U}, \mathbf{P})$ вероятностное пространство, в котором элементарными событиями являются всевозможные последовательности из нулей и единиц объема n . Если $\omega \in \Omega$, положим $\mathbf{P}(\omega) = p^k (1-p)^{n-k}$, если ω содержит k единиц. Ясно, что $\mathbf{P}(\Omega) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = 1$. Пусть событие A_k состоит из всех исходов, содержащих k единиц. Вероятности каждого такого исхода равна $p^k (1-p)^{n-k}$, поэтому вероятность события A_k равна

$$\mathbf{P}(A_k) = C_n^k p^k (1-p)^{n-k}.$$

Набор вероятностей $\{C_n^k p^k (1-p)^{n-k}\}_{k=0}^n$ называется биномиальным распределением вероятностей, а вероятностная схема – биномиальной схемой.

В случае $n = 1$ вероятностная схема называется схемой Бернулли.

Геометрические вероятности

Пусть Ω – область евклидова пространства \mathbf{R}^n с конечным n -мерным объемом. За алгебру событий примем σ -алгебру \mathcal{B} борелевских подмножеств Ω . Событиями назовем борелевские подмножества Ω . За вероятность события $A \in \mathcal{B}$ примем

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|},$$

где через $|V|$ обозначается n -мерный объем множества V .

Тогда тройка $(\Omega, \mathcal{B}, \mathbf{P})$ является вероятностным пространством.

Пример 1. Задача Бюффона. На плоскости, расчерченной параллельными прямыми, находящимися на расстоянии a друг от друга, случайно бросается игла длины $\ell < a$. Найти вероятность $\mathbf{P}(A)$ события $A = \{\text{игла пересечет какую-нибудь из параллельных прямых}\}$.

Обозначим через y расстояние от середины иглы до ближайшей прямой, x – острый угол между иглой и перпендикуляром к параллельным прямым. Координаты (x, y) определяют положение иглы относительно параллельных прямых и удовлетворяют условиям $0 \leq x \leq \pi/2$, $0 \leq y \leq a/2$. На плоскости (x, y) они образуют прямоугольник Ω . Игла пересечет одну из параллельных прямых, если x и y удовлетворяют неравенству

$$0 \leq y \leq \frac{\ell}{2} \cos x.$$

Искомая вероятность равна

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{\int_0^{\pi/2} \frac{\ell}{2} \cos x dx}{a/2 \cdot \pi/2} = \frac{2\ell}{a\pi}.$$

1.4. Условные вероятности, независимость

Условные вероятности

Пусть при N испытаниях события A , B и AB произошли с частотами N_A , N_B и N_{AB} . Если имеет место устойчивость частот

$$\frac{N_A}{N} \approx \mathbf{P}(A), \quad \frac{N_B}{N} \approx \mathbf{P}(B) \quad \text{и} \quad \frac{N_{AB}}{N} \approx \mathbf{P}(AB)$$

и $\mathbf{P}(B) > 0$, то относительная частота N_{AB}/N_B события A при условии, что произошло событие B также устойчива

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} \approx \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}.$$

Определение 5. Пусть $\mathbf{P}(B) > 0$. Условной вероятностью $\mathbf{P}(A|B)$ события A при условии, что произошло событие B , назовем отношение

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}.$$

Условная вероятность $\mathbf{P}(A|B)$ обозначается также $\mathbf{P}_B(A)$.

Если случайное событие B из некоторого вероятностного пространства $(\Omega, \mathcal{U}, \mathbf{P})$ фиксировано, то условная вероятность $\mathbf{P}_B(A)$ как функция случайного события A определяет новое вероятностное пространство $(\Omega, \mathcal{U}, \mathbf{P}_B)$.

Теорема 1.1 (Вероятность произведения событий). Пусть события A_1, \dots, A_n таковы, что $\mathbf{P}(A_1 \dots A_{n-1}) > 0$. Тогда

$$\mathbf{P}(A_1 \dots A_n) = \mathbf{P}(A_1) \mathbf{P}_{A_1}(A_2) \mathbf{P}_{A_1 A_2}(A_3) \dots \mathbf{P}_{A_1 \dots A_{n-1}}(A_n). \quad (1)$$

Доказывается методом математической индукции.

формула полной вероятности

Определение 6. Систему событий A_1, A_2, \dots, A_n будем называть конечным разбиением, если они попарно несовместны и

$$A_1 + A_2 + \dots + A_n = \Omega.$$

Теорема 1.2 (Формула полной вероятности). Если A_1, \dots, A_n – разбиение и все $\mathbf{P}(A_k) > 0$, то для любого события B имеет место формула

$$\mathbf{P}(B) = \sum_{k=1}^n \mathbf{P}(A_k) \mathbf{P}(B|A_k), \quad (2)$$

называемая формулой полной вероятности.

Доказательство. Ясно, что

$$B = B\Omega = BA_1 + BA_2 + \dots + BA_n,$$

поэтому

$$\mathbf{P}(B) = \sum_{k=1}^n \mathbf{P}(BA_k).$$

Применяя к $\mathbf{P}(BA_k)$ теорему умножения получаем (2). \square

Формула Байеса

Теорема 1.3. Если выполнены условия теоремы о полной вероятности и $\mathbf{P}(B) > 0$, то имеет место формула

$$\mathbf{P}(A_k|B) = \frac{\mathbf{P}(A_k) \mathbf{P}(B|A_k)}{\sum_{i=1}^n \mathbf{P}(A_i) \mathbf{P}(B|A_i)}.$$

Доказательство. По теореме умножения

$$\mathbf{P}(A_k B) = \mathbf{P}(A_k) \mathbf{P}(B|A_k) = \mathbf{P}(B) \mathbf{P}(A_k|B),$$

поэтому

$$\mathbf{P}(A_k|B) = \frac{\mathbf{P}(A_k) \mathbf{P}(B|A_k)}{\mathbf{P}(B)}$$

Осталось применить к знаменателю $\mathbf{P}(B)$ формулу полной вероятности. \square

Назовем события A_k гипотезами. Пусть событие B результат некоторого эксперимента. Вероятности $\mathbf{P}(A_k)$ – это *априорные* вероятности гипотез, вычисляемые до проведения опыта. Условные вероятности $\mathbf{P}(A_k|B)$ – это *апостериорные* вероятности, вычисляемые после проведения опыта.

Независимые события

Пусть вероятность события $\mathbf{P}(B)$ не равна нулю. Естественно считать, что событие A не зависит от события B , если $\mathbf{P}(A|B) = \mathbf{P}(A)$. Если при этом и $\mathbf{P}(A) > 0$, то

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B) \mathbf{P}(A|B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B) \mathbf{P}(A)}{\mathbf{P}(A)} = \mathbf{P}(B),$$

т.е. B не зависит от A . Из теоремы умножения тогда следует, что для независимых друг от друга событий A и B имеет место равенство $\mathbf{P}(AB) = \mathbf{P}(A) \mathbf{P}(B)$. Это приводит к следующему определению независимости.

Определение 7. События A и B называются независимыми, если

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B).$$

Если равенство не выполняется, то события будем называть зависимыми.

Обычно при построении вероятностной модели независимость событий не устанавливается, а постулируется на основе причинной независимости реальных прообразов.

Определение 8. События A_1, A_2, \dots, A_n называются независимыми, если для любых $1 \leq i_1 < i_2 < \dots < i_m \leq n$, $2 \leq m \leq n$, выполняются равенства

$$\mathbf{P}(A_{i_1} A_{i_2} \dots A_{i_m}) = \mathbf{P}(A_{i_1})\mathbf{P}(A_{i_2}) \dots \mathbf{P}(A_{i_m}).$$

В противном случае события называются зависимыми.

Из попарной независимости

$$\mathbf{P}(A_k A_m) = \mathbf{P}(A_k)\mathbf{P}(A_m) \quad k \neq m.$$

не следует независимости.

Пример 2. Пусть из чисел 2, 3, 5, 30 выбирается одно, причем все исходы равновозможны и имеют вероятность $1/4$. Введем через $A_k = \{\text{выбранное число делится на } k\}$. Легко видеть, что события A_2, A_3, A_5 попарно независимы, но зависимы в совокупности, так как

$$\mathbf{P}(A_2) = \mathbf{P}(A_3) = \mathbf{P}(A_5) = 1/2,$$

$$\mathbf{P}(A_2 A_3) = \mathbf{P}(A_2 A_5) = \mathbf{P}(A_3 A_5) = 1/4 \quad \text{и} \quad \mathbf{P}(A_2 A_3 A_5) = 1/4 \neq 1/8.$$

Из совместной независимости

$$\mathbf{P}(A_1 A_2 \dots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2) \dots \mathbf{P}(A_n).$$

не следует независимости.

Пример 3. Рассмотрим пространство Ω , состоящее из 36 упорядоченных пар (i, j) , где $i, j = 1, \dots, 6$. Будем считать, что все пары равновозможны. Тогда для событий $A = \{(i, j) : j = 1, 2, 5\}$, $B = \{(i, j) : j = 4, 5, 6\}$ и $C = \{(i, j) : i + j = 9\}$ имеют место соотношения

$$\mathbf{P}(AB) = \frac{1}{6} \neq \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(B),$$

$$\mathbf{P}(AC) = \frac{1}{36} \neq \frac{1}{18} = \mathbf{P}(A)\mathbf{P}(C),$$

$$\mathbf{P}(BC) = \frac{1}{12} \neq \frac{1}{18} = \mathbf{P}(B)\mathbf{P}(C).$$

Но в то же время

$$\mathbf{P}(ABC) = \frac{1}{36} = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

Лемма 1.2. 1° Если события A и B независимы, то события \bar{A} и B также независимы; 2° Если события A_1 и B и события A_2 и B также независимы и $A_1 A_2 = \emptyset$, то $A_1 + A_2$ и B независимы.

Определение 9. Алгебры (σ -алгебры) событий $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n \subset \mathcal{U}$ в вероятностном пространстве $(\Omega, \mathcal{U}, \mathbf{P})$ называются независимыми, если для любых $A_i \in \mathcal{U}_i$, $i = \overline{1, n}$, выполняется равенство

$$\mathbf{P}(A_1 A_2 \dots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2) \dots \mathbf{P}(A_n).$$

1.5. Независимые испытания

Под испытанием понимаем некоторый эксперимент, исходами которого служат случайные события. Поэтому моделью испытания является вероятностное пространство. Пусть проведены n испытаний, то есть даны вероятностные пространства

$$(\Omega_1, \mathcal{U}_1, \mathbf{P}_1), \dots, (\Omega_n, \mathcal{U}_n, \mathbf{P}_n).$$

Если эти испытания причинно независимы, то в вероятностном пространстве, являющемся моделью последовательности независимых испытаний, σ -алгебры \mathcal{U}_k , $k = 1, \dots, n$, должны быть независимыми. Мы можем об этом говорить только рассматривая σ -алгебры \mathcal{U}_k как σ -подалгебры σ -алгебры некоторого общего вероятностного пространства. Такое вероятностное пространство всегда можно построить. Прделаем это в случае, когда вероятностные пространства $(\Omega_k, \mathcal{U}_k, \mathbf{P}_k)$ конечны.

Пусть при каждом $k = 1, \dots, n$ $(\Omega_k, \mathcal{U}_k, \mathbf{P}_k)$ – конечное вероятностное пространство, где $\Omega_k = \{\omega_k\}$ – пространство элементарных событий, \mathcal{U}_k – алгебра всех подмножеств из Ω_k , $\mathbf{P}_k(A) = \sum_{\omega_k \in A} \mathbf{P}_k(\omega_k)$ – вероятность. Положим $\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n)\}$, где $\omega_k \in \Omega_k$, \mathcal{U} – алгебра всех подмножеств множества Ω ,

$$\mathbf{P}(\omega) = \mathbf{P}_1(\omega_1) \cdots \mathbf{P}_n(\omega_n), \quad \mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Введенная вероятность \mathbf{P} называется прямым произведением вероятностей \mathbf{P}_k и обозначается $\mathbf{P} = \mathbf{P}_1 \times \dots \times \mathbf{P}_n$. Аналогично, $\mathcal{U} = \mathcal{U}_1 \times \dots \times \mathcal{U}_n$ есть прямое произведение алгебр. В построенном вероятностном пространстве событие вида $A = A_1 \times \dots \times A_n$ будем называть прямоугольником. Из определения вероятности следует, что вероятность прямоугольника A равна

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega) = \sum_{\omega_1 \in A_1} \mathbf{P}_1(\omega_1) \cdots \sum_{\omega_n \in A_n} \mathbf{P}_n(\omega_n) = \prod_{k=1}^n \mathbf{P}_k(A_k).$$

Пусть \mathcal{U}'_k обозначает подалгебру всех прямоугольников вида

$$A'_k = \Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n \in \mathcal{U}'_k.$$

Между событиями $A'_i \in \mathcal{U}'_i$ и $A_i \in \mathcal{U}_i$ устанавливается естественный изоморфизм $A'_i \sim A_i$, сохраняющий вероятности $\mathbf{P}(A'_i) = \mathbf{P}_i(A_i)$. Поэтому события A_i из вероятностного пространства $(\Omega_i, \mathcal{U}_i, \mathbf{P}_i)$ будем отождествлять с изоморфными событиями A'_i из $(\Omega, \mathcal{U}, \mathbf{P})$. Так как $A = A_1 \times \dots \times A_n = \bigcap_{k=1}^n A'_k$, то из определения вероятности \mathbf{P} получаем

$$\mathbf{P}\left(\bigcap_{k=1}^n A'_k\right) = \prod_{k=1}^n \mathbf{P}(A'_k),$$

т.е. алгебры $\mathcal{U}'_1, \dots, \mathcal{U}'_n$ независимы. Построенное пространство будем называть прямым произведением вероятностных пространств $(\Omega_k, \mathcal{U}_k, \mathbf{P}_k)$, $k = 1, \dots, n$.

Еще раз о схеме Бернулли

Схема Бернулли — частный случай независимых испытаний с 2-я исходами в каждом испытании.

Пусть вероятностное пространство $(\Omega_i, \mathcal{U}_i, \mathbf{P}_i)$ таково, что $\Omega_i = \{0, 1\}$, $\mathcal{U}_i = \{\emptyset, \{0\}, \{1\}, \Omega_i\}$, $\mathbf{P}\{0\} = q$, $\mathbf{P}\{1\} = p = 1 - q$, для каждого $i = 1, \dots, n$.

В прямом произведении $(\Omega, \mathcal{U}, \mathbf{P})$ элементарный исход $\omega \in \Omega$ имеет вид $\omega = (\omega_1, \dots, \omega_n)$, $\omega_i = 0, 1$. Вероятность $\mathbf{P}\{\omega\}$ равна $\prod_{i=1}^n p^{\omega_i} (1 - p)^{1 - \omega_i}$. Исход $\{1\}$ будем называть *успехом*, противоположный

исход $\{0\}$ - *неуспехом*. Обозначим через $B_k = \{\omega : \omega_1 + \dots + \omega_n = k\}$ событие, состоящее в том, что при n независимых испытаниях в схеме Бернулли произошло ровно k успехов. Поскольку для любого исхода $\omega \in B_k$ имеем $\mathbf{P}\{\omega\} = p^k(1-q)^{n-k}$, то

$$\mathbf{P}(B_k) = C_n^k p^k q^{n-k}, k = 0, 1, \dots, n.$$

Напомним, что набор вероятностей $\{C_n^k p^k q^{n-k}\}_{k=0}^n$ называется *биномиальным распределением*.

Полиномиальная схема

Допустим, что в каждом независимом испытании возможно появление одного из r попарно несовместных исходов. Введем вероятностное пространство $(\Omega_i, U_i, \mathbf{P}_i)$, $i = 1, \dots, n$, полагая, что $\Omega_i = \{1, \dots, r\}$, p_1, \dots, p_r - вероятности исходов, $p_1 + \dots + p_r = 1$, и алгебра U_i состоит из всех подмножеств множества Ω_i .

В прямом произведении (Ω, U, \mathbf{P}) вероятностных пространств $(\Omega_i, U_i, \mathbf{P}_i)$, $i = 1, \dots, n$ элементарное событие $\omega \in \Omega$ имеет вид $\omega = (\omega_1, \dots, \omega_n)$, где ω_i - номер исхода при i -ом испытании. Полагаем $\mathbf{P}\{\omega\} = p_{\omega_1} p_{\omega_2} \dots p_{\omega_n}$.

Введем событие $B_{n_1, \dots, n_r} = \{ \text{в } n \text{ независимых испытаниях произошло ровно } n_k \text{ } k\text{-х исходов, } n_1 + n_2 + \dots + n_r = n \}$. Так как для любого $\omega \in B_{n_1, \dots, n_r}$

$$\mathbf{P}\{\omega\} = \prod_{k=1}^r p_k^{n_k}$$

и количество элементов в B_{n_1, \dots, n_r} равно $\frac{n!}{n_1! \dots n_r!}$, то $\mathbf{P}(B_{n_1, \dots, n_r}) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}$. Набор вероятностей $\{\mathbf{P}(B_{n_1, \dots, n_r}), n_1 + \dots + n_r = n\}$, называется *полиномиальным распределением*. Описанная схема с r исходами называется *полиномиальной*.

Глава 2

Случайные величины

2.1. Случайные величины (конечная схема)

Начнем с определения случайной величины.

Определение 10. Рассмотрим конечное вероятностное пространство (Ω, U, \mathbf{P}) . Числовую функцию от элементарного события $\xi = \xi(\omega)$, $\omega \in \Omega$ будем называть случайной величиной.

Пример 1. Число успехов при n испытаниях в схеме Бернулли — случайная величина.

Пример 2. Пусть в урне имеется N шаров, из них M белых, остальные черные. По схеме выбора без возвращения из урны извлекаются n шаров. Число извлеченных белых шаров в выборке объема n является случайной величиной.

Пример 3. С каждым событием $A \in U$ можно связать случайную величину

$$I_A = I_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A, \end{cases}$$

называемую индикатором события A .

Легко проверить следующие свойства индикаторов:

$$I_\emptyset = 0, I_\Omega = 1, I_{AB} = I_A I_B, I_{\bar{A}} = 1 - I_A.$$

Если события A_1, \dots, A_n попарно несовместны, то

$$I_{\sum_{k=1}^n A_k} = \sum_{k=1}^n I_{A_k}.$$

Определение 11. Пусть $x_1 < x_2 < \dots < x_n$ — значения, принимаемые случайной величиной ξ . Через A_i обозначим случайное событие $A_i = \{\omega \in \Omega : \xi(\omega) = x_i\}$. Тогда $\alpha_\xi = \{A_i\}_{i=1}^n$ есть разбиение Ω . Разбиение α_ξ порождает алгебру событий U_ξ , элементами которой являются события

$$\{\xi \in B\} = \{\omega : \xi(\omega) \in B\},$$

где B — любое числовое множество. Разбиение α_ξ и алгебру U_ξ будем называть порожденными случайной величиной ξ . Любое случайное событие $\{\xi \in B\}$ представимо в виде сумм $\sum_i A_i$, где суммирование ведется по тем i , для которых $x_i \in B$.

Нетрудно видеть, что случайная величина ξ представляется в виде

$$\xi = \xi(\omega) = \sum_{i=1}^n x_i I_{A_i}(\omega).$$

Определение 12. Законом распределения случайной величины будем называть функцию числового множества $B : B \rightarrow \mathbf{P}\{\xi \in B\}$. Закон распределения ξ определяется значениями x_1, \dots, x_n и вероятностями $p_i = \mathbf{P}\{\xi = x_i\} = p_i$, $i = \overline{1, n}$, $\sum_i p_i = 1$. Закон распределения удобно записать в форме таблицы:

ξ	x_1	x_2	\dots	x_n
P	p_1	p_2	\dots	p_n

В теории вероятности часто говорят о случайной величине с законом распределения, не указывая вероятностного пространства (Ω, U, \mathbf{P}) . Предполагается, что такое вероятностное пространство существует.

Простейшим вероятностным пространством, связанным с законом распределения $\mathbf{P}\{\xi \in B\}$ будет множество элементарных событий $\Omega = \{x_1, \dots, x_n\}$ с элементарными вероятностями $\mathbf{P}(x_i) = p_i$. Случайная величина определяется соотношением $\xi(x_i) = x_i$.

Примеры законов распределения.

1. *Биномиальный закон* для числа успехов μ в n независимых испытаниях в схеме Бернулли:

$$\mathbf{P}\{\mu = m\} = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \dots, n.$$

2. *Равномерное распределение* на $\{1, 2, \dots, N\}$:

$$\mathbf{P}\{\xi = m\} = \frac{1}{N}, \quad m = 1, 2, \dots, N.$$

Математическое ожидание

Определение 13. Математическое ожидание случайной величины $\xi = \xi(\omega)$ на вероятностном пространстве (Ω, U, \mathbf{P}) определяется как сумма

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega) \mathbf{P}(\omega)$$

Перечислим *свойства математического ожидания*. Пусть случайная величина ξ принимает значения x_1, \dots, x_k с вероятностями $p_1 = \mathbf{P}\{\xi = x_1\}, \dots, p_k = \mathbf{P}\{\xi = x_k\}$.

1) $MI_A = \mathbf{P}(A)$, так как

$$MI_A = \sum_{\omega \in \Omega} I_A(\omega) \mathbf{P}(\omega) = \sum_{\omega \in A} \mathbf{P}(\omega) = \mathbf{P}(A).$$

2) аддитивность: $M(\xi + \eta) = M\xi + M\eta$. По индукции получаем, что $M(\xi_1 + \dots + \xi_k) = M\xi_1 + \dots + M\xi_k$.

3) для любой константы c

$$M(c\xi) = cM\xi, \quad Mc = c.$$

4) если $\xi \geq \eta$, то $M\xi \geq M\eta$, в частности, если $\xi \geq 0$ и $M\xi = 0$, то $\mathbf{P}\{\xi = 0\} = 1$.

Действительно, если $\xi \geq 0$ и $M\xi = 0$, то $\forall \omega \in \Omega$ имеем $\xi(\omega) \mathbf{P}(\omega) = 0$, откуда из $\mathbf{P}(\omega) > 0$ следует $\xi(\omega) = 0$.

5) математическое ожидание ξ выражается через закон распределения случайной величины ξ формулой

$$M\xi = \sum_{i=1}^k x_i \mathbf{P}\{\xi = x_i\} = \sum_{i=1}^k x_i p_i.$$

В самом деле, $\xi = \sum_{i=1}^k x_i I_{\{\xi=x_i\}}$. Из свойств 1 – 3 следует

$$M\xi = \sum_{i=1}^k x_i MI_{\{\xi=x_i\}} = \sum_{i=1}^k x_i \mathbf{P}\{\xi = x_i\}.$$

6) если $g(x)$ - некоторая функция и $\eta = g(\xi)$ случайная величина, то

$$M\eta = Mg(\xi) = \sum_{i=1}^k g(x_i) \mathbf{P}\{\xi = x_i\}.$$

Доказывается как и свойство 5, надо только воспользоваться равенством

$$g(\xi) = \sum_{i=1}^k g(x_i) I_{\{\xi=x_i\}}.$$

Полагая $g(x) = x^n$, получаем

$$M\xi^n = \sum_{i=1}^k x_i^n \mathbf{P}\{\xi = x_i\}.$$

Пусть в ходе эксперимента, осуществляемого N раз, случайная величина ξ значение x_1 принимает N_1 раз, \dots , x_k — N_k раз. Среднее значение, принимаемое случайной величиной имеет вид

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i N_i = \sum_{i=1}^k x_i \frac{N_i}{N}.$$

Если N велико и эксперимент проводится в одних и тех же условиях, то относительная частота $\frac{N_i}{N}$ близка к вероятностям $p_i = \mathbf{P}\{\xi = x_i\}$. Поэтому среднее значение \bar{x} колеблется около $M\xi$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i N_i \approx \sum_{i=1}^k x_i p_i = M\xi$$

Таким образом, математическое ожидание статистически истолковывается как среднее значение случайной величины.

Определение 14. Математическое ожидание $M\xi^n$ называется n -м моментом случайной величины ξ . Абсолютным n -м моментом называется $M|\xi|^n$. Обозначим $M\xi$ через m . Центральным моментом n -го порядка называется $M(\xi - m)^n$ и абсолютным центральным моментом n -го порядка — $M|\xi - m|^n$.

Центральный момент второго порядка $M(\xi - m)^2$ называется дисперсией и обозначается через $D\xi$. Корень квадратный из дисперсии $\sqrt{D\xi}$ называется среднеквадратическим отклонением.

Дисперсия обладает следующими свойствами:

- 1) $D\xi = M\xi^2 - (M\xi)^2$;
- 2) $D\xi \geq 0$ и $D\xi = 0$ тогда и только тогда, когда $\mathbf{P}\{\xi = c\} = 1$;
- 3) для любой константы c имеют место равенства

$$D(c\xi) = c^2 D\xi, \quad D(\xi + c) = D\xi;$$

- 4) Неравенство Коши-Буняковского. Для любых двух величин ξ и η

$$|M\xi\eta| \leq \sqrt{M\xi^2 M\eta^2}.$$

Теорема 2.1 (вероятность суммы событий).

$$\begin{aligned} \mathbf{P}\left(\bigcup_{k=1}^n A_k\right) &= \sum_{k=1}^n \mathbf{P}(A_k) - \sum_{1 \leq k_1 < k_2 \leq n} \mathbf{P}(A_{k_1} A_{k_2}) + \sum_{1 \leq k_1 < k_2 < k_3 \leq n} \mathbf{P}(A_{k_1} A_{k_2} A_{k_3}) - \dots + \\ & \quad (-1)^{n-1} \mathbf{P}(A_1 A_2 \dots A_n) \end{aligned}$$

Доказательство. Для индикатора объединения событий A_1, \dots, A_n имеет место равенство

$$I_{\bigcup_{k=1}^n A_k} = 1 - \prod_{k=1}^n (1 - I_{A_k}) = \sum_{k=1}^n I(A_k) -$$

$$\sum_{1 \leq k_1 < k_2 \leq n} I_{A_{k_1} A_{k_2}} + \sum_{1 \leq k_1 < k_2 < k_3 \leq n} I_{A_{k_1} A_{k_2} A_{k_3}} - \dots + (-1)^{n-1} I_{A_1 A_2 \dots A_n}.$$

Осталось вычислить математическое ожидание обеих частей равенства и воспользоваться его аддитивностью. \square

Независимость случайных величин

Если ξ - случайная величина и $\{x_1, \dots, x_n\}$ принимаемые ею значения, то система случайных событий $A_i = \{\xi = x_i\}$, $i = 1, \dots, n$, является разбиением. Порождаемую случайной величиной ξ алгебру будем обозначать через U_ξ .

Определение 15. Случайные величины ξ_1, \dots, ξ_m независимы, если независимы порожденные ими алгебры $U_{\xi_1}, \dots, U_{\xi_m}$, т.е. для любых $B_1, \dots, B_m \subset \mathbb{R}$ выполняется равенство

$$\mathbf{P}\{\xi_1 \in B_1, \dots, \xi_m \in B_m\} = \prod_{k=1}^m \mathbf{P}\{\xi_k \in B_k\}$$

Дадим эквивалентное определение независимости.

Определение 16. Пусть ξ_i принимает значения $\{x_{i1}, \dots, x_{in_i}\}$. Случайные величины ξ_1, \dots, ξ_m независимы, если для любых $x_{1j_1}, \dots, x_{mj_m}$, $1 \leq j_k \leq n_k$,

$$\mathbf{P}\{\xi_1 = x_{1j_1}, \dots, \xi_m = x_{mj_m}\} = \prod_{k=1}^m \mathbf{P}\{\xi_k = x_{kj_k}\}.$$

Теорема 2.2. Если случайные величины ξ_1, \dots, ξ_n независимы и $g_i(x)$ - числовые функции, то случайные величины $\eta_i = g_i(\xi_i)$ также независимы.

Доказательство. Действительно,

$$\{\eta_i \in B_i\} = \{g_i(\xi_i) \in B_i\} = \{\xi_i \in g_i^{-1}(B_i)\}.$$

Поэтому $A_{g_i(\xi_i)} \subset A_{\xi_i}$. \square

Мультипликативное свойство математического ожидания

Доказательство следующей теоремы опирается на лемму, которая является обобщением предыдущей теоремы 2.2.

Лемма 2.1. Пусть случайные величины $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m$ независимы. Для любых числовых функций $g_1(x_1, \dots, x_n)$ и $g_2(y_1, \dots, y_m)$ случайные величины $\zeta_1 = g_1(\xi_1, \dots, \xi_n)$ и $\zeta_2 = g_2(\eta_1, \dots, \eta_m)$ независимы.

Теорема 2.3. Если случайные величины ξ_1, \dots, ξ_n независимы, то

$$M(\xi_1 \dots \xi_n) = \prod_{i=1}^n M\xi_i$$

Доказательство. Пусть ξ и η независимы и

$$\xi = \sum_{i=1}^k x_i I_{A_i}, \quad \text{где } A_i = \{\xi = x_i\}$$

$$\eta = \sum_{j=1}^m y_j I_{B_j}, \quad \text{где } B_j = \{\eta = y_j\}.$$

В силу аддитивности математического ожидания имеем

$$\begin{aligned} \xi\eta &= \sum_{i=1}^k \sum_{j=1}^m x_i y_j I_{A_i B_j}, & M\xi\eta &= \sum_{i=1}^k \sum_{j=1}^m x_i y_j \mathbf{P}(A_i B_j) = \\ & & \sum_{i=1}^k \sum_{j=1}^m x_i y_j \mathbf{P}(A_i) \mathbf{P}(B_j) &= \sum_{i=1}^k x_i \mathbf{P}(A_i) \sum_{j=1}^m y_j \mathbf{P}(B_j) = M\xi M\eta \end{aligned}$$

Общий случай можно доказать по индукции, если положить $\xi = \xi_1 \dots \xi_{n-1}$ и $\eta = \xi_n$ и воспользоваться независимостью ξ и η (см. лемму 2.1). \square

Теорема 2.4 (аддитивное свойство дисперсии). *Если случайные величины ξ_1, \dots, ξ_n независимы, то*

$$D(\xi_1 + \dots + \xi_n) = D\xi_1 + \dots + D\xi_n.$$

Доказательство. Пусть ξ и η независимы. Имеем

$$\begin{aligned} D(\xi + \eta) &= M((\xi + \eta) - M(\xi + \eta))^2 = M((\xi - M\xi) + (\eta - M\eta))^2 = \\ &= M(\xi - M\xi)^2 + M(\eta - M\eta)^2 + 2M(\xi - M\xi)(\eta - M\eta). \end{aligned}$$

Учитывая, что $\xi - M\xi$ и $\eta - M\eta$ независимы, имеем

$$M(\xi - M\xi)(\eta - M\eta) = M(\xi - M\xi)M(\eta - M\eta) = 0$$

Отсюда следует $D(\xi + \eta) = D\xi + D\eta$. Общий случай получается по индукции. \square

Пример 4. *Равномерное распределение на $\{1, 2, \dots, N\}$*

$$\begin{aligned} \mathbf{P}\{\xi = m\} &= \frac{1}{N}, \quad m = 1, 2, \dots, N \\ M\xi &= \frac{N+1}{2} \quad D\xi = \frac{N^2-1}{12}, \quad \text{так как} \\ \sum_{k=1}^n k^2 &= \frac{1}{3}(n+1)^3 - \frac{1}{2}(n+1)^2 + \frac{1}{6}(n+1). \end{aligned}$$

Пример 5. *Биномиальное распределение для числа успехов μ_n в n независимых испытаниях*

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m (1-p)^{n-m}.$$

Представим μ_n в виде суммы $\xi_1 + \dots + \xi_n$, где ξ_k индикатор успеха в k -ом испытании. Тогда

$$\begin{aligned} M\mu_n &= \sum_{k=1}^n M\xi_k = np \quad D\mu_n = \sum_{k=1}^n D\xi_k = npq, \quad \text{так как} \\ D\xi_k &= M\xi_k^2 - (M\xi_k)^2 = p - p^2 = pq. \end{aligned}$$

Пример 6. *Гипергеометрическое распределение.*

Пусть ξ - число белых шаров в выборке без возвращения объема n из урны с N шарами, из которых M шаров белого цвета. Тогда

$$\mathbf{P}\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m \leq \min\{n, M\}$$

Представим ξ в виде сумм $\xi_1 + \dots + \xi_n$, где ξ_k – индикатор события: k -м извлечен белый шар. Имеем $M\xi_k = \frac{M}{N}$. Поэтому $M\xi = \sum_{k=1}^n M\xi_k = n\frac{M}{N}$. Далее, $\xi^2 = \sum_{k=1}^n \xi_k + \sum_{k \neq l} \xi_k \xi_l$. Ясно, что $M\xi_k^2 = M\xi_k$. В предположении $k < l$ имеем:

$$M\xi_k \xi_l = MI_{\{\xi_k=1, \xi_l=1\}} = \mathbf{P}\{\xi_k = 1, \xi_l = 1\} = \mathbf{P}\{\xi_k = 1\} \mathbf{P}\{\xi_l = 1 | \xi_k = 1\} = \frac{M}{N} \frac{(M-1)}{(N-1)}$$

В результате,

$$M\xi^2 = M\xi + \sum_{k \neq l} M\xi_k \xi_l = n\frac{M}{N} + n(n-1) \frac{M(M-1)}{N(N-1)}$$

и

$$D\xi = M\xi^2 - (M\xi)^2 = n\frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}.$$

Совместное распределение случайных величин, случайные векторы

Пусть ξ, η случайные величины на (Ω, U, \mathbf{P}) , ξ принимает значения $\{x_1, \dots, x_n\}$, η – значения $\{y_1, \dots, y_m\}$. Набор чисел $p_{ij} = \mathbf{P}\{\xi = x_i, \eta = y_j\}$, $i = 1, \dots, n$ и $j = 1, \dots, m$, будем называть *совместным распределением* случайных величин ξ и η . Совместным законом распределения случайных величин ξ и η будем называть вероятность

$$\mathbf{P}\{(\xi, \eta) \in B\} = \sum_{x_i, y_j \in B} \mathbf{P}\{\xi = x_i, \eta = y_j\} = \sum_{x_i, y_j \in B} p_{ij},$$

где B любое подмножество в \mathbb{R}^2 .

Распределение вероятностей определяется таблицей

	y_1	y_2	\dots
x_1	p_{11}	p_{12}	\dots
x_2	p_{21}	p_{22}	\dots
\dots	\dots	\dots	\dots

Из двумерного закона распределения можно получить одномерный закон распределения для случайной величины ξ : $\mathbf{P}\{\xi = x_i\} = \sum_j p_{ij}$ и для случайной величины η : $\mathbf{P}\{\eta = y_j\} = \sum_i p_{ij}$.

Аналогичным образом для n случайных величин определяется n -мерный закон распределения

$$\mathbf{P}_{\xi_1, \dots, \xi_n}(B) = \mathbf{P}\{(\xi_1, \dots, \xi_n) \in B\},$$

где B – множество точек n -мерного пространства \mathbb{R}^n .

Случайные величины ξ_1, \dots, ξ_n будем трактовать как компоненты случайного вектора $\xi = (\xi_1, \dots, \xi_n)$.

Определение независимости для случайных величин

$$\mathbf{P}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \prod_{i=1}^n \mathbf{P}\{\xi_i \in B_i\}$$

обобщается на случайные вектора $\xi_i = (\xi_{i1}, \dots, \xi_{im_i})$, где $B_i \in \mathbb{R}^{m_i}$. Для таких независимых случайных векторов будет справедлива теорема.

Теорема 2.5. Если случайные вектора ξ_1, \dots, ξ_n независимы и $g_k(x)$ – числовые функции, то случайные величины $\eta_k = g_k(\xi_k)$ независимы.

Зависимые случайные величины. Коэффициент корреляции

Случайная величина ξ называется нормированной, если $M\xi = 0$ и $D\xi = 1$. Пусть ξ имеет конечное математическое ожидание и дисперсию. Случайная величина

$$\xi_1 = \frac{\xi - M\xi}{\sqrt{D\xi}}$$

с нулевым математическим ожиданием и единичной дисперсией является нормированной.

Пусть ξ и η — случайные величины и ξ_1 и η_1 — их нормировки. Коэффициент корреляции $\rho(\xi, \eta)$ между ξ и η определяется соотношением

$$\rho(\xi, \eta) = M(\xi_1 \eta_1) = \frac{M(\xi - M\xi)(\eta - M\eta)}{\sqrt{D\xi \cdot D\eta}}.$$

Числитель носит название ковариации между ξ и η и обозначается

$$Cov(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta).$$

Из неравенства Коши-Буняковского $(M\xi_1 \eta_1)^2 \leq M\xi_1^2 M\eta_1^2$ следует, что всегда $|\rho(\xi, \eta)| \leq 1$. Это же свойство можно доказать непосредственно:

$$0 \leq D(\xi_1 \pm \eta_1) = M(\xi_1 \pm \eta_1)^2 = 2 \pm 2\rho(\xi, \eta).$$

Если ξ и η независимы, то $\rho(\xi, \eta) = 0$. Обратное неверно. Пусть ξ и η — независимые случайные величины и $M\xi = M\eta = 0$. Определим $\zeta = \xi\eta$. Тогда ξ и ζ зависимы, вообще говоря. Но

$$M\xi\zeta = M\xi^2\eta = M\xi^2 M\eta = 0 = M\xi M\eta.$$

Пример 7. Пусть случайные величины ξ и η равномерно распределены на $\{1, 2, 3\}$ и независимы. Положим $\zeta = \xi\eta$. Имеем

$$\mathbf{P}\{\zeta < 3, \xi < 2\} = \mathbf{P}\{\xi = 1, \eta = 1\} + \mathbf{P}\{\xi = 1, \eta = 2\} = \frac{2}{9};$$

$$\mathbf{P}\{\zeta < 3\} = \frac{1}{3}, \quad \mathbf{P}\{\xi < 2\} = \frac{1}{3}.$$

Поэтому $\mathbf{P}\{\zeta < 3, \xi < 2\} \neq \mathbf{P}\{\zeta < 3\}\mathbf{P}\{\xi < 2\}$, т.е. ζ и ξ зависимы.

Если $Cov(\xi, \eta) = 0$, то случайные величины ξ и η называются некоррелированными.

Коэффициент корреляции равен по модулю единице тогда и только тогда, когда существуют числа $a \neq 0$ и b такие, что $\mathbf{P}\{\eta = a\xi + b\} = 1$.

Достаточность. Пусть $\mathbf{P}\{\eta = a\xi + b\} = 1$, $M\xi = \alpha$ и $D\xi = \beta^2$. Тогда

$$\rho\{\xi, \eta\} = M\left(\frac{\xi - \alpha}{\beta} \frac{a\xi + b - a\alpha - b}{|a|\beta}\right) = \text{sign}(a).$$

Необходимость. а) Пусть $\rho(\xi, \eta) = 1$. Тогда $D(\xi_1 - \eta_1) = 2(1 - \rho(\xi, \eta)) = 0$. Поэтому $\mathbf{P}\{\xi_1 - \eta_1 = \text{const}\} = 1$.

б) Пусть $\rho(\xi, \eta) = -1$. Тогда $D(\xi_1 + \eta_1) = 0$ и $\mathbf{P}\{\xi_1 + \eta_1 = \text{const}\} = 1$.

Линейная среднеквадратическая регрессия

Пусть ξ и η случайные величины и g вещественная функция на \mathbb{R} . Следующее понятие и нижеследующая теорема носят общий характер и не зависят от конечности рассматриваемой схемы.

Определение 17. Случайная величина $g(\xi)$ называется наилучшим приближением η в смысле наименьших квадратов, если $M(\eta - g(\xi))^2$ принимает наименьшее значение. Случайная величина $g(\xi)$ называется среднеквадратической регрессией случайной величины η на случайную величину ξ .

Теорема 2.6. *Линейная среднеквадратическая регрессия $g(\xi)$ величины η на величину ξ имеет вид*

$$g(\xi) = m_2 + \rho \frac{\sigma_2}{\sigma_1} (\xi - m_1).$$

Здесь $m_1 = M\xi$, $m_2 = M\eta$, $\sigma_1^2 = D\xi$, $\sigma_2^2 = D\eta$, $\rho = \rho(\xi, \eta)$ — коэффициент корреляции.

Доказательство. Найдем α и β такие что $\Phi(\alpha, \beta) = M(\eta - \alpha - \beta\xi)^2$ принимает наименьшее возможное значение. Имеем

$$\begin{aligned} \Phi(\alpha, \beta) &= M\eta^2 + \alpha^2 + \beta^2 M\xi^2 - 2\alpha M\eta - 2\beta M\xi\eta + 2\alpha\beta M\xi = \\ &= \sigma_2^2 + m_2^2 + \alpha^2 + \beta^2(\sigma_1^2 + m_1^2) - 2\alpha m_2 - 2\beta(\text{cov}(\xi, \eta) + m_1 m_2) + 2\alpha\beta m_1. \end{aligned}$$

Исследуем функцию $\Phi(\alpha, \beta)$ на экстремум:

$$\frac{\partial \Phi}{\partial \alpha} = -2(m_2 - \alpha - \beta m_1) = 0.$$

$$\frac{\partial \Phi}{\partial \beta} = 2[\beta(\sigma_1^2 + m_1^2) - (\text{cov}(\xi, \eta) + m_1 m_2) + \alpha m_1] = 0.$$

Отсюда находим, что

$$\beta = \frac{\text{cov}(\xi, \eta)}{\sigma_1^2} = \rho \frac{\sigma_2}{\sigma_1}, \quad \alpha = m_2 - \beta m_1 = m_2 - \rho \frac{\sigma_2}{\sigma_1} m_1.$$

В это же время

$$\begin{vmatrix} \frac{\partial^2 \Phi}{\partial \alpha^2} & \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} & \frac{\partial^2 \Phi}{\partial \beta^2} \end{vmatrix} = \begin{vmatrix} 2 & 2m_1 \\ 2m_1 & 2(\sigma_1^2 + m_1^2) \end{vmatrix} = 4\sigma_1^2 > 0.$$

Таким образом, при найденных α и β функция $\Phi(\alpha, \beta)$ принимает наименьшее значение. Значит,

$$g(\xi) = m_2 + \rho \frac{\sigma_2}{\sigma_1} (\xi - m_1)$$

является линейной регрессией η на ξ . \square

Коэффициент $\beta = \rho \frac{\sigma_2}{\sigma_1}$ называется *коэффициентом регрессии η на ξ* . Прямая

$$\frac{y - m_2}{\sigma_2} = \rho \frac{x - m_1}{\sigma_1}$$

называется *прямой регрессии η на ξ* . Нетрудно найти наименьшее значение $M(\eta - \alpha - \beta\xi)^2$;

$$\min_{\alpha, \beta} M(\eta - \alpha - \beta\xi)^2 = \sigma_2^2(1 - \rho^2).$$

Эта величина называется *остаточной дисперсией* величины η относительно ξ . Разность $\eta_1 = \eta - g(\xi)$ называется остатком величины η относительно величины ξ . Случайные величины η_1 и ξ являются некоррелированными. В самом деле,

$$\begin{aligned} M(\xi - m_1)(\eta - m_2 - \rho \frac{\sigma_2}{\sigma_1} (\xi - m_1)) = \\ M(\xi - m_1)(\eta - m_2) - \rho \frac{\sigma_2}{\sigma_1} M(\xi - m_1)^2 = \text{cov}(\xi, \eta) - \rho \sigma_1 \sigma_2 = 0. \end{aligned}$$

Аналогично можно рассмотреть линейную регрессию величины ξ на величину η :

$$L(\eta) = m_1 + \rho \frac{\sigma_1}{\sigma_2} (\eta - m_2).$$

Коэффициент регрессии ξ на η равен $\rho \frac{\sigma_1}{\sigma_2}$, и прямая линейной регрессии ξ на η определяется соотношением

$$\frac{y - m_2}{\sigma_2} = \frac{1}{\rho} \frac{x - m_1}{\sigma_1}$$

Остаточная дисперсия величины ξ относительно η равна $\sigma_1^2(1 - \rho^2)$.

Обе прямые регрессии проходят через точку (m_1, m_2) и совпадают только при $\rho = \pm 1$.

Неравенство Чебышева

Теорема 2.7. Для любого $x > 0$ имеет место неравенства

$$\mathbf{P}\{|\xi| \geq x\} \leq \frac{M|\xi|}{x}$$

$$\mathbf{P}\{|\xi - M\xi| \geq x\} \leq \frac{D\xi}{x^2}.$$

Доказательство. Воспользуемся неравенством

$$|\xi| = |\xi|I_{\{|\xi| \geq x\}} + |\xi|I_{\{|\xi| < x\}} \geq |\xi|I_{\{|\xi| \geq x\}} \geq xI_{\{|\xi| \geq x\}}.$$

Вычисляя математическое ожидание от обеих частей этого неравенства, получаем

$$M|\xi| \geq xMI_{\{|\xi| \geq x\}} = x\mathbf{P}\{|\xi| \geq x\}.$$

Другое неравенство получается из первого, если его применить к случайной величине $\eta = (\xi - M\xi)^2$, помня, что $M\eta = D\xi$. \square

Рассмотрим последовательность случайных величин ξ_1, ξ_2, \dots . Будем считать, что ξ_1, \dots, ξ_n определены на вероятностном пространстве $(\Omega_n, U_n, \mathbf{P}_n)$, причем при каждом $k < n$ случайные величины ξ_k имеют распределение вероятностей, не зависящее от n .

Пример 8. ξ_k - индикатор успеха k -го испытания при испытаниях в схеме Бернулли, $k \leq n$.

На самом деле, любую последовательность независимых случайных величин можно определить на одном бесконечном вероятностном пространстве.

Следующая теорема имеет общий характер.

Теорема 2.8 (Чебышев). Если ξ_1, ξ_2, \dots независимы и существует такая константа $c > 0$, что $D\xi_n \leq c$, $n = 1, 2, \dots$, тогда при любом $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P}\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n} - \frac{M\xi_1 + \dots + M\xi_n}{n}\right| < \varepsilon\right\} = 1.$$

Доказательство. Обозначим через $\zeta_n = \xi_1 + \dots + \xi_n$ и применим к $\frac{\zeta_n}{n}$ второе неравенство Чебышева

$$1 \geq \mathbf{P}\left\{\left|\frac{\zeta_n}{n} - \frac{M\zeta_n}{n}\right| < \varepsilon\right\} \geq 1 - \frac{D\zeta_n}{\varepsilon^2 n^2} \geq 1 - \frac{c}{\varepsilon^2 n},$$

так как $D\zeta_n = \sum_{k=1}^n D\xi_k \leq nc$. В пределе получим нужное соотношение. \square

Следующая теорема является следствием предыдущей.

Теорема 2.9 (закон больших чисел в форме Чебышева). Если ξ_1, ξ_2, \dots независимы и одинаково распределены, $M\xi_n = a$, $D\xi_n = \sigma^2 < \infty$, то при любом $x > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| < x \right\} = 1.$$

В частности, для схемы Бернулли верна

Теорема 2.10 (Бернулли). Пусть μ_n - число успехов при n испытаниях в схеме Бернулли с вероятностью успеха $p \in (0, 1)$ в каждом испытании. Тогда при любом $x > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \left| \frac{\mu_n}{n} - p \right| \leq x \right\} = 1.$$

Доказательство. Представим μ_n в виде суммы независимых слагаемых $\xi_1 + \dots + \xi_n$, где ξ_i - индикатор успеха при i -ом испытании. Поскольку $M\xi_i = p$ и $D\xi_i = p(1-p)$, то к μ_n применима предыдущая теорема. \square

Условное математическое ожидание

Рассмотрим случайные величины ξ и η , принимающие значения $\{x_1, \dots, x_m\}$ и $\{y_1, \dots, y_n\}$. Условным законом распределения η при заданном значении $\xi = x_k$ назовем набор условных вероятностей

$$\mathbf{P}\{\eta = y_l | \xi = x_k\} = \frac{\mathbf{P}\{\eta = y_l, \xi = x_k\}}{\mathbf{P}\{\xi = x_k\}}, \quad l = 1, \dots, n.$$

Условным математическим ожиданием η при заданном $\xi = x_k$ будет тогда сумма

$$M(\eta | \xi = x_k) = \sum_{l=1}^n y_l \mathbf{P}\{\eta = y_l | \xi = x_k\} = \frac{\sum_{l=1}^n y_l \mathbf{P}\{\eta = y_l, \xi = x_k\}}{\mathbf{P}\{\xi = x_k\}}.$$

Можем считать $M\{\eta | \xi = x_k\}$ значениями случайной величины $M(\eta | \xi)$, являющейся функцией от ξ . Случайную величину $M(\eta | \xi)$ будем называть условным математическим ожиданием η при заданной случайной величине ξ .

Математическое ожидание случайной величины $M(\eta | \xi)$ тогда равно

$$M[M(\eta | \xi)] = \sum_{k=1}^m \mathbf{P}\{\xi = x_k\} M\{\eta | \xi = x_k\}.$$

Имеет место следующее утверждение.

Теорема 2.11. Верно равенство:

$$M[M(\eta | \xi)] = M\eta.$$

Доказательство. Имеем

$$\begin{aligned} M[M(\eta | \xi)] &= \sum_{k=1}^m \mathbf{P}\{\xi = x_k\} M\{\eta | \xi = x_k\} = \\ &= \sum_{k=1}^m \sum_{l=1}^n y_l \mathbf{P}\{\eta = y_l, \xi = x_k\} = \sum_{l=1}^n y_l \mathbf{P}\{\eta = y_l\} = M\eta. \quad \square \end{aligned}$$

В результате приходим к формуле

$$M\eta = \sum_{k=1}^m \mathbf{P}\{\xi = x_k\} M\{\eta | \xi = x_k\},$$

частным случаем которой является формула полной вероятности.

Локальная предельная теорема Муавра-Лапласа

Теорема 2.12. Пусть μ - биномиально распределенная случайная величина с вероятностью успеха p , $0 < p < 1$. Если $\sigma = \sqrt{npq} \rightarrow \infty$, то для любого $c > 0$ равномерно по всем x (удовлетворяющим $|x| \leq c$ вида $x = \frac{m-np}{\sqrt{npq}}$, где m - целое неотрицательное) имеет место равенство

$$\mathbf{P}\{\mu = m\} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2}} \left(1 + O\left(\frac{1}{\sigma}\right)\right)$$

Доказательство. Пусть $m = np + x\sigma$. Имеем

$$\mathbf{P}\{\mu = m\} = \mathbf{P}\left\{\frac{\mu - np}{\sigma} = x\right\} = \frac{n!}{m!(n-m)!} p^m q^{n-m}, \quad q = 1 - p.$$

Воспользуемся формулой Стирлинга

$$n! = \sqrt{2\pi n} n^n e^{-n} e^{\theta_n}, \quad \text{где } |\theta_n| < \frac{1}{12n}.$$

Ясно, что n , $m = np + x\sqrt{npq} = np\left(1 + \frac{xq}{\sigma}\right)$ и $n - m = nq - x\sqrt{npq} = nq\left(1 - \frac{xq}{\sigma}\right)$ стремится к $+\infty$ при $n \rightarrow +\infty$. Поэтому

$$\begin{aligned} \ln \mathbf{P}\{\mu = m\} &= \ln n! - \ln m! - \ln(n-m)! + m \ln p + (n-m) \ln q = \\ &= n \ln n - m \ln m - (n-m) \ln(n-m) + m \ln p + (n-m) \ln q + \frac{1}{2} \ln \frac{n}{2\pi m(n-m)} + \theta_n - \theta_m - \theta_{n-m}. \end{aligned}$$

Так как $\ln \frac{n}{m(n-m)} = \ln \frac{1}{npq} - \ln\left(1 + \frac{xq}{\sigma}\right) - \ln\left(1 - \frac{xq}{\sigma}\right) = 2 \ln \frac{1}{\sigma} + O\left(\frac{1}{\sigma}\right)$, находим, что

$$\begin{aligned} \ln \mathbf{P}\{\mu = m\} &= -m \ln \frac{m}{np} - (n-m) \ln \frac{n-m}{nq} + \ln \frac{1}{\sigma\sqrt{2\pi}} + O\left(\frac{1}{\sigma}\right) = \\ &= -(np + x\sigma) \ln\left(1 + \frac{xq}{\sigma}\right) - (nq - x\sigma) \ln\left(1 - \frac{xq}{\sigma}\right) + \ln \frac{1}{\sigma\sqrt{2\pi}} + O\left(\frac{1}{\sigma}\right) = \\ &= \ln \frac{1}{\sigma\sqrt{2\pi}} - (np + x\sigma) \left(\frac{xq}{\sigma} - \frac{x^2 q^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right) - (nq - x\sigma) \left(-\frac{xq}{\sigma} - \frac{x^2 q^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right) + O\left(\frac{1}{\sigma}\right) = \\ &= \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right). \quad \square \end{aligned}$$

Аппроксимация в локальной теореме Муавра-Лапласа удовлетворительная, если

$$\frac{m - np}{\sqrt{npq}} = o((npq)^{\frac{1}{6}})$$

и тем хуже, чем больше $|\frac{1}{2} - p|$.

Интегральная предельная теорема Муавра-Лапласа

Теорема 2.13. При $\sigma = \sqrt{npq} \rightarrow \infty$ равномерно по $-\infty \leq a < b \leq +\infty$.

$$\mathbf{P}\left\{a \leq \frac{\mu - np}{\sqrt{npq}} \leq b\right\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \rightarrow 0.$$

Здесь μ - число успехов в схеме Бернулли при n независимых испытаниях с вероятностью успеха p в каждом испытании, $q = 1 - p$.

Доказательство. Выберем положительное $c > 0$ и пусть $|a| \leq c$ и $|b| \leq c$.

Через m_1 и m_2 обозначим соответственно $\lfloor np + a\sqrt{npq} \rfloor$ и $\lfloor np + b\sqrt{npq} \rfloor$, где $\lfloor x \rfloor$ - такое наименьшее целое, что $x \leq \lfloor x \rfloor$, а $\lceil x \rceil$ такое наибольшее целое, что $\lceil x \rceil \leq x$.

Положим $m = np + x_m\sqrt{npq}$, тогда $\Delta x_m = x_{m+1} - x_m = \frac{1}{\sqrt{npq}}$. Согласно локальной предельной теореме

$$\mathbf{P}\left\{a \leq \frac{\mu - np}{\sqrt{npq}} \leq b\right\} = \sum_{m=m_1}^{m_2} \mathbf{P}\{\mu = m\} = \sum_{m=m_1}^{m_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_m^2} \Delta x_m (1 + o(\frac{1}{\sigma})).$$

Справа в равенстве стоит интегральная сумма, сходящаяся равномерно по a и b при $\sigma \rightarrow \infty$ к интегралу $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$, если $|a| \leq c$ и $|b| \leq c$.

Так как

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1,$$

верно равенство

$$\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx = 1 - \frac{1}{\sqrt{2\pi}} \int_{|x|>c} e^{-\frac{x^2}{2}} dx.$$

В общем случае пусть $\xi_n = \frac{\mu - np}{\sqrt{npq}}$. Учитывая также, что

$$\mathbf{P}\{|\xi_n| > c\} = 1 - \mathbf{P}\{|\xi_n| \leq c\}.$$

Поэтому

$$\left| \mathbf{P}\{|\xi_n| > c\} - \frac{1}{\sqrt{2\pi}} \int_{|x|>c} e^{-\frac{x^2}{2}} dx \right| = \left| \mathbf{P}\{|\xi_n| \leq c\} - \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx \right|.$$

Выберем $\epsilon > 0$. Найдем $c > 0$ такое, что

$$\frac{1}{\sqrt{2\pi}} \int_{|x|<c} e^{-\frac{x^2}{2}} dx < \frac{\epsilon}{8}$$

и зафиксируем его. По только что доказанному найдется n_1 , что для всех $n \geq n_1$

$$\left| \mathbf{P}\{|\xi_n| \leq c\} - \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx \right| < \frac{\epsilon}{8}.$$

Поэтому при тех же $n \geq n_1$ имеем

$$\mathbf{P}\{|\xi_n| > c\} \leq \frac{\epsilon}{4}.$$

Выберем теперь любой интервал $[a, b]$ и обозначим $[A, B] = [a, b] \cap [-c, c]$. Согласно доказанному найдется n_2 такое, что для всех $n \geq n_2$ имеет место неравенство

$$\left| \mathbf{P}\{\xi_n \in [A, B]\} - \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{x^2}{2}} dx \right| < \frac{\epsilon}{2}.$$

Из неравенства

$$\left| \mathbf{P}\{\xi_n \in [a, b]\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \leq \mathbf{P}\{|\xi_n| > c\} + \frac{1}{\sqrt{2\pi}} \int_{|x|>c} e^{-\frac{x^2}{2}} dx$$

$$+ \left| \mathbf{P}\{\xi_n \in [A, B]\} - \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{x^2}{2}} dx \right|$$

получаем, что при $n \geq n_0 = \max(n_1, n_2)$

$$\left| \mathbf{P}\{\xi_n \in [a, b]\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \leq \epsilon$$

равномерно по всем $a \leq b$. \square

Неравенство Берри-Эссена

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{p^2 + q^2}{\sqrt{npq}},$$

где $F_n(x) = \mathbf{P}\left\{\frac{\mu - np}{\sqrt{npq}} \leq x\right\}$ и $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$, позволяет оценить точность приближения в интегральной теореме Муавра-Лапласа.

Без доказательства сформулируем аналог теоремы Муавра - Лапласа для полиномиальной схемы.

Пусть μ_i – число появлений i -го события в n испытаниях, $i = 1, \dots, k$, с вероятностью успеха p_i , $\mu_1 + \dots + \mu_k = n$. Заметим, что для

$$x_i = \frac{\mu_i - np_i}{\sqrt{np_i q_i}}$$

выполняется соотношение

$$\sum_{i=1}^k x_i \sqrt{np_i q_i} = 0.$$

Теорема 2.14. Если в полиномиальной схеме вероятности исходов p_i , $0 < p_i < 1$, то какова бы ни была измеримая по Жордану область G на гиперплоскости $\sum_{i=1}^k x_i \sqrt{np_i q_i} = 0$ равномерно относительно G при $n \rightarrow \infty$ имеет место соотношение

$$\mathbf{P}(G) \rightarrow \frac{\sqrt{q_1 \cdot \dots \cdot q_k}}{\sqrt{(2\pi)^{k-1} \sum_{i=1}^k p_i q_i}} \int_G \exp\left\{-\frac{1}{2} \sum_{i=1}^k q_i x_i^2\right\} dx,$$

где dx – элемент объема на G .

Предельная теорема Пуассона

Пусть μ – биномиально распределенная случайная величина,

$$\mathbf{P}\{\mu = m\} = C_n^m p^m q^{n-m}, \quad q = 1 - p.$$

Теорема 2.15 (Пуассона). Если $n \rightarrow \infty$ и $p \rightarrow 0$ так, что $np \rightarrow a$, то для любого фиксированного $m = 0, 1, \dots$ имеет место соотношение

$$\mathbf{P}\{\mu = m\} = C_n^m p^m q^{n-m} \rightarrow \frac{a^m}{m!} e^{-a}.$$

Доказательство. Действительно,

$$\mathbf{P}\{\mu = m\} = \frac{(np)^m}{m!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) (1-p)^{n-m}$$

При $n \rightarrow \infty$ и $np \rightarrow a$ предел $(1-p)^n$ равен e^{-a} . Поэтому

$$\mathbf{P}\{\mu = m\} \rightarrow \frac{a^m}{m!} e^{-a}. \quad \square$$

Можно показать, что имеет место следующая оценка

$$|\mathbf{P}\{\mu = m\} - \frac{a^m}{m!} e^{-a}| \leq np^2 = \frac{a^2}{n},$$

где $a = np$.

Применение предельных теорем

Приближение, задаваемое теоремой Пуассона, называется *пуассоновским*. Приближение с помощью теорем Муавра-Лапласа называется *нормальным*. Для распределения Пуассона

$$\mathbf{P}\{\mu = m\} = \frac{a^m}{m!} e^{-a}$$

и интеграла

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du,$$

называемого интегралом Лапласа, имеются таблицы.

Пусть, например, нам нужно вычислить вероятность $\mathbf{P}\{m_1 \leq \mu \leq m_2\}$ в схеме Бернулли в n независимых испытаниях с вероятностью успеха p .

Вычислим $x_{m_1} = \frac{m_1 - np}{\sqrt{npq}}$, $x_{m_2} = \frac{m_2 - np}{\sqrt{npq}}$ и положим

$$\mathbf{P}\{m_1 \leq \mu \leq m_2\} \approx \frac{1}{\sqrt{2\pi}} \int_{x_{m_1}}^{x_{m_2}} e^{-\frac{x^2}{2}} dx = \Phi_0(x_{m_2}) - \Phi_0(x_{m_1}).$$

При этом мы допускаем погрешность, которую можно оценить по формуле Берри-Эссена. Вместе с тем эту погрешность можно уменьшить, если в правой части приближенного равенства немного изменить пределы интегрирования, полагая

$$\mathbf{P}\{m_1 \leq \mu \leq m_2\} \approx \Phi_0(x_{m_2 + \frac{1}{2}}) - \Phi_0(x_{m_1 - \frac{1}{2}}),$$

где $x_{m_2 + \frac{1}{2}} = \frac{m_2 + 1/2 - np}{\sqrt{npq}}$ и $x_{m_1 - \frac{1}{2}} = \frac{m_1 - 1/2 - np}{\sqrt{npq}}$.

Предельная теорема Муавра-Лапласа позволяет дать приближенный ответ на следующие вопросы:

Пусть μ_n – число успехов при n испытаниях в схеме Бернулли с вероятностью успеха p , $0 < p < 1$, в каждом испытании.

I. Чему равна вероятность того, что частота наступления события отклоняется от вероятности не более чем на α ?

$$\begin{aligned} \mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| \leq \alpha\right\} &= \mathbf{P}\left\{-\alpha\sqrt{\frac{n}{pq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \alpha\sqrt{\frac{n}{pq}}\right\} \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\alpha\sqrt{\frac{n}{pq}}}^{\alpha\sqrt{\frac{n}{pq}}} \exp\left\{-\frac{x^2}{2}\right\} dx = 2\Phi_0\left(\alpha\sqrt{\frac{n}{pq}}\right). \end{aligned}$$

II. Какое наименьшее число испытаний нужно произвести для того, чтобы с вероятностью, не меньшей β , частота наступления события отклонялась от вероятности не более чем на α ?

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| \leq \alpha\right\} \approx 2\Phi_0\left(\alpha\sqrt{\frac{n}{pq}}\right) \geq \beta.$$

III. При данной вероятности β и числе испытаний n определить границу возможных значений $\left|\frac{\mu_n}{n} - p\right|$?

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| \leq \alpha\right\} \approx 2\Phi_0\left(\alpha\sqrt{\frac{n}{pq}}\right) = \beta.$$

2.2. Случайные величины (общий случай)

Определение случайной величины

Пусть $(\Omega, \mathcal{U}, \mathbf{P})$ – произвольное вероятностное пространство.

Определение 18. Функция $\xi = F(\omega)$, определенная на Ω и принимающая вещественные значения, называется случайной величиной, если для любого $x \in \mathbb{R}$ множество $\{\omega \in \Omega : \xi(\omega) \leq x\} = \{\xi \leq x\}$ является случайным событием, то есть принадлежит \mathcal{U} .

Определение 19. Функция $F(x) = F_\xi(x) = \mathbf{P}\{\xi \leq x\}$ называется функцией распределения случайной величины ξ .

По свойству непрерывности вероятности \mathbf{P} имеем $\mathbf{P}(\xi < x) = F(x - 0)$. Поэтому

- 1) $\mathbf{P}(x_1 < \xi \leq x_2) = F(x_2) - F(x_1)$;
- 2) $\mathbf{P}(\xi = x) = F(x) - F(x - 0)$;
- 3) $\mathbf{P}(x_1 \leq \xi \leq x_2) = F(x_2) - F(x_1 - 0)$;
- 4) $\mathbf{P}(x_1 < \xi < x_2) = F(x_2 - 0) - F(x_1)$;
- 5) $\mathbf{P}(x_1 \leq \xi < x_2) = F(x_2 - 0) - F(x_1 - 0)$.

Теорема 2.16. Функция $F(x)$ не убывает, $F(x)$ непрерывна справа и удовлетворяет соотношениям

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1, \quad F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0.$$

Доказательство. Ясно, что $B_n = \{x < \xi \leq x + \frac{1}{n}\} \downarrow \emptyset$. Поэтому $\mathbf{P}(B_n) = F(x + \frac{1}{n}) - F(x) \rightarrow 0$. Введем $A_n = \{n - 1 < \xi \leq n\}$. Имеем $\Omega = \sum_{-\infty}^{\infty} A_n$. Поэтому

$$1 = \lim_{N \rightarrow \infty} \sum_{-N}^N \mathbf{P}(A_n) = \lim_{N \rightarrow \infty} [F(N) - F(-N)]. \quad \square$$

Определение 20. σ -алгебра \mathcal{B} подмножеств \mathbb{R} , порожденная интервалами вида $a < x \leq b$, называется борелевской. Элементы борелевской σ -алгебры называются борелевскими множествами.

Если ξ – случайная величина, то система множеств $\{\xi^{-1}(B), B \in \mathcal{B}\}$ является σ -алгеброй, обозначается символом \mathcal{U}_ξ и называется σ -алгеброй, порожденной случайной величиной ξ . \mathcal{U}_ξ является σ -подалгеброй σ -алгебры \mathcal{U} , так как элементами \mathcal{B} являются множества, полученные не более чем счетным числом объединений, пересечений и разностей промежутков вида $a < x \leq b$.

Определение 21. Функция вероятности $\mathbf{P}_\xi(B)$, определяемая для всех $B \in \mathcal{B}$ соотношением $\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}$, называется распределением вероятностей случайной величины ξ .

Тройка $(\mathbb{R}, \mathcal{B}, \mathbf{P}_\xi)$ является новым вероятностным пространством, порожденным случайной величиной ξ .

Теорема 2.17 (теорема о продолжении функции вероятности). Пусть на алгебре \mathcal{V}_0 подмножеств Ω определена вероятность \mathbf{P} , удовлетворяющая аксиомам

$$1^\circ \mathbf{P}(A) \geq 0 \text{ для всех } A \in \mathcal{V}_0;$$

$$2^\circ \mathbf{P}(\Omega) = 1;$$

$$3^\circ \mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B), \text{ если } A \cap B = \emptyset;$$

$$4^\circ \text{ Если события } A_n \text{ в последовательности } \{A_n\} \text{ попарно несовместны и } \sum_{n=1}^{\infty} A_n \in \mathcal{V}_0, \text{ то}$$

$$\mathbf{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

Тогда вероятность \mathbf{P} можно единственным образом продолжить на минимальную σ -алгебру \mathcal{V} , порожденную \mathcal{V}_0 .

Система промежутков $\{(x_1, x_2]\}$ порождает алгебру \mathcal{B}_0 , состоящую из всевозможных конечных сумм вида $\sum_{k=1}^n (x_{1k}, x_{2k}]$. Соотношение $F(x_2) - F(x_1) = \mathbf{P}_\xi((x_1, x_2])$ определяет на алгебре \mathcal{B}_0 σ -аддитивную функцию множества, удовлетворяющую условиям $1^\circ - 4^\circ$. По теореме о продолжении вероятности введенная таким образом функция множества продолжается единственным образом на минимальную σ -алгебру, совпадающую с σ -алгеброй борелевских множеств \mathcal{B} . Поэтому вероятность $\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}$, $B \in \mathcal{B}$, совпадает с мерой, порожденной функцией распределения F_ξ и представляется интегралом Лебега-Стилтьеса

$$\mathbf{P}_\xi(B) = \int_B dF_\xi(x).$$

Пусть $F(x)$ имеет не более счетного числа точек разрыва x_1, x_2, \dots , $\mathbf{P}\{\xi = x_k\} = p_k$ и $\sum_k p_k = 1$. В этом случае ξ называется дискретной случайной величиной, говорят также, что ξ имеет дискретное распределение.

Пример 9. Биномиальное распределение

$$\mathbf{P}\{\xi = k\} = C_n^k p^k q^{n-k}, \quad k = 0, \dots, n, \quad p \in (0, 1), \quad q = 1 - p.$$

Пример 10. Пуассоновское распределение

$$\mathbf{P}\{\xi = k\} = \frac{a^k}{k!} e^{-a}, \quad a > 0, \quad k = 0, 1, \dots$$

Пример 11. Геометрическое распределение

$$\mathbf{P}\{\xi = k\} = p(1-p)^{k-1}, \quad k = 1, 2, \dots, \quad p \in (0, 1).$$

Суммируемая функция $p_\xi(x)$ называется плотность случайной величины ξ , если

$$\mathbf{P}\{x_1 < \xi \leq x_2\} = F_\xi(x_2) - F_\xi(x_1) = \int_{x_1}^{x_2} p_\xi(x) dx,$$

Ясно, что

$$1^\circ. F'_\xi(x) = p_\xi(x) \text{ в точках непрерывности } p_\xi(x);$$

$$2^\circ. F_\xi(x) = \int_{-\infty}^x p_\xi(x) dx.$$

Из свойств функции распределения следует, что

$$p_{\xi}(x) \geq 0, \quad \int_{-\infty}^{\infty} p_{\xi}(x) dx = 1.$$

Случайная величина ξ , имеющая плотность, называется абсолютно непрерывной случайной величиной, говорят также, что ξ имеет абсолютно непрерывное распределение. При этом

$$\mathbf{P}_{\xi}(B) = \int_B p_{\xi}(x) dx.$$

Пример 12. Нормальное (гауссовское) распределение.

Случайная величина ξ имеет нормальное распределение с параметрами (a, σ) , $-\infty < a < \infty$, $\sigma > 0$, если ξ абсолютно непрерывна и ее плотность имеет вид

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Пример 13. Равномерное распределение.

Случайная величина ξ равномерно распределена на отрезке $[a, b]$, если ξ абсолютно непрерывна и ее плотность имеет вид

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b]. \end{cases}$$

Пример 14. γ -распределение.

Случайная величина ξ имеет γ -распределение, если ее плотность представляется в виде

$$p_{\xi}(x) = \begin{cases} 0, & x < 0 \\ \frac{\lambda^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, & x \geq 0 \end{cases}$$

с параметрами $\alpha > 0$ и $\lambda > 0$.

В частности, при $\alpha = 1$ получаем плотность показательного распределения. Тогда случайная величина ξ называется показательно распределенной.

Если существует такое случайное событие $B \in \mathcal{B}$, имеющее нулевую лебегову меру, такое, что $\mathbf{P}_{\xi}(B) = 1$, то случайная величина ξ называется сингулярной или имеющей сингулярное распределение.

В общем случае любая функция распределения $F(x)$ представима в виде

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x),$$

где $a_i \geq 0$, $a_1 + a_2 + a_3 = 1$, $F_1(x)$ – дискретная функция распределения, $F_2(x)$ – функция распределения, имеющая плотность, $F_3(x)$ – сингулярная функция распределения.

Определение 22. Функция $g(u)$ называется борелевской функцией, если для любого $B \in \mathcal{B}$ полный прообраз $g^{-1}(B)$ принадлежит \mathcal{B} .

Теорема 2.18. Если ξ – случайная величина и $g(u)$ – борелевская функция, то $\eta = g(\xi)$ является случайной величиной.

Доказательство. Пусть $B \in \mathcal{B}$. Тогда $g^{-1}(B) = B_1 \in \mathcal{B}$. Так как $\eta^{-1}(B) = \xi^{-1}(B_1) \in \mathcal{U}$, то η – случайная величина. \square

Пример 15. Пусть $g(x) = x^2$, $F_{\xi}(x)$ – абсолютно непрерывная функция распределения с плотностью $p_{\xi}(x)$ и $\eta = \xi^2$. При $x \geq 0$ из соотношения

$$F_{\eta}(x) = \mathbf{P}\{\xi^2 \leq x\} = \mathbf{P}\{-\sqrt{x} \leq \xi \leq \sqrt{x}\} = F_{\xi}(\sqrt{x}) - F_{\xi}(-\sqrt{x})$$

получаем, что η абсолютно непрерывна и

$$p_\eta(x) = \frac{1}{2\sqrt{x}} (p_\xi(\sqrt{x}) + p_\xi(-\sqrt{x})).$$

Многомерные распределения

Пусть ξ_1, \dots, ξ_n – случайные величины. Вектор ξ со случайными компонентами ξ_1, \dots, ξ_n будем называть случайным вектором. Функция вида

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \mathbf{P}\{\xi_1 \leq x_1, \dots, \xi_n \leq x_n\}$$

называется функцией распределения случайного вектора. Сформулируем некоторые свойства функции распределения:

$$1^\circ \lim_{x_n \rightarrow +\infty} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1, \dots, \xi_{n-1}}(x_1, \dots, x_{n-1}) \text{ (согласованность).}$$

$$2^\circ \lim_{x_n \rightarrow -\infty} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = 0.$$

$$3^\circ F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \text{ не убывает и непрерывна справа по каждому аргументу.}$$

$$4^\circ F_{\xi_1, \dots, \xi_n}(+\infty, \dots, +\infty) = 1.$$

$$5^\circ \mathbf{P}\{x_1 < \xi_1 \leq x_1 + h_1, \dots, x_n < \xi_n \leq x_n + h_n\} = \Delta_{h_1, \dots, h_n} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) =$$

$$\sum_{\theta_1, \dots, \theta_n=0}^1 (-1)^{n+\theta_1+\dots+\theta_n} F_{\xi_1, \dots, \xi_n}(x_1 + \theta_1 h_1, \dots, x_n + \theta_n h_n).$$

Последнее соотношение доказывается методом математической индукции.

Рассмотрим алгебру \mathcal{B}_0^n всевозможных конечных объединений прямоугольников вида

$\prod_{i=1}^n \{a_i < \xi_i \leq b_i\}$. Наименьшая σ -алгебра, содержащая \mathcal{B}_0^n , называется σ -алгеброй \mathcal{B}^n борелевских множеств. σ -алгебра \mathcal{B}^n является замыканием алгебры \mathcal{B}_0^n относительно не более чем счетного числа операций объединения, пересечения и дополнения. Поэтому σ -алгебра $\mathcal{U}_\xi = \{\xi \in B, B \in \mathcal{B}^n\} = \{U = \xi^{-1}(B), B \in \mathcal{B}^n\}$, порожденная случайным вектором $\xi = (\xi_1, \dots, \xi_n)$ содержится в исходной σ -алгебре \mathcal{U} вероятностного пространства $(\Omega, \mathcal{U}, \mathbf{P})$, $\mathcal{U}_\xi \subset \mathcal{U}$.

Элементы алгебры \mathcal{B}_0^n представляются в виде объединения конечного числа попарно непересекающихся прямоугольников вида $\prod_{i=1}^n \{a_i < \xi_i \leq b_i\}$. Пусть $B = \sum_k B_k$, где B_k – попарно непересекающиеся прямоугольники. Определим функцию множества, положив $\mathbf{P}_\xi(B) = \sum_k \mathbf{P}\{\xi \in B_k\}$ (см. 5°). Доказывается, что так определенная функция множества \mathbf{P}_ξ не зависит от представления $B \in \mathcal{B}_0^n$ в виде объединения попарно непересекающихся прямоугольников и σ -аддитивна на алгебре \mathcal{B}_0^n . С помощью теоремы о продолжении вероятности функция \mathbf{P}_ξ продолжается единственным образом до функции вероятности на σ -алгебре \mathcal{B}^n и полностью определяется функцией распределения F_ξ . Так как $\mathcal{U}_\xi \subset \mathcal{U}$, имеет место равенство $\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}$ для всех $B \in \mathcal{B}^n$ и вероятность $\mathbf{P}_\xi(B)$ представляется интегралом Лебега-Стилтьеса

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\} = \int_B dF_\xi(x).$$

Тройка $(\mathbb{R}^n, \mathcal{B}^n, \mathbf{P}_\xi)$ является новым вероятностным пространством. Вероятностная функция \mathbf{P}_ξ называется распределением вероятностей случайного вектора $\xi = (\xi_1, \dots, \xi_n)$.

Дискретное распределение задается конечным или счетным набором точек $x^k = (x_1, \dots, x_n)$ таких, что неотрицательные числа $p_k = \mathbf{P}\{\xi = x^k\}$ удовлетворяют $\sum_k p_k = 1$. Вероятность случайного события $B \in \mathcal{B}^n$ вычисляется по формуле

$$\mathbf{P}_\xi(B) = \sum_{x^k \in B} p_k.$$

Многомерной плотностью распределения называется такая функция $p_\xi(x)$, $x \in \mathbb{R}^n$, что $\mathbf{P}_\xi(B) = \int_B p_\xi(x) dx$, $B \in \mathcal{B}^n$. Плотность распределения неотрицательна, $p_\xi(x) \geq 0$ и нормирована

условием $\int_{\mathbb{R}^n} p_\xi(x) dx = 1$. Плотность распределения p_ξ и функция распределения F_ξ связаны соотношением

$$p_\xi(x_1, \dots, x_n) = \frac{\partial^n F_\xi(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

в точках непрерывности p_ξ .

Пример 16. Распределение (распределение вероятностей) вектора (ξ_1, \dots, ξ_n) называется *невырожденным нормальным*, если случайный вектор (ξ_1, \dots, ξ_n) имеет плотность вида

$$p(x_1, \dots, x_n) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}Q(x_1, \dots, x_n)\right\},$$

где $Q(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij}x_i x_j$ – положительно определенная невырожденная форма, $|A|$ – определитель матрицы $\|a_{ij}\|$.

Пример 17. Распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется *равномерным* в области $\Omega \subset \mathbb{R}^n$, если его плотность распределения равна

$$p(x_1, \dots, x_n) = \begin{cases} |\Omega|^{-1}, & x \in \Omega, \\ 0, & x \notin \Omega. \end{cases}$$

В этом случае для любого борелевского множества $B \subset \mathbb{R}^n$ имеем

$$\mathbf{P}_\xi(B) = \frac{|B \cap \Omega|}{|\Omega|}.$$

Предложение 2.1. Пусть определитель $\det C$ матрицы C порядка k отличен от нуля и случайный вектор ξ имеет плотность $p_\xi(x)$. Тогда случайный вектор $\eta = \xi C$ имеет плотность $p_\eta(y)$ вида

$$p_\eta(y) = \frac{1}{\det C} p_\xi(yC^{-1}).$$

Доказательство. В самом деле, для любого $A \in \mathcal{B}^k$ имеем

$$\begin{aligned} \mathbf{P}\{\xi \in A\} &= \int_A p_\xi(x) dx = \int_{\{y=xC; x \in A\}} p_\xi(yC^{-1}) \frac{1}{|\det C|} dy = \\ &= \int_{\{y=xC; x \in A\}} p_\eta(y) dy = \mathbf{P}\{\eta \in \{xC; x \in A\}\}. \quad \square \end{aligned}$$

Математическое ожидание случайной величины

Пусть $F_\xi(x) = \mathbf{P}\{\xi \leq x\}$ – функция распределения случайной величины ξ на вероятностном пространстве $(\Omega, \mathcal{U}, \mathbf{P})$ и $\mathbf{P}_\xi(B)$, $B \in \mathcal{B}$, распределение вероятностей случайной величины ξ .

Определение 23. Математическим ожиданием случайной величины ξ называется число $M\xi$, равное

$$M\xi = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega),$$

если последний интеграл существует.

Пусть g борелевская функция на \mathbb{R} и $\int_{\mathbb{R}} |g(x)| \mathbf{P}_{\xi}(dx) < \infty$. Так как функции $g(x)$ $x \in \mathbb{R}$ и $g(\xi(\omega))$, $\omega \in \Omega$, равноизмеримы в пространствах с мерой $(\mathbb{R}, \mathcal{B}, \mathbf{P}_{\xi})$ и $(\Omega, \mathcal{U}, \mathbf{P})$, то есть $\mathbf{P}_{\xi}\{g \leq t\} = \mathbf{P}\{g(\xi) \leq t\}$, то имеет место равенство

$$\int_{\Omega} g(\xi(\omega)) \mathbf{P}(d\omega) = \int_{\mathbb{R}} g(x) \mathbf{P}_{\xi}(dx).$$

Мера \mathbf{P}_{ξ} порождается функцией распределения F_{ξ} , поэтому

$$Mg(\xi) = \int_{\Omega} g(\xi(\omega)) \mathbf{P}(d\omega) = \int_{\mathbb{R}} g(x) \mathbf{P}_{\xi}(dx) = \int_{\mathbb{R}} g(x) dF_{\xi}(x).$$

Последний интеграл носит название интеграла Лебега-Стилтьеса. В частности,

$$M\xi = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) = \int_{\mathbb{R}} x \mathbf{P}_{\xi}(dx) = \int_{\mathbb{R}} x dF_{\xi}(x)$$

Для дискретной случайной величины ξ и борелевской функции g имеем

$$\int_{\mathbb{R}} g(x) dF_{\xi}(x) = \sum_k g(x_k) [F_{\xi}(x_k) - F_{\xi}(x_k - 0)] = \sum_k g(x_k) \mathbf{P}\{\xi = x_k\}.$$

Для абсолютно непрерывной случайной величины ξ имеем

$$\int_{\mathbb{R}} g(x) dF_{\xi}(x) = \int_{\mathbb{R}} g(x) p_{\xi}(x) dx,$$

где p_{ξ} – плотность распределения случайной величины ξ .

Из определения интеграла Лебега следует, что математическое ожидание $M\xi$ конечно тогда и только тогда, когда $M|\xi| < +\infty$ ($M\xi$ может существовать, но не быть конечным).

Таким образом, если ξ – дискретная случайная величина, то $M\xi = \sum_k x_k \mathbf{P}\{\xi = x_k\}$. Если F_{ξ} имеет плотность, то $M\xi = \int_{\mathbb{R}} x p_{\xi}(x) dx$.

Свойства математического ожидания:

1. Если $\mathbf{P}\{\xi = c\} = 1$, то $M\xi = c$;
2. $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$, если существуют два из трех математических ожиданий;
3. Если $\xi_1 \leq \xi_2$, то $M\xi_1 \leq M\xi_2$, в частности, $|M\xi_1| \leq M|\xi_2|$;
4. Если I_A – индикатор события A , то $MI_A = \mathbf{P}(A)$;
5. Если $\xi \geq 0$ и $M\xi = 0$, то $\xi = 0$ с вероятностью 1;

Доказательство. Ясно, что $\{\xi > 0\} = \bigcup_n \{\xi \geq \frac{1}{n}\}$ и $\{\xi \geq \frac{1}{n}\} \subset \{\xi \geq \frac{1}{n+1}\}$. Поэтому

$$0 = M\xi = M\xi I_{\{\xi=0\}} + M\xi I_{\{\xi>0\}} = M\xi I_{\{\xi>0\}} \geq M\xi I_{\{\xi \geq \frac{1}{n}\}} \geq \frac{1}{n} M I_{\{\xi \geq \frac{1}{n}\}} = \frac{1}{n} \mathbf{P}\{\xi \geq \frac{1}{n}\}.$$

Из свойства непрерывности вероятности получаем, что $\mathbf{P}\{\xi > 0\} = 0$. \square

6. Для дисперсии $D\xi = M(\xi - M\xi)^2$ имеет место равенство $D\xi = M\xi^2 - (M\xi)^2$.

Пример 18. Пусть ξ – нормально распределенная случайная величина с параметрами (a, σ) . Тогда

$$M\xi = \int_{\mathbb{R}} (t - a) \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t - a)^2}{2\sigma^2}\right\} dt + a \int_{\mathbb{R}} \exp\left\{-\frac{(t - a)^2}{2\sigma^2}\right\} dt = a,$$

и

$$D\xi = \sigma.$$

Если ξ нормально распределена с параметрами $(0, 1)$, то $\eta = \sigma\xi + a$ нормально распределена с параметрами (a, σ) .

Пример 19. Пусть ξ – случайная величина, распределенная по закону Пуассона с параметром μ . Тогда

$$M\xi = \sum_n n \frac{\mu^n}{n!} e^{-\mu} = \mu$$

и

$$D\xi = \mu.$$

Пример 20. Пусть случайная величина ξ равномерно распределена на отрезке $[a, b]$. Тогда

$$M\xi = \frac{1}{a-b} \int_a^b x dx = \frac{a+b}{2} \text{ и } D\xi = \frac{(b-a)^2}{12}.$$

7. Если $0 \leq \xi_n \nearrow \xi$, то $M\xi_n \rightarrow M\xi$ (теорема о монотонной сходимости).

Следствие. Если ряд $\sum_{n=1}^{\infty} \xi_n$ состоит из неотрицательных случайных величин, то

$$M\left(\sum_n \xi_n\right) = \sum_n M\xi_n.$$

8. Если $\lim_{n \rightarrow \infty} \xi_n = \xi$ и для любого $\omega \in \Omega$ выполняется неравенство $|\xi_n(\omega)| \leq \eta(\omega)$, где $M\eta \leq +\infty$, то $\lim_{n \rightarrow \infty} M\xi_n = M\xi$ (теорема Лебега о мажорируемой сходимости).

9. Если случайная величина ξ имеет плотность $p_\xi(x)$, $g(x)$ – борелевская функция и

$$\int_{\mathbb{R}} |g(x)| p_\xi(x) dx < +\infty,$$

то

$$Mg(\xi) = \int_{\mathbb{R}} g(x) p_\xi(x) dx.$$

Независимые случайные величины

Одним из важнейших понятий теории вероятностей является понятие независимости случайных величин.

Определение 24. Случайные величины ξ_1, \dots, ξ_n называются *независимыми*, если

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n).$$

Определение 25. Случайные величины последовательности $\{\xi_n\}_{n=1}^{\infty}$, заданные на вероятностном пространстве $(\Omega, \mathcal{U}, \mathbf{P})$, называются независимыми, если для любого числа k

$$F_{\xi_1, \dots, \xi_k}(x_1, \dots, x_k) = F_{\xi_1}(x_1) \dots F_{\xi_k}(x_k).$$

Отметим также, что для последовательности независимых случайных величин будут независимыми любые два события, относящиеся к непересекающимся группам случайных величин этой последовательности.

Теорема 2.19. Пусть ξ_1, \dots, ξ_n – независимые случайные величины. Если $F_{\xi_i}(x)$ есть функция распределения абсолютно непрерывной случайной величины ξ_i с плотностью $p_{\xi_i}(x)$, $i = 1, \dots, n$, то $F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ имеет плотность $p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, равную $p_{\xi_1}(x_1) \cdots p_{\xi_n}(x_n)$. Обратно, если $F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ абсолютно непрерывна, то и $F_{\xi_i}(x)$ тоже абсолютно непрерывны для всех i . При этом

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = p_{\xi_1}(x_1) \cdots p_{\xi_n}(x_n)$$

почти всюду.

Доказательство. Если функция распределения каждой случайной величины абсолютно непрерывна и

$$F_{\xi_i}(x_i) = \int_{-\infty}^{x_i} p_{\xi_i}(t_i) dt_i,$$

то совместная функция распределения имеет вид

$$\begin{aligned} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) &= F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n) = \\ &= \int_{-\infty}^{x_1} p_{\xi_1}(t_1) dt_1 \cdots \int_{-\infty}^{x_n} p_{\xi_n}(t_n) dt_n = \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1}(t_1) \cdots p_{\xi_n}(t_n) dt_1 \cdots dt_n. \end{aligned}$$

Обратно, предположим, что

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) dt_1 \cdots dt_n,$$

Тогда

$$\begin{aligned} F_{\xi_1}(x_1) &= F_{\xi_1, \dots, \xi_n}(x_1, \infty, \dots, \infty) = \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) dt_1 \cdots dt_n, \end{aligned}$$

так что

$$p_{\xi_1}(t_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{\xi_1, \dots, \xi_n}(t_1, t_2, \dots, t_n) dt_2 \cdots dt_n.$$

Равенства имеет место почти всюду, поэтому функции $p_{\xi_i}(x_i)$ определены почти всюду. \square

Дадим другое определение независимости случайных величин.

Определение 26. Случайные величины ξ_1, \dots, ξ_n называется независимыми, если для любых борелевских $B_i \in \mathcal{B}$, $i = 1, \dots, n$, имеет место равенство

$$\mathbf{P}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \mathbf{P}\{\xi_1 \in B_1\} \cdots \mathbf{P}\{\xi_n \in B_n\}.$$

Приведенные определения независимости случайных величин эквивалентны. Для того, чтобы доказать эквивалентность воспользуемся следующей теоремой теории меры.

Определение 27. Алгебры событий \mathcal{A}_1 и \mathcal{A}_2 называются независимыми, если для любых событий $A_1 \in \mathcal{A}_1$ и $A_2 \in \mathcal{A}_2$ имеет место равенство

$$\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2).$$

Теорема 2.20. σ -алгебры \mathcal{U}_1 и \mathcal{U}_2 , порожденные соответственно независимыми алгебрами событий \mathcal{A}_1 и \mathcal{A}_2 независимы.

Теперь мы можем сформулировать утверждение, доказывающую эквивалентность определений 25 и 26.

Теорема 2.21. Случайные величины ξ_1 и ξ_2 независимы тогда и только тогда, когда независимы σ -алгебры \mathcal{U}_{ξ_1} и \mathcal{U}_{ξ_2} , порожденные этими случайными величинами.

Доказательство. Пусть ξ_1 и ξ_2 независимы и Δ и Λ – полуинтервалы вида $(a_1, a_2]$ и $(b_1, b_2]$. Имеет место цепочка равенств

$$\begin{aligned} \mathbf{P}\{\xi_1 \in \Delta, \xi_2 \in \Lambda\} &= \mathbf{P}\{\xi_1 \in (a_1, a_2], \xi_2 \in (b_1, b_2]\} = \\ &= F_{\xi_1, \xi_2}(a_2, b_2) - F_{\xi_1, \xi_2}(a_1, b_2) - F_{\xi_1, \xi_2}(a_2, b_1) + F_{\xi_1, \xi_2}(a_1, b_1) = \\ &= (F_{\xi_1}(a_2) - F_{\xi_1}(a_1))(F_{\xi_2}(b_2) - F_{\xi_2}(b_1)) = \mathbf{P}\{\xi \in \Delta\}\mathbf{P}\{\xi_2 \in \Lambda\}. \end{aligned}$$

Далее, если $\Delta_i, i = 1, \dots, n$, и $\Lambda_j, j = 1, \dots, m$ – две системы непересекающихся интервалов, то

$$\begin{aligned} \mathbf{P}\{\xi_1 \in \cup_{i=1}^n \Delta_i, \xi_2 \in \cup_{j=1}^m \Lambda_j\} &= \\ &= \sum_{i,j} \mathbf{P}\{\xi_1 \in \Delta_i, \xi_2 \in \Lambda_j\} = \sum_{i,j} \mathbf{P}\{\xi_1 \in \Delta_i\}\mathbf{P}\{\xi_2 \in \Lambda_j\} = \\ &= \mathbf{P}\{\xi_1 \in \cup_{i=1}^n \Delta_i\}\mathbf{P}\{\xi_2 \in \cup_{j=1}^m \Lambda_j\}. \end{aligned}$$

Таким образом, алгебры, порожденные полуинтервалами Δ_i и Λ_j независимы. Из теоремы 2.20 следует независимость \mathcal{U}_{ξ_1} и \mathcal{U}_{ξ_2} .

Обратное утверждение следует из равенств

$$\begin{aligned} F_{\xi_1, \xi_2}(x_1, x_2) &= \mathbf{P}\{\xi_1 \leq x_1, \xi_2 \leq x_2\} = \\ &= \mathbf{P}\{\xi_1 \leq x_1\}\mathbf{P}\{\xi_2 \leq x_2\} = F_{\xi_1}(x_1)F_{\xi_2}(x_2). \quad \square \end{aligned}$$

Эта теорема справедлива для любого числа случайных величин.

Следствие. Пусть φ_1, φ_2 – борелевские функции и ξ_1 и ξ_2 – независимые случайные величины. Тогда $\eta_1 = \varphi_1(\xi_1)$ и $\eta_2 = \varphi_2(\xi_2)$ – также независимые случайные величины.

Доказательство. Очевидно, что $\mathcal{U}_{\eta_1} \subset \mathcal{U}_{\xi_1}$ и $\mathcal{U}_{\eta_2} \subset \mathcal{U}_{\xi_2}$. Поэтому из независимости σ -алгебр \mathcal{U}_{ξ_1} и \mathcal{U}_{ξ_2} следует независимость \mathcal{U}_{η_1} и \mathcal{U}_{η_2} . \square

Замечание. Если $\xi = (\xi_1, \dots, \xi_n)$, $\eta = (\eta_1, \dots, \eta_m)$ – независимые случайные векторы, $\varphi_1(x_1, \dots, x_n)$ и $\varphi_2(y_1, \dots, y_m)$ – борелевские функции, то случайные величины $\varphi_1(\xi_1, \dots, \xi_n)$, $\varphi_2(\eta_1, \dots, \eta_m)$ независимы.

Определения и теоремы о независимых случайных величинах остаются в силе и для случайных векторов.

Замечание. Из теоремы 2.21 следует, что если компоненты ξ_1 и ξ_2 случайного вектора $\xi = (\xi_1, \xi_2)$ независимы, то для борелевских множеств $B = B_1 \times B_2 = \{(x_1, x_2) : x_1 \in B_1, x_2 \in B_2\}$, $B_1, B_2 \in \mathcal{B}$, выполняется равенство

$$\mathbf{P}\{\xi \in B\} = \mathbf{P}\{\xi_1 \in B_1, \xi_2 \in B_2\} = \mathbf{P}\{\xi_1 \in B_1\}\mathbf{P}\{\xi_2 \in B_2\}.$$

Говорят тогда, что мера \mathbf{P}_ξ равна произведению мер \mathbf{P}_{ξ_1} и \mathbf{P}_{ξ_2} . Интеграл

$$\int_B g(x_1, x_2) \mathbf{P}_{\xi_1, \xi_2}(dx_1, dx_2)$$

в этом случае может быть записан в виде повторного интеграла:

$$\iint_{B_1 \times B_2} g(x_1, x_2) \mathbf{P}_{\xi_1}(dx_1) \mathbf{P}_{\xi_2}(dx_2).$$

Если $g(x_1, x_2) = g_1(x_1)g_2(x_2)$, тогда

$$\int_B g(x_1, x_2) \mathbf{P}_{\xi_1, \xi_2}(dx_1, dx_2) = \int_{B_1} g_1(x_1) \mathbf{P}_{\xi_1}(dx_1) \cdot \int_{B_2} g_2(x_2) \mathbf{P}_{\xi_2}(dx_2)$$

при условии, что интегралы сходятся.

Теорема 2.22 (формула композиции). Пусть ξ и η – независимые абсолютно непрерывные случайные величины, p_ξ и p_η – их плотности. Тогда

$$p_{\xi+\eta}(x) = \int_{\mathbb{R}} p_\xi(x-y)p_\eta(y)dy = \int_{\mathbb{R}} p_\xi(y)p_\eta(x-y)dy.$$

Доказательство. Плотность $p_{\xi,\eta}(x,y)$ совместного распределения случайного вектора с компонентами ξ и η равна произведению плотностей $p_\xi(x)$ и $p_\eta(y)$. Тогда

$$\begin{aligned} F_{\xi+\eta}(z) &= \mathbf{P}\{\xi + \eta \leq z\} = \int_{x+y \leq z} p_\xi(x)p_\eta(y)dx dy = \text{(по теореме Фубини)} \\ &= \int_{-\infty}^{\infty} p_\xi(x) \left(\int_{-\infty}^{z-x} p_\eta(y)dy \right) dx = \int_{-\infty}^{\infty} F_\eta(z-x)p_\xi(x)dx = \\ &= \int_{-\infty}^{\infty} p_\xi(x) \left(\int_{-\infty}^z p_\eta(y-x)dy \right) dx = \int_{-\infty}^z dz \int_{-\infty}^{\infty} p_\xi(x)p_\eta(y-x)dx \end{aligned}$$

Таким образом,

$$F_{\xi+\eta}(z) = \int_{-\infty}^{\infty} F_\eta(z-x)p_\xi(x)dx$$

и

$$p_{\xi+\eta}(z) = \int_{-\infty}^{\infty} p_\eta(z-x)p_\xi(x)dx. \quad \square$$

Пример 21. Пусть ξ_1 и ξ_2 независимы и равномерно распределены на отрезке $[0, 1]$. Тогда

$$p_{\xi_1}(x) = p_{\xi_2}(x) = \begin{cases} 1, & x \in [0, 1], \\ 0, & x \notin [0, 1] \end{cases}$$

и

$$p_{\xi_1+\xi_2}(z) = \int_0^1 p_{\xi_1}(x-t)dt = \begin{cases} 0, & x \notin [0, 2], \\ x, & x \in [0, 1], \\ 2-x, & x \in [1, 2]. \end{cases}$$

Мультипликативное свойство

Докажем здесь мультипликативное свойство математического ожидания.

Теорема 2.23. Если случайные величины ξ и η независимы и имеют конечное математическое ожидание $M\xi$ и $M\eta$, то

$$M\xi\eta = M\xi \cdot M\eta.$$

Приведем набросок доказательства. Случайную величину ξ будем называть простой, если она представима в виде

$$\xi = \sum_{k=1}^m x_k I_{A_k},$$

где A_k – попарно несовместные случайные события.

1. Если $\xi \geq 0$, то существует такая последовательность простых неотрицательных случайных величин ξ_n , $\xi_n \geq 0$, что $\xi_n \nearrow \xi$ почти наверное. В качестве ξ_n можно рассмотреть

$$\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\{\frac{k-1}{2^n} < \xi \leq \frac{k}{2^n}\}} + n I_{\{\xi > n\}}.$$

Ясно, что $0 \leq \xi_n \leq \xi_{n+1} \leq \xi$ и при $\xi_n(\omega) \leq n$ имеет место неравенство $\xi(\omega) \leq \xi_n(\omega) + \frac{1}{2^n}$.

2. Пусть ξ и η независимые простые случайные величины, принимающие значения $x_1 < \dots < x_m$ и $y_1 < \dots < y_n$. Тогда

$$\xi = \sum_{k=1}^m x_k I_{\{\xi=x_k\}}, \quad \eta = \sum_{k=1}^n y_k I_{\{\eta=y_k\}}$$

и

$$\mathbf{P}\{\xi = x_k, \eta = y_l\} = \mathbf{P}\{\xi = x_k\} \cdot \mathbf{P}\{\eta = y_l\}.$$

В этом случае, как нам известно, $M\xi\eta = M\xi \cdot M\eta$ (см. теорему 2.3).

3. Напомним, что если ξ и η независимы и g_1 и g_2 – борелевские функции, то случайные величины $g_1(\xi)$ и $g_2(\eta)$ также независимы.

В обозначениях пункта 1 имеем

$$M(\xi\eta) = \lim_{n \rightarrow \infty} M(\xi_n \eta_n) = \lim_{n \rightarrow \infty} M\xi_n \cdot \lim_{n \rightarrow \infty} M\eta_n = M\xi \cdot M\eta.$$

4. В общем случае $\xi = \xi^+ - \xi^-$ и $\eta = \eta^+ - \eta^-$, где ξ^\pm и η^\pm – независимые как функции ξ и η и неотрицательные случайные величины. Поэтому

$$M(\xi\eta) = M(\xi^+ - \xi^-)(\eta^+ - \eta^-) = M(\xi^+ - \xi^-) \cdot M(\eta^+ - \eta^-) = M\xi \cdot M\eta. \quad \square$$

Следствие. Если случайные величины ξ_1, \dots, ξ_n независимы и имеют конечное математическое ожидание, то

$$M(\xi_1 \cdots \xi_n) = M\xi_1 \cdots M\xi_n.$$

Условное математическое ожидание

Пусть $(\Omega, \mathcal{U}, \mathbf{P})$ – вероятностное пространство, $B \subset \mathcal{U}$ и $\mathbf{P}(B) > 0$. Рассмотрим $(\Omega, \mathcal{U}, \mathbf{P}_B)$, где $\mathbf{P}_B(A) = \mathbf{P}(A|B)$. Пусть ξ случайная величина на $(\Omega, \mathcal{U}, \mathbf{P})$, а значит и на $(\Omega, \mathcal{U}, \mathbf{P}_B)$.

Определение 28. Условным математическим ожиданием случайной величины ξ относительно случайного события B , называется число, равное

$$M_B \xi = M(\xi|B) = \int_{\Omega} \xi(\omega) \mathbf{P}_B(d\omega).$$

Имеем

$$\begin{aligned} M(\xi|B) &= \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega|B) = \frac{1}{\mathbf{P}(B)} \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega \cap B) = \\ &= \frac{1}{\mathbf{P}(B)} \int_B \xi(\omega) \mathbf{P}(d\omega) = \frac{M(\xi; B)}{\mathbf{P}(B)}, \end{aligned}$$

где

$$M(\xi; B) = \int_B \xi(\omega) \mathbf{P}(d\omega).$$

Пусть $\{B_n\}$ – разбиение Ω . Тогда

$$M\xi = \sum_n \int_{B_n} \xi(\omega) \mathbf{P}(d\omega) = \sum_n M(\xi; B_n) = \sum_n \mathbf{P}(B_n) M(\xi|B_n). \quad (2.1)$$

Эта формула является обобщением формулы полной вероятности для математического ожидания. В самом деле, если $\xi = I_A$ – индикатор события A , то $M\xi = MI_A = \mathbf{P}(A)$ и

$$M(\xi|B_n) = M(I_A|B_n) = \frac{1}{\mathbf{P}(B_n)} \int_{B_n} I_A(\omega) \mathbf{P}(d\omega) = \frac{1}{\mathbf{P}(B_n)} \int_{AB_n} \mathbf{P}(d\omega) = \frac{\mathbf{P}(AB_n)}{\mathbf{P}(B_n)} = \mathbf{P}(A|B_n).$$

Заменяя в (2.1) ξ на I_A и $M(\xi|B_n)$ на $\mathbf{P}(A|B_n)$, получим формулу полной вероятности.

Пример 22. Рабочий обслуживает n одноципных станков, расположенных прямолинейно на расстоянии 1 друг от друга. Рабочий обслуживает их подходу к ним в порядке очередности. Найти средний переход между станками.

Через B_k обозначим случайное событие – рабочий находится у станка с номером k . Тогда $\mathbf{P}(B_k) = 1/n$. Вероятность того, что от станка k рабочий направляется к станку i , $i = 1, \dots, n$, также равна $1/n$. Длина перехода при этом составляет $|i - k|$ и

$$M(\xi|B_k) = \sum_{i=1}^{k-1} \frac{1}{n}(k-i) + \sum_{i=k+1}^n \frac{1}{n}(i-k) = \frac{k(k-1)}{2n} + \frac{(n-k)(n-k+1)}{2n}$$

По формуле (2.1)

$$M\xi = \sum_{k=1}^n \mathbf{P}(B_k) M(\xi|B_k) = \sum_{k=1}^n \frac{1}{n^2} k(k-1).$$

Так как

$$\sum_{k=1}^n k(k-1) = \frac{1}{3}(n-1)n(n+1),$$

то

$$M\xi = \frac{n-1}{3} \left(1 + \frac{1}{n}\right).$$

Таким образом, если обслуживается n станков, то средняя длина перехода рабочего составляет приблизительно $\frac{n-1}{3}$.

Перейдем к рассмотрению абсолютно непрерывного вектора (ξ, η) . Через Δ_y будем обозначать интервал, содержащий y . Условная функция распределения случайной величины ξ при условии $\eta = y$ определяется как предел

$$\lim_{|\Delta_y| \rightarrow 0} \mathbf{P}\{\xi \leq x | \eta \in \Delta_y\} = F_{\xi}(x | \eta = y).$$

Если случайный вектор (ξ, η) имеет плотность $p_{\xi\eta}(u, v)$, то

$$F_{\xi}(x|\eta = y) = \lim_{|\Delta_y| \rightarrow 0} \frac{\frac{1}{|\Delta_y|} \int_{\Delta_y} \int_{-\infty}^x p_{\xi\eta}(u, v) dudv}{\frac{1}{|\Delta_y|} \int_{\Delta_y} \int_{-\infty}^{\infty} p_{\xi\eta}(u, v) dudv} = \frac{\int_{-\infty}^x p_{\xi\eta}(u, y) du}{\int_{-\infty}^{\infty} p_{\xi\eta}(u, y) du}.$$

Условная плотность распределения $p_{\xi}(x|\eta = y)$ находится дифференцированием

$$p_{\xi}(x|\eta = y) = \frac{d}{dy} F_{\xi}(x|\eta = y) = \frac{p_{\xi\eta}(x, y)}{\int_{-\infty}^{\infty} p_{\xi\eta}(u, y) du}. \quad (2.2)$$

Условное математическое ожидание ξ при заданном значении $\eta = y$ вычисляется по формуле

$$M(\xi|\eta = y) = \int_{-\infty}^{\infty} xp_{\xi}(x|\eta = y) dx$$

и может быть рассмотрено как значение случайной величины $M(\xi|\eta)$, принимаемое на элементарных событиях $\omega \in \{\eta = y\}$.

Так как

$$p_{\eta}(y) = \int_{-\infty}^{\infty} p_{\xi\eta}(u, y) du$$

из (2.2) находим, что

$$p_{\xi\eta}(x, y) = p_{\eta}(y)p_{\xi}(x|\eta = y).$$

Отсюда нетрудно получить, что при любых a и b

$$\mathbf{P}\{a \leq \xi \leq b\} = \int_{-\infty}^{\infty} p_{\eta}(y) \left(\int_a^b p_{\xi}(x|\eta = y) dx \right) dy.$$

Как и для случайных величин, определенных на конечном вероятностном пространстве, имеет место равенство $M(M(\xi|\eta)) = M\xi$. Действительно,

$$M(M(\xi|\eta)) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} xp_{\xi}(x|\eta = y) dx \right) p_{\eta}(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp_{\xi\eta}(x, y) dy dx = \int_{-\infty}^{\infty} xp_{\xi}(x) dx = M\xi.$$

Поэтому имеет место равенство

$$M\xi = \int_{-\infty}^{\infty} p_{\eta}(y) M(\xi|\eta = y) dy, \quad (2.3)$$

которое как и аналогичное равенство для случайных величин в конечном вероятностном пространстве может рассматриваться как обобщение формулы полной вероятности.

2.3. Характеристические функции

Определение и простейшие свойства характеристических функций

Пусть $\zeta = \xi + i\eta$, где ξ и η действительные случайные величины, у которой существуют и конечны $M\xi$ и $M\eta$. Математическое ожидание комплексной случайной величины ζ определим соотношением

$$M\zeta = M\xi + iM\eta.$$

Основные свойства математического ожидания, включая свойства аддитивности и мультипликативности остаются в силе. Докажем неравенство

$$|M\zeta| \leq M|\zeta|.$$

Существует действительное α такое, что

$$|M\zeta| = M(e^{i\alpha}\zeta).$$

Пусть $e^{i\alpha}\zeta = \xi_1 + i\eta_1$, где ξ_1 и η_1 – действительные случайные величины. Имеем

$$|M\zeta| = M\xi_1 \leq M|\xi_1| \leq M|e^{i\alpha}\zeta| = M|\zeta|.$$

Определение 29. Характеристической функцией случайной величины ξ называется функция $f_\xi(t)$ действительного переменного t , равная

$$f_\xi(t) = Me^{it\xi} = M \cos(\xi t) + iM \sin(\xi t) = \int_{-\infty}^{\infty} e^{itx} dF_\xi(x).$$

Если случайная величина ξ дискретна, то

$$f_\xi(t) = \sum_k e^{itx_k} \mathbf{P}\{\xi = x_k\}.$$

Если случайная величина абсолютно непрерывна, то

$$f_\xi(t) = \int_{-\infty}^{\infty} e^{itx} p_\xi(x) dx,$$

то есть характеристическая функция $f_\xi(t)$ в этом случае совпадает с преобразованием Фурье плотности $p_\xi(x)$ случайной величины ξ .

Некоторые свойства характеристических функций:

1. $|f_\xi(t)| \leq 1$ для всех $t \in \mathbb{R}$ и $f_\xi(0) = 1$.
2. $f_\xi(t)$ равномерно непрерывна по t .

Предварительно докажем лемму.

Лемма 2.2. При действительных x и любом целом $n \geq 1$ имеет место неравенство

$$\left| e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} \right| \leq \frac{|x|^n}{n!}$$

Доказательство. Доказательство по индукции. Имеем

$$|e^{ix} - 1| = \left| \int_0^x e^{iu} du \right| \leq \int_0^{|x|} du \leq |x|.$$

Предположим, что неравенство верно при $m = n$. Докажем его справедливость при $m = n + 1$. Воспользуемся соотношением

$$\int_0^x \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} \right) du = \frac{1}{i} \left(e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right).$$

Поэтому

$$\left| e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right| = \left| \int_0^x \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} \right) du \right| \leq \int_0^{|x|} \frac{u^n}{n!} du = \frac{|x|^{n+1}}{(n+1)!}. \quad \square$$

Теперь нетрудно доказать свойство 2. Положим $A = \{|\xi| \leq x\}$. Тогда

$$\begin{aligned} |f(t+h) - f(t)| &= |M e^{it\xi} (e^{ih\xi} - 1)| \leq M |e^{ih\xi} - 1| I_A + M |e^{ih\xi} - 1| I_{\bar{A}} \leq \\ &\leq 2M I_{\bar{A}} + |h| M |\xi| I_A \leq x|h| + 2\mathbf{P}\{|\xi| > x\}. \end{aligned}$$

Фиксируем $\varepsilon > 0$ и выберем x так, чтобы $\mathbf{P}\{|\xi| > x\} < \frac{\varepsilon}{4}$. Положим теперь $\delta = \frac{\varepsilon}{2x}$. Тогда для всех h , для которых $|h| < \delta$ выполняется неравенство $|f(t+h) - f(t)| < \varepsilon$. \square

3. Если $\eta = a\xi + b$, где a и b действительные константы, то $f_\eta(t) = e^{itb} f_\xi(at)$.

Определение 30. Комплекснозначные случайные величины $\xi = \xi_1 + i\xi_2$ и $\eta = \eta_1 + i\eta_2$ независимы, если независимы σ -алгебры $\mathcal{A}_{\xi_1, \xi_2}$ и $\mathcal{A}_{\eta_1, \eta_2}$. Иначе, для любых борелевских $B_1, B_2 \subset \mathbb{R}^2$ выполняется равенство

$$\mathbf{P}\{(\xi_1, \xi_2) \in B_1, (\eta_1, \eta_2) \in B_2\} = \mathbf{P}\{(\xi_1, \xi_2) \in B_1\} \mathbf{P}\{(\eta_1, \eta_2) \in B_2\}$$

Данное определение независимости случайных величин легко обобщается на случай конечного числа комплекснозначных случайных величин.

4. Если ξ_1, \dots, ξ_n независимы, то

$$f_{\xi_1 + \dots + \xi_n}(t) = \prod_{k=1}^n f_{\xi_k}(t).$$

Доказывается с помощью мультипликативного свойства математического ожидания.

5. Обозначим через $m_n = M\xi^n$. Если m_n конечно, то существуют все производные $f_\xi^{(k)}(t)$ порядка $k \leq n$ и

$$f_\xi^{(k)}(0) = i^k M\xi^k. \quad (2.4)$$

Кроме того, имеет место разложение

$$f_\xi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} m_k + R_n(t), \quad (2.5)$$

где $R_n(t) = o(|t|^n)$ при $t \rightarrow 0$.

Доказательство. Продифференцировав формально характеристическую функцию $f_\xi(t)$ k раз получим

$$f_\xi^{(k)}(t) = i^k M\xi^k e^{it\xi} = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF_\xi(x). \quad (2.6)$$

Обосновать законность дифференцирования под знаком интеграла можно, рассуждая по индукции. Пусть соотношение (2.6) имеет место при $k < n$. Поскольку

$$\frac{f_\xi^{(k)}(t+h) - f_\xi^{(k)}(t)}{h} = i^k M\xi^k \frac{e^{it\xi}(e^{ih\xi} - 1)}{h}, \quad (2.7)$$

$$\left| \xi^k e^{it\xi} \frac{(e^{ih\xi} - 1)}{h} \right| \leq |\xi|^{k+1} \text{ и } M|\xi^{k+1}| < \infty,$$

то в правой части (2.7) по теореме Лебега о мажорируемой сходимости можно перейти к пределу при $h \rightarrow 0$. Полагая в (2.6) $t = 0$, приходим к (2.4). Разложение (2.5) следует из теоремы Тейлора. \square

6. Имеет место равенство $f_\xi(-t) = \overline{f_\xi(t)}$.

7. Если $\xi \geq 0$, то $f_\xi(\lambda)$ определена при $\text{Im}\lambda \geq 0$. При этом $|f_\xi(\lambda)| \leq 1$, $f_\xi(\lambda)$, регулярна в $\text{Im}\lambda > 0$ и непрерывна, включая границу $\text{Im}\lambda = 0$,

$$f_\xi(\lambda) = \int_0^\infty e^{i\lambda x} dF_\xi(x).$$

Непрерывность доказывается как в свойстве 2. Регулярность следует из возможности дифференцирования под знаком интеграла.

Дискретную случайную величину ξ , принимающую только целые неотрицательные значения, будем называть *целочисленной* случайной величиной. Закон распределения целочисленной случайной величины определяется вероятностями

$$p_n = \mathbf{P}\{\xi = n\}, \quad n = 0, 1, \dots,$$

для которой $\sum_{n=0}^\infty p_n = 1$.

Определение 31. *Вероятностной производящей функцией* называется сумма ряда

$$\varphi_\xi(s) = \sum_{n=0}^\infty p_n s^n, \quad (2.8)$$

равная $M s^\xi$. Ряд в (2.8) абсолютно сходится при $|s| \leq 1$.

8. Если $\varphi_\xi(s) = M s^\xi$ – производящая функция целочисленной случайной величины, то

$$f_\xi(t) = \varphi_\xi(e^{it}).$$

Здесь же укажем еще одно свойство производящих функций Пусть ξ_1, ξ_2, \dots – последовательность целочисленных независимых одинаково распределенных случайных величин с производящей функцией $\varphi_\xi(s)$ и ν – независимая от них целочисленная случайная величина с производящей функцией $\varphi_\nu(s)$. Через ζ_ν обозначим сумму случайного числа случайных величин:

$$\zeta_\nu = \xi_1 + \dots + \xi_\nu, \quad \nu \geq 1 \text{ и } \zeta_0 = 0.$$

Теорема 2.24. *Производящая функция φ_{ζ_ν} случайной величины ζ_ν равна суперпозиции*

$$\varphi_{\zeta_\nu}(s) = \varphi_\nu(\varphi_\xi(s)).$$

Доказательство. Имеем

$$\varphi_{\zeta_\nu}(s) = M s^{\zeta_\nu} = M [M s^{\zeta_\nu} | \nu] = M [\varphi_\xi(s)]^\nu = \varphi_\nu(\varphi_\xi(s)),$$

так как

$$M\{s^{\xi_1 + \dots + \xi_\nu} | \nu = n\} = M s^{\xi_1 + \dots + \xi_n} = [\varphi_\xi(s)]^n. \quad \square$$

Рассмотрим несколько примеров.

Пример 23. Если $\mathbf{P}\{\xi = c\} = 1$, то $f_\xi(t) = e^{itc}$.

Пример 24. Для нормально распределенной случайной величины ξ с параметрами $(0, 1)$ имеем

$$f_{\xi}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} dx = I$$

Положим $z = x - it$ и продолжим выкладку, сославшись на теорему Коши,

$$I = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty-it}^{\infty-it} e^{-\frac{z^2}{2}} dz = e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = e^{-\frac{t^2}{2}}.$$

Пусть теперь случайная величина η нормально распределена с параметрами (a, σ) . Тогда случайная величина $\xi = (\eta - a)/\sigma$ нормально распределена с параметрами $(0, 1)$. Поэтому

$$f_{\eta}(t) = e^{ita} e^{-\frac{(t\sigma)^2}{2}} = e^{ita - \frac{t^2\sigma^2}{2}}.$$

Пример 25. Пусть ξ равномерно распределена на отрезке $[0, 1]$. Тогда

$$f_{\xi}(t) = \int_0^1 e^{itx} dx = \frac{e^{it} - 1}{it}.$$

Пример 26. Для случайной величины ξ , распределенной по Пуассону с параметром μ имеем

$$\varphi_{\xi}(s) = \sum_{k=0}^{\infty} s^k \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} e^{s\mu} = e^{[\mu(s-1)]}.$$

Пример 27. Для биномиально распределенной случайной величины ξ с законом распределения $\mathbf{P}\{\xi = m\} = C_n^m p^m (1-p)^{n-m}$ имеем

$$\varphi_{\xi}(s) = \sum_{m=0}^n C_n^m p^m (1-p)^{n-m} s^m = (ps + 1 - p)^n$$

Формула обращения

Далеко не всякая функция $f(t)$, $t \in \mathbb{R}$, может быть характеристической функцией некоторой случайной величины.

Определение 32. Непрерывная функция $f(t)$ вещественного аргумента t называется неотрицательно определенной на \mathbb{R} , если каковы бы ни были вещественные t_1, \dots, t_n и комплексные ζ_1, \dots, ζ_n выполняется неравенство

$$\sum_{j=1}^n \sum_{k=1}^n f(t_j - t_k) \zeta_j \bar{\zeta}_k \geq 0.$$

Теорема 2.25 (Бохнер–Хинчин). *Для того, чтобы непрерывная функция $f(t)$ такая, что $f(0) = 1$, была характеристической необходимо и достаточно, чтобы она была неотрицательно определенной.*

Необходимость неотрицательной определенности проверяется несложно:

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n f(t_j - t_k) \zeta_j \bar{\zeta}_k &= \sum_{j=1}^n \sum_{k=1}^n \left\{ \int_{\mathbb{R}} e^{ix(t_j - t_k)} dF(x) \right\} \zeta_j \bar{\zeta}_k = \\ &= \int_{\mathbb{R}} \sum_{j=1}^n \sum_{k=1}^n e^{ix(t_j - t_k)} \zeta_j \bar{\zeta}_k dF(x) = \int_{\mathbb{R}} \left(\sum_{j=1}^n e^{ix t_j} \zeta_j \right) \overline{\left(\sum_{k=1}^n e^{ix t_k} \zeta_k \right)} dF(x) = \\ &= \int_{\mathbb{R}} \left| \sum_{k=1}^n e^{ix t_k} \zeta_k \right|^2 dF(x) \geq 0. \quad (2.9) \end{aligned}$$

Сформулируем здесь же теорему Герглотца, дискретный вариант теоремы Бохнера-Хинчина.

Теорема 2.26. *Если последовательность c_n , $n = 0, \pm 1, \dots$, обладает тем свойством, что при любом выборе комплексных чисел ζ_1, \dots, ζ_N и произвольном N*

$$\sum_{k=1}^N \sum_{j=1}^N c_{j-k} \zeta_j \bar{\zeta}_k \geq 0$$

то последовательность c_n может быть записана в форме

$$c_n = \int_{-\pi}^{\pi} e^{inx} d\sigma(x),$$

где $\sigma(x)$ – неубывающая функция с ограниченной вариацией.

Если случайная величина ξ имеет плотность $p_\xi(x)$, то характеристическая функция $f_\xi(t)$, является преобразованием Фурье плотности $p_\xi(x)$. Если предположить, что интеграл $\int_{-\infty}^{\infty} |f_\xi(t)| dt$ конечен, то имеет место формула обращения для преобразования Фурье

$$p_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f_\xi(t) dt.$$

Докажем формулу обращения в общем случае.

Теорема 2.27 (формула обращения). *Если $F(x)$ – функция распределения случайной величины ξ , а $f(t)$ – характеристическая функция для ξ , то для любых точек x, y непрерывности $F(x)$ имеет место равенство*

$$\mathbf{P}\{x \leq \xi \leq y\} = F(y) - F(x) = \frac{1}{2\pi} \lim_{A \rightarrow \infty} \int_{-A}^A \frac{e^{-itx} - e^{-ity}}{it} f(t) dt.$$

Доказательство. Для определенности будем считать, что $x < y$. Имеем

$$I_A = \frac{1}{2\pi} \int_{-A}^A \frac{e^{-itx} - e^{-ity}}{it} f(t) dt = \frac{1}{2\pi} \int_{-A}^A \int_{-\infty}^{\infty} \frac{e^{it(u-x)} - e^{it(u-y)}}{it} dF(u) dt.$$

Подинтегральная функция в повторном интеграле ограничена. Поэтому по теореме Фубини можно

переставить местами интегралы:

$$\begin{aligned}
 I_A &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-A}^A \frac{e^{it(u-x)} - e^{it(u-y)}}{it} dt \right] dF(u) = \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_0^A \frac{e^{it(u-x)} - e^{-it(u-x)} - e^{it(u-y)} + e^{-it(u-y)}}{it} dt \right] dF(u) = \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_0^A \left(\frac{\sin(u-x)t}{t} - \frac{\sin(u-y)t}{t} \right) dt \right] dF(u) = \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left[\int_{A(u-y)}^{A(u-x)} \frac{\sin t}{t} dt \right] dF(u).
 \end{aligned}$$

Из курса анализа известно, что

$$\lim_{\substack{A \rightarrow \infty \\ B \rightarrow \infty}} \frac{1}{\pi} \int_{-B}^A \frac{\sin t}{t} dt = 1.$$

Отсюда следует, что функция

$$S_A(u) = \frac{1}{\pi} \int_{A(u-y)}^{A(u-x)} \frac{\sin t}{t} dt$$

ограничена. Поэтому

$$\lim_{A \rightarrow \infty} I_A = \int_{-\infty}^{\infty} \lim_{A \rightarrow \infty} S_A(u) dF(u).$$

Так как

$$\lim_{A \rightarrow \infty} S_A(u) = \begin{cases} 1, & x < u < y \\ \frac{1}{2}, & u = x \text{ или } u = y \\ 0, & u < x \text{ или } u > y, \end{cases}$$

получаем

$$\lim_{A \rightarrow \infty} I_A = \int_x^y dF(u) = F(y) - F(x). \quad \square$$

Следствие. Характеристическая функция случайной величины однозначно определяет функцию распределения этой случайной величины.

Приведем несколько полезных примеров применения доказанной теоремы.

Пример 28. Пусть ξ_1 и ξ_2 независимые случайные величины, распределенные по нормальному закону с параметрами (a_1, σ_1) и (a_2, σ_2) . Тогда случайная величина $\xi_1 + \xi_2$ также нормально распределена с параметрами $(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Действительно,

$$\begin{aligned}
 f_{\xi_1 + \xi_2}(t) &= f_{\xi_1}(t) f_{\xi_2}(t) = \\
 &= e^{ita_1 - \frac{t^2 \sigma_1^2}{2}} e^{ita_2 - \frac{t^2 \sigma_2^2}{2}} = e^{it(a_1 + a_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}}
 \end{aligned}$$

Пример 29. Сумма двух биномиально распределенных независимых случайных величин ξ_1 и ξ_2 с одинаковым параметром p является биномиально распределенной случайной величиной.

В самом деле,

$$\varphi_{\xi_1+\xi_2}(s) = \varphi_{\xi_1}(s)\varphi_{\xi_2}(s) = (ps+q)^{n_1}(ps+q)^{n_2} = (ps+q)^{n_1+n_2}.$$

Пример 30. Сумма двух независимых пуассоновских случайных величин ξ_1 и ξ_2 является пуассоновской случайной величиной.

Имеем

$$\varphi_{\xi_1+\xi_2}(s) = \varphi_{\xi_1}(s)\varphi_{\xi_2}(s) = e^{\lambda_1(s-1)}e^{\lambda_2(s-1)} = e^{(\lambda_1+\lambda_2)(s-1)}.$$

Как мы видели раньше, если ξ_1 и ξ_2 , распределены равномерно на $[0, 1]$ и независимы, то F_{ξ_1} и $F_{\xi_1+\xi_2}$ существенно разные функции.

Теорема непрерывности

Таким образом, между множеством функций распределения и множеством их характеристических функций имеется взаимно однозначное соответствие. Ограничимся здесь формулировкой теоремы, из которой следует, что это соответствие не только взаимно однозначное, но и взаимно непрерывное.

Теорема 2.28 (теорема непрерывности). Пусть $f_n(t)$ есть характеристическая функция случайной величины ξ_n и $F_{\xi_n}(x) = F_n(x)$. Если при $n \rightarrow \infty$ $f_n(t) \rightarrow f(t)$ в каждой точке $t \in \mathbb{R}$ и $f(t)$ непрерывна при $t = 0$, то

1. $f(t)$ является характеристической функцией некоторой функции распределения $F(x)$;
2. $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности предельной функции.

Обратно, если $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности функции $F(x)$ и $F(x)$ является функцией распределения, то $f_n(t) \rightarrow f(t)$ в каждой точке $t \in \mathbb{R}$, при этом $f(t)$ есть характеристическая функция для функции распределения $F(x)$.

Для целочисленных случайных величин теорема непрерывности формулируется следующим образом:

Теорема 2.29. Пусть

$$\varphi_r(s) = \sum_{n=0}^{\infty} p_n^{(r)} s^n, \quad r = 1, 2, \dots,$$

есть последовательность вероятностных производящих функций и

$$\varphi(s) = \sum_{n=0}^{\infty} p_n s^n,$$

является производящей функцией последовательности $\{p_n\}$. Для того, чтобы при каждом n

$$\lim_{r \rightarrow \infty} p_n^{(r)} = p_n,$$

необходимо и достаточно, чтобы при всех $s \in [0, 1]$

$$\lim_{r \rightarrow \infty} \varphi_r(s) = \varphi(s).$$

Пример 31. Воспользуемся теоремой непрерывности для доказательства предельной теоремы Пуассона.

Пусть $a_n = np_n \rightarrow a$. Тогда

$$(sp_n + 1 - p_n)^n = \left(s \frac{a_n}{n} + 1 - \frac{a_n}{n}\right)^n = \left(1 + \frac{a_n}{n}(s-1)\right)^n \rightarrow e^{a(s-1)}.$$

Поэтому

$$\lim_{n \rightarrow \infty} C_n^m \left(\frac{a_n}{n}\right)^m \left(1 - \frac{a_n}{n}\right)^{n-m} = \frac{a^m}{m!} e^{-a}.$$

2.4. Центральная предельная теорема

Ранее была доказана интегральная теорема Муавра - Лапласа утверждающая, что распределение числа успехов μ в схеме Бернулли при $n \rightarrow \infty$ и постоянном $0 < p < 1$ обладает следующим предельным свойством

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\mu - M\mu}{\sqrt{D\mu}} \leq x \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Предельной функцией является функция нормального распределения вида

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du,$$

которая связана с интегралом Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du$$

соотношением $\Phi(x) = \frac{1}{2} + \Phi_0(x)$.

Этот результат является частным случаем общей *центральной предельной теоремы*. Пусть ξ_n , $n = 1, \dots$, независимы, $M\xi_n = a$, $D\xi_n = \sigma^2$. Положим $S_n = \xi_1 + \dots + \xi_n$ и введем

$$\zeta_n = \frac{S_n - an}{\sigma\sqrt{n}} = \frac{\frac{S_n}{n} - a}{\sigma} \sqrt{n}.$$

Теорема 2.30. Если $0 < \sigma^2 < \infty$, то при $n \rightarrow \infty$

$$\mathbf{P}\{\zeta_n \leq x\} \rightarrow \Phi(x)$$

равномерно относительно $x \in (-\infty, \infty)$.

Доказательство. Можем считать, что $M\xi_k = 0$, иначе можно было бы рассмотреть последовательность $\{\xi'_n = \xi_n - a\}_{n=1}^{\infty}$, при этом последовательность $\{\zeta_n\}$ не изменилась бы. Согласно теореме непрерывности достаточно доказать, что

$$f_{\zeta_n}(t) \rightarrow e^{-\frac{t^2}{2}}, \quad t \in (-\infty, \infty).$$

Положим $f_{\xi_n}(t) = f(t)$. Имеем

$$f_{\zeta_n}(t) = f^n\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Так как

$$f(t) = f(0) + tf'(0) + \frac{t^2}{2}f''(0) + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2),$$

поэтому

$$\begin{aligned} \log f_{\zeta_n}(t) &= n \log \left[1 - \frac{\sigma^2}{2} \left(\frac{t}{\sigma\sqrt{n}} \right)^2 + o\left(\frac{t^2}{n}\right) \right] \sim \\ &\sim n \left(-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right) = -\frac{t^2}{2} + o(1) \rightarrow -\frac{t^2}{2}, \end{aligned} \tag{2.10}$$

так как $n o\left(\frac{t^2}{n}\right) \rightarrow 0$ при $n \rightarrow \infty$ и любом $t \in (-\infty, \infty)$.

Если последовательность функций распределения $F_n(x)$ сходится к непрерывной функции распределения $F(x)$, то сходимость $F_n(x)$ к $F(x)$ равномерна. Достаточно показать, что у каждой точки расширенной числовой прямой $\overline{\mathbb{R}}$ имеется окрестность, в которой сходимость равномерна. Пусть $-\infty < x < y < \infty$ и $\varepsilon > 0$. Найдется n_0 такое, что для всех $n \geq n_0$ выполняются неравенства

$$\begin{aligned} F_n(x) - \varepsilon < F(x) < F_n(x) + \varepsilon, \\ F_n(y) - \varepsilon < F(y) < F_n(y) + \varepsilon. \end{aligned}$$

Тогда для всех $t \in [x, y]$ имеем

$$\begin{aligned} F(x) - \varepsilon < F(x) \leq F(t) \leq F(y) < F(y) + \varepsilon, \\ F(x) - \varepsilon < F_n(x) \leq F_n(t) \leq F_n(y) < F(y) + \varepsilon. \end{aligned}$$

Отсюда следует, что для всех $n \geq n_0$ и $t \in [x, y]$

$$|F(t) - F_n(t)| < (F(y) - F(x)) + 2\varepsilon. \quad (2.11)$$

Для бесконечно удаленной точки $x = -\infty$ имеем

$$\begin{aligned} 0 \leq F(t) \leq F(y) < F(y) + \varepsilon, \\ 0 \leq F_n(t) \leq F_n(y) < F(y) + \varepsilon. \end{aligned}$$

Поэтому для всех $t \in (-\infty, y]$ при $n \geq n_0$ верно неравенство

$$|F(t) - F_n(t)| < F(y) + \varepsilon. \quad (2.12)$$

Аналогично рассматривается случай другой бесконечно удаленной точки $y = \infty$. Найдется такое n_0 , что для всех $n \geq n_0$ выполняется неравенство

$$F(y) - \varepsilon \leq F_n(y) \leq F(y) + \varepsilon.$$

Поэтому для всех $t \in (y, \infty]$ при $n \geq n_0$ имеют место соотношения

$$\begin{aligned} F(y) - \varepsilon \leq F_n(y) \leq F_n(t) \leq 1 \\ F(y) - \varepsilon \leq F(y) \leq F(t) \leq 1. \end{aligned}$$

Отсюда получаем, что

$$|F_n(t) - F(t)| \leq 1 - F(y) + \varepsilon. \quad (2.13)$$

Из неравенств (2.11), (2.12) и (2.13) следует возможность выбора у каждой точки расширенной числовой прямой $\overline{\mathbb{R}}$ окрестности, в которой сходимость равномерна. \square

Центральная предельная теорема имеет место также при некоторых более общих условиях, налагаемых на последовательность независимых случайных величин. Сформулируем одну из подобного рода теорем.

Пусть теперь ξ_k , $k = 1, \dots$, независимы и имеют не обязательно одно и то же распределение. Введем обозначения

$$M\xi_k = a_k, \quad D\xi_k = b_k^2, \quad M|\xi_k - a_k|^3 = c_k^3$$

и

$$A_n = \sum_{k=1}^n a_k, \quad B_n^2 = \sum_{k=1}^n b_k^2, \quad C_n^3 = \sum_{k=1}^n c_k^3.$$

Теорема 2.31 (Теорема Ляпунова). *Если случайные величины ξ_1, ξ_2, \dots независимы, a_k, b_k, c_k конечны и $C_n/B_n \rightarrow 0$, то*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\xi_1 + \dots + \xi_n - A_n}{B_n} \leq x \right\} = \Phi(x).$$

Хорошо известной иллюстрацией сформулированной теоремы является рассеивание артиллерийских снарядов при стрельбе по цели. На траекторию полета снаряда действует большое количество независимых факторов, причем влияние каждого фактора само по себе невелико. Это отклонения в количестве заряда, отклонения в весе и размерах снаряда, отклонения во влажности и температуре воздуха, направление и сила ветра на разных высотах и т.д. В результате, отклонение снаряда от точки прицеливания удивительно точно описывается нормальным законом.

Нужно не забывать, что центральная предельная теорема не доказывает нормальность рассеяния снаряда, а дает только гипотезе нормальности некоторое теоретическое обоснование. Ведь математические теории строятся не на реальных явлениях, а на математических моделях. Поэтому в применении теории вероятностей всегда следует заботиться о том, чтобы рассматриваемая модель правильно отражала соответствующее явление.

Пример 32. (Ошибки измерения). При измерении некоторой величины a мы получаем приближенное значение ξ . Сделанная ошибка $\delta = \xi - a$ может быть представлена в виде суммы двух ошибок

$$\delta = (\xi - M\xi) + (M\xi - a),$$

первая из которых $\xi - M\xi$ называется случайной ошибкой, а вторая $M\xi - a$ — систематической ошибкой. Хорошие методы измерения не должны иметь систематической ошибки. Будем полагать $M\xi = a$. Случайная ошибка δ имеет нулевое математическое ожидание, $M\delta = 0$. Пусть $D\delta = \sigma^2$. Для уменьшения этой ошибки производят n независимых измерений ξ_1, \dots, ξ_n и принимают за оценку измеримой величины a среднее арифметическое

$$\hat{a} = \frac{\xi_1 + \dots + \xi_n}{n}.$$

Допускаемую при этом погрешность можно оценить с помощью центральной предельной теоремы. Согласно этой теореме сумма $\xi_1 + \dots + \xi_n$ одинаково распределенных независимых случайных величин с $M\xi_n = a$ и $D\xi_n = \sigma^2 > 0$ асимптотически нормально распределена с параметрами $(an, \sigma\sqrt{n})$. Поэтому \hat{a} при больших n асимптотически нормальна с параметрами $(a, \sigma/\sqrt{n})$ и

$$\mathbf{P}\{|\hat{a} - a| \leq \varepsilon\} \approx \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon\sqrt{n}/\sigma}^{\varepsilon\sqrt{n}/\sigma} e^{-\frac{u^2}{2}} du$$

Дисперсия $D(\hat{a} - a)$ погрешности $\hat{a} - a$ уменьшится в сравнении с дисперсией ξ и станет равной

$$D(\hat{a} - a) = \frac{D\xi_1 + \dots + D\xi_n}{n^2} = \frac{\sigma^2}{n}.$$

2.5. Многомерные характеристические функции

Определение и свойства

Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор и $t = (t_1, \dots, t_n) \in \mathbb{R}^n$.

Определение 33. Характеристическая функция $f_\xi(t)$ случайного вектора ξ определяется соотношением

$$f_\xi(t) = M e^{i(t, \xi)} = M e^{i \sum_{k=1}^n t_k \xi_k}.$$

Перечислим свойства характеристических функций.

1. При всех $t \in \mathbb{R}^n$ имеем $|f_\xi(t)| \leq 1$, $f_\xi(0) = 1$.
2. $f_\xi(-t) = \overline{f_\xi(t)}$.
3. $f_\xi(t)$ равномерно непрерывна по t .

Доказательство. Обозначим через A событие $A = \{|\xi_k| \leq x, k = 1, \dots, n\}$ и пусть $h \in \mathbb{R}^n$. Оценим приращение

$$\begin{aligned} |f_\xi(t+h) - f_\xi(t)| &= |Me^{i(t,\xi)}(e^{i(h,\xi)} - 1)| \leq M|e^{i(h,\xi)} - 1| = \\ &= M|e^{i(h,\xi)} - 1|I_A + M|e^{i(h,\xi)} - 1|I_{\bar{A}} \leq M|(h,\xi)|I_A + 2MI_{\bar{A}} \leq \\ &\leq x|h| + 2\mathbf{P}\{\xi \notin [-x, x]^n\}, \end{aligned}$$

где $|h| = \sum_{k=1}^n |h_k|$ и $[-x, x]^n$ – прямоугольник в \mathbb{R}^n :

$$\{t = (t_1, \dots, t_n) \in \mathbb{R}^n : |t_k| \leq x, k = 1, \dots, n\}.$$

Пусть дано $\varepsilon > 0$. Выберем x так, чтобы выполнялось неравенство $\mathbf{P}\{\xi \notin [-x, x]^n\} < \varepsilon/4$. Тогда при $|h| < \varepsilon/(2x)$ имеем $|f_\xi(t+h) - f_\xi(t)| < \varepsilon$. \square

4. Если ξ_1, \dots, ξ_n – независимые случайные векторы и $\zeta = \xi_1 + \dots + \xi_n$, то

$$f_\zeta(t) = \prod_{k=1}^n f_{\xi_k}(t).$$

5. Если известна характеристическая функция $f_\xi(t)$ случайного вектора $\xi = (\xi_1, \dots, \xi_n)$, то характеристическая функция случайного вектора $(\xi_{k_1}, \dots, \xi_{k_j})$ можно получить, положив все t_k , за исключением t_{k_1}, \dots, t_{k_j} , равными нулю.

6. Если $\eta = \xi C$ – линейное преобразование с матрицей $C = \|c_{\alpha\beta}\|$, $\alpha = 1, \dots, m$; $\beta = 1, \dots, k$, то

$$f_\eta(t) = f_\xi(tC'),$$

где C' – сопряженная с C матрица.

Доказательство.

$$f_\eta(t) = Me^{i(t,\eta)} = Me^{i(t,\xi C)} = Me^{it(C'\xi')} = Me^{i(tC')\xi'} = Me^{i(tC',\xi)} = f_\xi(tC'),$$

здесь ξ' обозначает вектор-столбец. \square

Пример 33. Если $\eta = \xi C + b$ – аффинное преобразование, то

$$f_\eta(t) = e^{i(t,b)} f_\xi(tC').$$

7. Если $\xi = (\xi_1, \dots, \xi_n)$ и конечен момент $M(|\xi_1| + \dots + |\xi_n|)^k$, то $f_\xi(t)$ имеет все производные порядка k и для всех натуральных k_1, \dots, k_n , для которых $k = k_1 + \dots + k_n$, выполняются соотношения

$$\frac{\partial^k f_\xi(0, \dots, 0)}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} = i^k M \xi_1^{k_1} \dots \xi_n^{k_n}.$$

и

$$f_\xi(t) = \sum_{\alpha=0}^k i^\alpha \sum_{\alpha_1 + \dots + \alpha_n = \alpha} \frac{t_1^{\alpha_1} \dots t_n^{\alpha_n}}{\alpha_1! \dots \alpha_n!} M \xi_1^{\alpha_1} \dots \xi_n^{\alpha_n} + o(|t|^k)$$

при $|t_1| + \dots + |t_n| \rightarrow 0$.

Свойство 7 доказывается так же, как и в одномерном случае.

Определение 34. Пусть $\zeta = (\xi_1, \dots, \xi_r)$ – случайный вектор с целочисленными случайными компонентами и $p_\alpha = \mathbf{P}\{\zeta = \alpha\} = \mathbf{P}\{\xi_1 = \alpha_1, \dots, \xi_r = \alpha_r\}$, где $\alpha = (\alpha_1, \dots, \alpha_r)$. Многомерная производящая функция $\varphi_\zeta(s_1, \dots, s_r)$ определяется соотношением

$$\varphi_\zeta(s_1, \dots, s_r) = Ms_1^{\xi_1} \dots s_r^{\xi_r} = \sum_{\alpha} p_\alpha s_1^{\alpha_1} \dots s_r^{\alpha_r}.$$

Характеристическая функция $f_\zeta(t_1, \dots, t_r)$ и производящая функция $\varphi_\zeta(s_1, \dots, s_r)$ связаны соотношением

$$f_\zeta(t_1, \dots, t_r) = \varphi_\zeta(e^{it_1}, \dots, e^{it_r}).$$

Определение 35. Случайный вектор $\zeta = (\xi_1, \dots, \xi_r)$ называется полиномиально распределенным, если

$$\mathbf{P}\{\zeta = \alpha\} = \mathbf{P}\{\xi_1 = \alpha_1, \dots, \xi_r = \alpha_r\} = \frac{n!}{\alpha_1! \dots \alpha_r!} p_1^{\alpha_1} \dots p_r^{\alpha_r}, \quad \text{где } p_1 + \dots + p_r = 1 \text{ и } \alpha_1 + \dots + \alpha_r = n.$$

Для полиномиального распределения

$$\varphi_\zeta(s_1, \dots, s_r) = (p_1 s_1 + \dots + p_r s_r)^n.$$

Пример 34. Пусть ξ_1, \dots, ξ_k – нормально распределенные независимые случайные величины с $M\xi_\alpha = a_\alpha$ и $D\xi_\alpha = b_\alpha$. Тогда

$$f_{\xi_1 \dots \xi_k}(t_1, \dots, t_k) = e^{i \sum_{\alpha=1}^k t_\alpha a_\alpha - \frac{1}{2} \sum_{\alpha=1}^k b_\alpha t_\alpha^2}.$$

Пусть случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ имеет непрерывную плотность $p_\xi(x)$ и суммируемую характеристическую функцию $f_\xi(t) \in L_1$. Тогда плотность $p_\xi(x)$ восстанавливается по характеристической функции $f_\xi(t)$ однозначно при помощи обратного преобразования Фурье

$$p_\xi(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i(t,x)} f_\xi(t) dt.$$

Формула обращения в общем случае является обобщением своего одномерного аналога.

Теорема 2.32. Пусть $\Delta = \{x : a_k < x_k < b_k\}$, $k = 1, \dots, n$, и функция распределения $F_\xi(x)$ непрерывна на границах Δ . Тогда

$$\mathbf{P}\{\xi \in \Delta\} = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^n} \int_{-T}^T \dots \int_{-T}^T \prod_{k=1}^n \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} f_\xi(t) dt_1 \dots dt_n.$$

Следствием этой теоремы является утверждение о взаимно однозначном соответствии между характеристическими функциями и функциями распределения.

Приведем здесь еще один критерий независимости случайных величин.

8. Для независимости ξ_1, \dots, ξ_n необходимо и достаточно, чтобы

$$f_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) = \prod_{k=1}^n f_{\xi_k}(t_k).$$

Необходимость следует из мультипликативного свойства математического ожидания. Достаточность следует из доказанной формулы обращения (см. теорему 27).

Как и в одномерном случае приведем только формулировку *теоремы непрерывности*.

Теорема 2.33. Для того, чтобы последовательность многомерных функций распределения $F^{(n)}(x)$ сходилась к некоторой функции распределения $F(x)$ в каждой точке непрерывности предельной функции, необходимо и достаточно, чтобы при любом t последовательность соответствующих характеристических функций $f^{(n)}(t)$ сходилась к пределу $f(t)$, непрерывному в точке $t = 0$. При этом $f(t)$ есть характеристическая функция распределения $F(x)$.

2.6. Многомерное нормальное распределение

Невырожденное нормальное распределение

Определение 36. Определим невырожденное центрированное многомерное нормальное распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ как распределение с плотностью

$$p_\xi(x) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} e^{-\frac{1}{2}Q(x)}, \quad x \in \mathbb{R}^n$$

где $Q(x) = x A x' = \sum_{i,j=1}^n a_{ij} x_i x_j$, $|A|$ – определитель положительно определенной матрицы $\|a_{i,j}\|$.

Здесь x' обозначает вектор-столбец. Для центрированного распределения ниже будет доказано, что $M\xi = 0$. Распределение случайного вектора $\xi + a$, где $a \in \mathbb{R}^n$ – постоянный вектор, также будем называть невырожденным нормальным распределением.

Найдем характеристическую функцию случайного вектора ξ . Покажем, что

$$f_\xi(t) = e^{-\frac{1}{2}t \overline{M} t'},$$

где $\overline{M} = A^{-1}$ есть матрица, обратная к A , совпадающая с матрицей $\|m_{ij}\|$ вторых моментов распределения ξ :

$$m_{ij} = M \xi_i \xi_j.$$

В самом деле,

$$f_\xi(t) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{itx' - \frac{1}{2}x A x'} dx_1 \dots dx_n$$

Выберем ортогональную матрицу C так, чтобы $D = C A C'$ была бы диагональной матрицей, и обозначим через μ_1, \dots, μ_n значения диагональных элементов. Сделаем замену переменных, положив $x = y C$ и $t = v C$. Тогда

$$\begin{aligned} |A| &= |D| = \prod_{k=1}^n \mu_k, \\ itx' - \frac{1}{2}x A x' &= ivy' - \frac{1}{2}y D y' = \\ &= i \sum_{k=1}^n v_k y_k - \frac{1}{2} \sum_{k=1}^n \mu_k y_k^2 \end{aligned} \quad (2.14)$$

Поэтому

$$\begin{aligned} f_\xi(t) &= \frac{\sqrt{|D|}}{(2\pi)^{n/2}} \prod_{k=1}^n \int_{-\infty}^{\infty} e^{iv_k y_k - \frac{1}{2}\mu_k y_k^2} dy_k = \prod_{k=1}^n e^{-\frac{1}{2}\frac{v_k^2}{\mu_k}} = e^{-\frac{1}{2}v E^{-1} v'} = \\ &= e^{-\frac{1}{2}t C' E^{-1} C t'} = e^{-\frac{1}{2}t A^{-1} t'}. \end{aligned}$$

С другой стороны, так как все моменты случайного вектора ξ существуют, то в окрестности точки $t = 0$ имеет место равенство

$$\begin{aligned} f_\xi(t) &= 1 - \frac{1}{2}t A^{-1} t' + o(|t|^2) = \\ &= 1 + it(M\xi)' + \frac{i^2}{2}t \overline{M} t' + o(|t|^2). \end{aligned} \quad (2.15)$$

Отсюда следует, что компоненты $M\xi_k$, $k = 1, \dots, n$, вектора $M\xi$ равны нулю и $A^{-1} = \overline{M}$.

Из доказанного вытекает очень важное свойство нормальных распределений: *компоненты нормально распределенного вектора $\xi = (\xi_1, \dots, \xi_n)$ независимы тогда и только тогда, когда коэффициенты корреляции $\rho(\xi_i, \xi_j) = 0$ при всех $i \neq j$.*

Доказательство. Действительно, если \overline{M} – диагональная матрица, то матрица $A = \overline{M}^{-1}$ также диагональная и $p_\xi(x)$ равна произведению плотностей компонент вектора ξ . Обратно, если ξ_1, \dots, ξ_n независимы, то A диагональна и, значит, диагональна матрица \overline{M} . \square

Пусть $\xi = (\xi_1, \dots, \xi_n)$ невырожденный центрированный нормально распределенный случайный вектор с плотностью $p_\xi(x)$

$$p_\xi(x) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} e^{-\frac{1}{2}xAx'}, \quad x \in \mathbb{R}^n.$$

Выберем ортогональную матрицу B такую, что $B'AB$ является диагональной матрицей. В этом случае, как мы видели, компоненты случайного вектора $\zeta = \xi B$ нормально распределены с нулевым математическим ожиданием и независимы. Поэтому $\xi = \zeta B'$, а значит, найдется невырожденная $(n \times n)$ -матрица C и случайный вектор η , компоненты которого независимы и нормально распределены с параметрами $(0, 1)$, такие, что $\xi = \eta C$.

В общем случае многомерным нормально распределенным является случайный вектор ξ имеющий представление

$$\xi = \eta C + a,$$

где компоненты случайного вектора η нормально распределены с параметрами $(0, 1)$ и независимы, C – невырожденная $(n \times n)$ -матрица и a – постоянный вектор, равный $M\xi$.

Определение 37. Распределение $\xi = (\xi_1, \dots, \xi_n)$ с плотностью

$$p_\xi(x) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2}$$

называется сферически нормальным распределением. Характеристическая функция сферически нормального распределения имеет вид

$$f_\xi(t) = e^{-\frac{\sigma^2}{2}(t,t)}.$$

Сферически нормальное распределение инвариантно относительно ортогональных преобразований $\eta = \xi C$, так как

$$f_\eta(t) = f_\xi(tC') = e^{-\frac{\sigma^2}{2}(tC', tC')} = e^{-\frac{\sigma^2}{2}(t,t)}.$$

Многомерное нормальное распределение (общий случай)

Пусть $\xi = (\xi_1, \dots, \xi_r)$ – r -мерный случайный вектор. Положим

$$\eta_i = \sum_{\ell=1}^r c_{\ell i} \xi_\ell + a_i, \quad i = 1, \dots, m, \quad (2.16)$$

или в векторной форме

$$\eta = \xi C + a, \quad (2.17)$$

где $\eta = (\eta_1, \dots, \eta_m)$, $a = (a_1, \dots, a_m)$ и $C = \|c_{ij}\|$ – $(r \times m)$ -матрица. Здесь вектор a и матрица C постоянны. Через $D[\xi] = D[\xi_1, \dots, \xi_r]$ будем обозначать ковариационную матрицу $\| \text{cov}(\xi_i, \xi_j) \|$.

Теорема 2.34. Для любой постоянной матрицы C в (2.17) имеет место равенство

$$D[\eta] = C' D[\xi] C. \quad (2.18)$$

Доказательство. Из (2.17) следует, что $\eta_i - M\eta_i = \sum_{\ell=1}^r c_{\ell i} (\xi_\ell - M\xi_\ell)$ и

$$(\eta_i - M\eta_i)(\eta_j - M\eta_j) = \sum_{k, \ell=1}^r c_{\ell i} c_{k j} (\xi_\ell - M\xi_\ell)(\xi_k - M\xi_k).$$

Вычисляя от обеих частей последнего равенства математическое ожидание, получаем

$$\text{cov}(\eta_i, \eta_j) = \sum_{\ell, k=1}^r c_{ki} \text{cov}(\xi_k, \xi_\ell) c_{\ell j} = \sum_{\ell, k=1}^r c'_{ik} \text{cov}(\xi_k, \xi_\ell) c_{\ell j},$$

где $c'_{ik} = c_{ki}$. \square

Определение 38. Многомерным нормальным распределением m случайных величин назовем распределение вектора $\eta = (\eta_1, \dots, \eta_m)$, где вектор η определен формулой (2.17), а компоненты ξ_1, \dots, ξ_r случайного вектора ξ независимы и каждая из них распределена нормально с параметрами $(0, 1)$.

Отметим некоторые общие свойства многомерного нормального распределения. Пусть вектор $\eta = (\eta_1, \dots, \eta_m)$ распределен нормально. Имеем:

1. Одномерные распределения координат η_i являются нормальными, если $D\eta_i > 0$.
2. Случайный вектор ζ , представимый в виде $\zeta = A\eta$, где A – постоянная матрица, имеет многомерное нормальное распределение.
3. Любая линейная комбинация $\zeta = a_1\eta_1 + \dots + a_m\eta_m$ компонент вектора η имеет нормальное распределение, если $D\zeta > 0$.

Эти свойства непосредственно следуют из определения и того факта, что сумма независимых нормально распределенных случайных величин нормально распределена.

Теорема 2.35. Для того, чтобы случайный вектор $\eta = (\eta_1, \dots, \eta_m)$ был нормально распределен, необходимо и достаточно, чтобы характеристическая функция случайного вектора η имела вид

$$f_\eta(t) = e^{ita' - \frac{1}{2}tBt'}, \quad (2.19)$$

где $a = (a_1, \dots, a_m)$ – постоянный вектор, $B = \|b_{\alpha\beta}\|$ – симметричная неотрицательно определенная $(m \times m)$ -матрица.

Доказательство. Найдем характеристическую функцию нормально распределенного случайного вектора. Согласно (2.16) и определению характеристической функции

$$f_\eta(t) = M e^{it\eta'} = M \exp \left\{ i \sum_{\ell=1}^r \left(\sum_{k=1}^m t_k c_{\ell k} \right) \xi_\ell + i \sum_{k=1}^m t_k a_k \right\}$$

Используя независимость случайных величин ξ_1, \dots, ξ_r , получаем

$$\begin{aligned} f_\eta(t) &= \exp \left\{ i \sum_{k=1}^m t_k a_k \right\} \cdot \prod_{\ell=1}^r M \exp \left\{ i \left(\sum_{k=1}^m t_k c_{\ell k} \right) \xi_\ell \right\} = \\ &= \exp(it a') \prod_{\ell=1}^r \exp \left\{ -\frac{1}{2} \left(\sum_{k=1}^m t_k c_{\ell k} \right)^2 \right\} = \\ &= \exp(it a') \exp \left\{ -\frac{1}{2} \sum_{k_1, k_2=1}^m t_{k_1} t_{k_2} \left(\sum_{\ell=1}^r c_{\ell k_1} c_{\ell k_2} \right) \right\}. \end{aligned}$$

Так как

$$b_{k_1 k_2} = \text{cov}(\eta_{k_1}, \eta_{k_2}) = \sum_{\ell=1}^r c_{\ell k_1} c_{\ell k_2} \text{cov}(\xi_{k_1}, \xi_{k_2}) = \sum_{\ell=1}^r c_{\ell k_1} c_{\ell k_2}, \quad (\text{см. теорему 2.34}),$$

то окончательно получаем

$$f_\eta(t) = \exp \left\{ it a' - \frac{1}{2} \sum_{k_1, k_2=1}^m b_{k_1 k_2} t_{k_1} t_{k_2} \right\}.$$

Пусть теперь (2.19) имеет место. Тогда каждая компонента η_k имеет характеристическую функцию вида

$$f_{\eta_k}(t) = e^{ita_k - \frac{b_{kk}}{2}t^2},$$

то есть нормально распределена с $M\eta_k = a_k$ и $D\eta_k = b_{kk}$. Далее удобнее будет перейти к центрированному вектору $\eta - a$. Сохраняя обозначения, будем считать, что $M\eta_k = 0$. Поскольку конечны все моменты $M\eta_\alpha^2$, то конечны и смешанные моменты $M\eta_\alpha\eta_\beta$, поэтому их можно вычислить, пользуясь равенством

$$\text{cov}(\eta_\alpha, \eta_\beta) = M\eta_\alpha\eta_\beta = -\frac{\partial^2 f_\eta}{\partial t_\alpha \partial t_\beta}(0) = b_{\alpha\beta}.$$

Таким образом, $B = \|b_{\alpha\beta}\|$ является ковариационной матрицей случайного вектора η .

Для любой $(m \times m)$ -матрицы $S = \|s_{\alpha\beta}\|$ случайный вектор $\zeta = \eta S$ имеет нулевое математическое ожидание $M\zeta = 0$, ковариационную матрицу $\|\text{cov}(\zeta_\alpha, \zeta_\beta)\| = S'BS$ и характеристическую функцию

$$f_\zeta(t) = f_\eta(tS') = e^{-\frac{1}{2}t'S'BS't'}.$$

Пусть S – такая ортогональная матрица, что

$$S'BS = \left\| \begin{array}{cccc} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_{mm} \end{array} \right\| = D, \quad d_{\alpha\alpha} \geq 0.$$

Тогда

$$f_\zeta(t) = e^{-\frac{1}{2}\sum_{\alpha=1}^m d_{\alpha\alpha}t_\alpha^2},$$

поэтому ζ_1, \dots, ζ_m независимы, причем для тех α , для которых $d_{\alpha\alpha} > 0$ случайные величины ζ_α нормально распределены с параметрами $(0, \sqrt{d_{\alpha\alpha}})$, а при $d_{\alpha\alpha} = 0$ с вероятностью 1 имеем $\zeta_\alpha = 0$.

Если ранг r матрицы B меньше m , то можем считать, что $d_{\alpha\alpha} \neq 0$ для $\alpha = 1, \dots, r$ и $d_{r+1, r+1} = \dots = d_{mm} = 0$. В этом случае $\mathbf{P}\{\zeta_{r+1} = \dots = \zeta_m = 0\} = 1$, то есть случайные величины η_1, \dots, η_m сосредоточены на пространстве, определяемом равенствами

$$\sum_{\beta=1}^m s_{\alpha\beta}\eta_\beta = 0, \quad \alpha = r+1, \dots, m.$$

Тогда

$$\eta_i = \sum_{k=1}^r s_{ki}\zeta_k, \quad i = 1, \dots, m.$$

Положим $\xi_\alpha = \zeta_\alpha/\sqrt{d_{\alpha\alpha}}$ и введем $(r \times m)$ -матрицу C с компонентами $c_{\alpha\beta} = \sqrt{d_{\alpha\alpha}}s_{\beta\alpha}$. Случайные величины ξ_α независимы и нормально распределены с параметрами $(0, 1)$. В результате, имеем

$$\eta = \xi C,$$

то есть

$$\eta_i = \sum_{\alpha=1}^r c_{\alpha i}\xi_\alpha, \quad i = 1, \dots, m. \quad \square$$

Теорема 2.36. Если $r = m$ и определитель $\det C \neq 0$, то r -мерное нормальное распределение η является невырожденным абсолютно непрерывным, и его плотность распределения имеет вид

$$p_\eta(x_1, \dots, x_r) = \frac{1}{(2\pi)^{r/2}\sqrt{\det B}} e^{-\frac{1}{2}Q(x-a)} \quad (\text{см. определение 36}),$$

где $B = \|b_{k\ell}\|$ – невырожденная $(r \times r)$ -матрица, $b_{k\ell} = \text{cov}(\eta_k, \eta_\ell)$, $x = (x_1, \dots, x_r)$, $a = (a_1, \dots, a_r)$, $a_k = M\eta_k$, $Q(x) = \sum b_{k\ell}^* x_k x_\ell$ – квадратичная форма, коэффициенты которой составляют матрицу $\|b_{k\ell}^*\|$, обратную к матрице B .

Доказательство. Плотность распределения случайного вектора ξ из теоремы 2.35 имеет вид

$$p_{\xi}(x_1, \dots, x_r) = \frac{1}{(\sqrt{2\pi})^r} e^{-\frac{1}{2} \sum_{k=1}^r x_k^2}.$$

Так как $\det C \neq 0$ преобразование вида

$$y_i = g_i(x_1, \dots, x_r) = \sum_{k=1}^r c_{ik} x_k + a_i \quad i = 1, \dots, m,$$

является взаимнооднозначным и его якобиан J равен $\det C$. Пусть $x_i = \sum_{\ell=1}^r c_{i\ell}^* (y_{\ell} - a_{\ell})$, где $c_{i\ell}^*$ – элементы матрицы, обратной к C . Тогда

$$p_{\eta}(y_1, \dots, y_r) = |J|^{-1} p_{\xi}(g^{-1}(y)) = (\sqrt{2\pi})^{-r} |J|^{-1} e^{\frac{1}{2} Q(y-a)},$$

где

$$Q(y-a) = \sum_{k=1}^r x_k^2 = \sum_{k=1}^r \left(\sum_{\ell=1}^r c_{k\ell}^* (y_{\ell} - a_{\ell}) \right)^2 = \sum_{\ell_1, \ell_2=1}^r (y_{\ell_1} - a_{\ell_1})(y_{\ell_2} - a_{\ell_2}) \left(\sum_{k=1}^r c_{k\ell_1}^* c_{k\ell_2}^* \right)$$

Так как $D[\xi] = E$ (E – единичная матрица) имеем $B = C' D[\xi] C = C' C$. Следовательно,

$$\| b_{\ell_1 \ell_2}^* \| = B^{-1} = C^{-1} (C')^{-1} = \left\| \sum_{k=1}^r c_{k\ell_1}^* c_{k\ell_2}^* \right\|$$

и $\det B = |\det C|^2$. Заменяя $|J|$ на $\sqrt{\det B}$ и $\sum_{k=1}^r c_{k\ell_1}^* c_{k\ell_2}^*$ на $b_{\ell_1 \ell_2}^*$, получим утверждение теоремы. \square

Предельная теорема для сумм независимых одинаково распределенных случайных векторов

Докажем здесь предельную теорему для сумм независимых одинаково распределенных случайных векторов.

Теорема 2.37. Пусть ξ_1, ξ_2, \dots – последовательность независимых одинаково распределенных случайных векторов $\xi_n = (\xi_{n1}, \dots, \xi_{nk})$ с $M\xi_{n\alpha} = a_{\alpha}$ и с конечными ковариациями $\text{cov}(\xi_{n\alpha}, \xi_{n\beta}) = b_{\alpha\beta}$, $\alpha, \beta = 1, \dots, k$. Обозначим $\zeta_n = \xi_1 + \dots + \xi_n$. Тогда функция распределения случайного вектора

$$\zeta'_n = \frac{\zeta_n - na}{\sqrt{n}}$$

слабо сходится к нормальной функции распределения с нулевым математическим ожиданием и матрицей ковариации $B = \| b_{\alpha\beta} \|$.

Доказательство. Обозначим через $f(t)$ характеристическую функцию случайного вектора $\tilde{\xi}_n = \xi_n - a$. Поскольку $M\tilde{\xi}_n = 0$ и $M\tilde{\xi}_{n\alpha}\tilde{\xi}_{n\beta} = b_{\alpha\beta}$, то по свойству характеристической функции

$$f\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n} \sum_{\alpha, \beta=1}^k b_{\alpha\beta} t_{\alpha} t_{\beta} + o\left(\frac{1}{n}\right).$$

Поэтому при любом t при $n \rightarrow \infty$

$$f_{\zeta'_n}(t) = \left[f\left(\frac{t}{\sqrt{n}}\right) \right]^n \rightarrow e^{-\frac{1}{2} \sum_{\alpha, \beta=1}^k b_{\alpha\beta} t_{\alpha} t_{\beta}},$$

отсюда и из теоремы 2.33 следует доказываемое утверждение. \square

Случайный вектор $\xi = (\xi_1, \dots, \xi_k)$ имеет полиномиальное распределение в n независимых испытаниях с вероятностями исходов $p = (p_1, \dots, p_k)$ в отдельном испытании, если

$$\mathbf{P}\{\xi_1 = n_1, \dots, \xi_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

где $n_1 + \dots + n_k = n$.

Важным частным случаем предыдущей теоремы является следующее утверждение

Теорема 2.38. Пусть $\xi = (\xi_1, \dots, \xi_k)$ имеет полиномиальное распределение в n испытаниях с вероятностями исходов $p = (p_1, \dots, p_k)$ в отдельном испытании. Распределение вектора $(\xi - np)/\sqrt{n}$ при $n \rightarrow \infty$ слабо сходится к нормальному с нулевым средним и матрицей ковариации $\|\delta_{\alpha\beta} p_\alpha - p_\alpha p_\beta\|$, где $\delta_{\alpha\beta}$ – символ Кронекера.

Доказательство. Случайный вектор ξ представим в виде суммы $\eta_1 + \dots + \eta_n$ независимых векторов $\eta_\alpha = (\eta_{\alpha 1}, \dots, \eta_{\alpha k})$, где $\eta_{\alpha\beta} = 1$, если при α -м испытании произошел исход β и $\eta_{\alpha\beta} = 0$ в противном случае. Совместный закон распределения случайных величин $\eta_{\alpha\beta}$ и $\eta_{\alpha\gamma}$ имеет вид

$$\begin{aligned} \mathbf{P}\{\eta_{\alpha\beta} = 1, \eta_{\alpha\gamma} = 1\} &= 0, \quad \mathbf{P}\{\eta_{\alpha\beta} = 1, \eta_{\alpha\gamma} = 0\} = p_\beta, \\ \mathbf{P}\{\eta_{\alpha\beta} = 0, \eta_{\alpha\gamma} = 1\} &= p_\gamma, \quad \mathbf{P}\{\eta_{\alpha\beta} = 0, \eta_{\alpha\gamma} = 0\} = 1 - p_\beta - p_\gamma \quad \beta \neq \gamma. \end{aligned}$$

Учитывая, что $M\eta_{\alpha\beta} = p_\beta$ и $\text{cov}(\eta_{\alpha\beta}, \eta_{\alpha\gamma}) = p_\beta \delta_{\beta\gamma} - p_\beta p_\gamma$, в силу предыдущей теоремы получаем требуемое утверждение. \square

2.7. Распределения, связанные с многомерным нормальным распределением

χ^2 -распределение

Пусть $\xi = (\xi_1, \dots, \xi_k)$ имеет сферически нормальное распределение. Тогда ξ_i , $i = 1, \dots, k$, являются нормально распределенными с параметрами $(0, 1)$ независимыми случайными величинами. Положим $\chi_k^2 = \xi_1^2 + \dots + \xi_k^2$ и $\chi_k = \sqrt{\xi_1^2 + \dots + \xi_k^2}$. Найдем функцию распределения и плотность случайной величины χ_k . Переходя к сферическим координатам, получим

$$\mathbf{P}\{\chi_k \leq r\} = \frac{1}{(2\pi)^{k/2}} \int_{|x| \leq r} e^{-\frac{1}{2} x x'} dx_1 \dots dx_k = c_k \int_0^r t^{k-1} e^{-\frac{1}{2} t^2} dt$$

Поэтому

$$p_{\chi_k}(r) = c_k r^{k-1} e^{-\frac{r^2}{2}}, \quad r \geq 0.$$

Постоянная c_k находится из условия

$$\int_0^\infty p_{\chi_k}(r) dr = 1.$$

Имеем

$$\begin{aligned} c_k \int_0^\infty t^{k-1} e^{-\frac{t^2}{2}} dt &= c_k 2^{\frac{k}{2}} \int_0^\infty t^{k-1} e^{-t^2} dt = c_k 2^{\frac{k}{2}-1} \int_0^\infty t^{\frac{k-2}{2}} e^{-t} dt = \\ &= c_k 2^{\frac{k}{2}-1} \int_0^\infty t^{\frac{k}{2}-1} e^{-t} dt = c_k 2^{\frac{k}{2}-1} \Gamma\left(\frac{k}{2}\right) = 1. \end{aligned} \tag{2.20}$$

Поэтому

$$p_{\chi_k}(r) = \frac{r^{k-1}}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})} e^{-\frac{r^2}{2}}, \quad r \geq 0.$$

Плотность распределения случайной величины χ_k^2 имеет вид

$$p_{\chi_k^2}(r) = \frac{1}{2\sqrt{r}} p_{\chi_k}(\sqrt{r}) = \frac{r^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} e^{-\frac{r}{2}}, \quad r \geq 0. \quad (2.21)$$

Случайная величина с плотностью (2.21) называется χ^2 -распределенной с k степенями свободы.

Распределение Стьюдента

Пусть случайные величины $\xi_0, \xi_1, \dots, \xi_k$ независимы и нормально распределены с параметрами $(0, 1)$. Случайная величина

$$\tau_k = \frac{\xi_0}{\sqrt{\frac{1}{k} \sum_{i=1}^k \xi_i^2}} = \frac{\xi_0 \sqrt{k}}{\chi_k}$$

называется отношением Стьюдента с k степенями свободы. Найдем функцию распределения и плотность τ_k :

$$S_k(x) = \mathbf{P}\{\tau_k \leq x\} = \iint_{\substack{\frac{u}{v} \leq x \\ v > 0}} p_{\xi_0 \sqrt{k}}(u) p_{\chi_k}(v) du dv = a_k \iint_{\substack{\frac{u}{v} \leq x \\ v > 0}} v^{k-1} e^{-\frac{1}{2}(\frac{u^2}{k} + v^2)} du dv, \quad (2.22)$$

где

$$a_k = \frac{1}{\sqrt{2\pi k}} \frac{1}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})}.$$

Сделаем замену переменных $u = yz$, $v = z$. Так как якобиан перехода J равен

$$J = \frac{\partial(u, v)}{\partial(y, z)} = z,$$

будем иметь

$$(2.22) = a_k \int_{-\infty}^x dy \int_0^{\infty} z^k e^{-\frac{z^2}{2}(\frac{y^2}{k} + 1)} dz. \quad (2.23)$$

После замены

$$w = z \sqrt{\frac{y^2}{k} + 1}$$

получаем

$$(2.23) = a_k \int_{-\infty}^x \left(\frac{y^2}{k} + 1\right)^{-\frac{k+1}{2}} dy \int_0^{\infty} w^k e^{-\frac{w^2}{2}} dw = a_k 2^{\frac{k-1}{2}} \Gamma\left(\frac{k+1}{2}\right) \int_{-\infty}^x \left(\frac{y^2}{k} + 1\right)^{-\frac{k+1}{2}} dy.$$

Поэтому плотность $p_{\tau_k}(x)$ равна

$$p_{\tau_k}(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \quad -\infty < x < \infty. \quad (2.24)$$

Случайная величина с плотностью (2.24) называется распределенной по Стьюденту с k степенями свободы.

Так как $\Gamma(1+x) \sim \sqrt{2\pi}x^x e^{-x}$ то нетрудно убедиться, что для каждого $x \in \mathbb{R}$

$$p_{\tau_k}(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ при } k \rightarrow \infty.$$

Распределение Стьюдента ввел англичанин Госсет, работавший на фирме, занимавшейся пивоварением.

Ф-распределение

Пусть $\xi_1, \dots, \xi_p, \eta_1, \dots, \eta_q$ – независимые $(0, 1)$ -нормально распределенные случайные величины. Введем случайную величину

$$F_{pq} = \frac{\frac{1}{p} \sum_{\alpha=1}^p \xi_\alpha^2}{\frac{1}{q} \sum_{\beta=1}^q \eta_\beta^2}.$$

Имеем

$$\mathbf{P}\left\{\frac{p}{q}F_{pq} \leq x\right\} = \frac{1}{2^{\frac{p+q}{2}} \Gamma(\frac{p}{2}) \Gamma(\frac{q}{2})} \int\int_{\substack{u \leq x \\ v \geq 0}} u^{\frac{p}{2}-1} v^{\frac{q}{2}-1} e^{-\frac{u+v}{2}} dudv. \quad (2.25)$$

Положим

$$a_{pq} = \frac{1}{2^{\frac{p+q}{2}} \Gamma(\frac{p}{2}) \Gamma(\frac{q}{2})}.$$

Сделаем замену переменных $u = yz, v = z$. Тогда

$$(2.25) = a_{pq} \int_0^x y^{\frac{p}{2}-1} dy \int_0^\infty z^{\frac{p+q}{2}-1} e^{-\frac{z(1+y)}{2}} dz = a_{pq} 2^{\frac{p+q}{2}} \Gamma(\frac{p+q}{2}) \int_0^x \frac{y^{\frac{p}{2}-1}}{(1+y)^{\frac{p+q}{2}}} dy,$$

полагая $u = \frac{z(1+y)}{2}$. Поэтому

$$\mathbf{P}\left\{\frac{p}{q}F_{pq} \leq x\right\} = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2}) \Gamma(\frac{q}{2})} \int_0^x \frac{y^{\frac{p}{2}-1}}{(1+y)^{\frac{p+q}{2}}} dy.$$

Плотность распределения F_{pq} имеет вид

$$p_{F_{pq}}(x) = p^{\frac{p}{2}} q^{\frac{q}{2}} \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2}) \Gamma(\frac{q}{2})} \frac{x^{\frac{p}{2}-1}}{(q+px)^{\frac{p+q}{2}}}, x \geq 0.$$

Распределения с такой плотностью называются F -распределениями Фишера.

2.8. Закон больших чисел

О сходимости случайных величин

Пусть $(\Omega, \mathcal{U}, \mathbf{P})$ есть вероятностное пространство, $\zeta_n, n = 1, 2, \dots$, и ζ – случайная величина.

Определение 39. Последовательность случайных величин $\{\zeta_n\}$ сходится по вероятности к ζ ($\zeta_n \xrightarrow{P} \zeta$), если для любого $\varepsilon > 0$

$$\mathbf{P}\{|\zeta_n - \zeta| > \varepsilon\} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Пример 35. Пусть μ_n – число успехов в серии n независимых испытаний с вероятностью успеха p в одном испытании. Согласно теореме Бернулли $\frac{\mu_n}{n} \xrightarrow{P} p$.

Определение 40. Последовательность случайных величин $\{\zeta_n\}$ сходится почти наверное к ζ , пишут $\zeta_n \xrightarrow{\text{п.н.}} \zeta$, если $\zeta_n(\omega) \rightarrow \zeta(\omega)$ на $A \in \mathcal{U}$ и $\mathbf{P}(\Omega \setminus A) = 0$.

Теорема 2.39. Из сходимости $\{\zeta_n\}$ к ζ п.н. следует сходимость по вероятности.

Доказательство. Пусть $\zeta_n(\omega) \rightarrow \zeta(\omega)$ на $A \in \mathcal{U}$. Для любого положительного ε имеет место включение

$$A \subset \bigcup_n \bigcap_{k \geq n} \{|\zeta_k - \zeta| \leq \varepsilon\}.$$

Поэтому верно соотношение

$$\Omega \setminus A \supset \bigcap_n \bigcup_{k \geq n} \{|\zeta_k - \zeta| > \varepsilon\}.$$

Так как

$$0 = \mathbf{P}(\Omega \setminus A) \geq \lim_{n \rightarrow \infty} \mathbf{P}\left\{\bigcup_{k \geq n} \{|\zeta_k - \zeta| > \varepsilon\}\right\},$$

из свойства монотонности меры следует, что

$$\lim_{n \rightarrow \infty} \mathbf{P}\{|\zeta_n - \zeta| > \varepsilon\} = 0. \quad \square$$

Обратное утверждение неверно. Но из сходящейся по вероятности последовательности случайных величин можно выбрать подпоследовательность, сходящуюся почти наверное.

Определение 41. Последовательность случайных величин $\{\zeta_n\}$ сходится к случайной величине ζ по распределению, или слабо сходится, если $F_{\zeta_n}(x) \rightarrow F_{\zeta}(x)$ в каждой точке непрерывности предельной функции распределения. Слабая сходимость F_{ζ_n} к F_{ζ} обозначается через $F_{\zeta_n} \Rightarrow F_{\zeta}$.

Теорема 2.40. Из сходимости $\{\zeta_n\}$ к ζ по вероятности следует сходимость по распределению.

Доказательство. Введем событие $A_n = \{|\zeta_n - \zeta| \leq \varepsilon\}$. Так как при $\omega \in A_n$

$$\zeta(\omega) - \varepsilon \leq \zeta_n(\omega) \leq \zeta(\omega) + \varepsilon,$$

то при каждом x имеем

$$\{\zeta_n \leq x\} \subseteq \{\zeta \leq x + \varepsilon\} \cup \bar{A}_n$$

и

$$\{\zeta \leq x - \varepsilon\} \subseteq \{\zeta_n \leq x\} \cup \bar{A}_n$$

Отсюда следует, что

$$\mathbf{P}\{\zeta \leq x - \varepsilon\} - \mathbf{P}(\bar{A}_n) \leq \mathbf{P}\{\zeta_n \leq x\} \leq \mathbf{P}\{\zeta \leq x + \varepsilon\} + \mathbf{P}(\bar{A}_n).$$

Поэтому

$$\mathbf{P}\{\zeta \leq x - \varepsilon\} \leq \liminf_{n \rightarrow \infty} \mathbf{P}\{\zeta_n \leq x\} \leq \limsup_{n \rightarrow \infty} \mathbf{P}\{\zeta_n \leq x\} \leq \mathbf{P}\{\zeta \leq x + \varepsilon\} \quad (2.26)$$

Если x есть точка непрерывности $F_{\zeta}(x)$, то из (2.26) получаем

$$\lim_{n \rightarrow \infty} F_{\zeta_n}(x) = F_{\zeta}(x). \quad \square$$

Из сходимости по распределению, вообще говоря, не следует сходимость по вероятности. Но если $F_{\zeta_n}(x)$ слабо сходится к вырожденному распределению, то имеет место обратное утверждение.

Теорема 2.41. Если $F_{\zeta_n} \Rightarrow F_\zeta$ и F_ζ вырождено в точке c , то есть

$$F_\zeta(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c, \end{cases}$$

то $\zeta_n \xrightarrow{P} c$.

Доказательство. Так как $F_{\zeta_n}(c + \varepsilon) \rightarrow 1$ и $F_{\zeta_n}(c - \varepsilon) \rightarrow 0$, то $\mathbf{P}\{c - \varepsilon < \zeta_n \leq c + \varepsilon\} \rightarrow 1$, то есть $\mathbf{P}\{|\zeta_n - c| > \varepsilon\} \rightarrow 0$. \square

Следующая теорема нам потребуется при изучении математической статистики.

Теорема 2.42. Пусть $\zeta_n = \xi_n + \eta_n$, $n = 1, 2, \dots$ – случайные величины. Если $\eta_n \xrightarrow{P} 0$ и последовательность функций распределения $F_{\xi_n}(x)$ слабо сходится к функции распределения $F(x)$, то последовательность функций распределения $F_{\zeta_n}(x)$ также слабо сходится к $F(x)$.

Доказательство. Пусть x – точка непрерывности $F(x)$ и $\varepsilon > 0$. Положим

$$A_n = \{\xi_n + \eta_n < x\}, \quad B_n = \{|\eta_n| < \varepsilon\}.$$

Учитывая, что

$$A_n = A_n B_n + A_n \bar{B}_n \subset A_n B_n + \bar{B}_n.$$

получаем

$$\mathbf{P}(A_n B_n) \leq \mathbf{P}(A_n) \leq \mathbf{P}(A_n B_n) + \mathbf{P}(\bar{B}_n).$$

Отсюда и из соотношения

$$\{(\xi_n < x - \varepsilon) \cap B_n\} \subset A_n B_n \subset \{\xi_n < x + \varepsilon\}.$$

находим

$$\mathbf{P}\{(\xi_n < x - \varepsilon) \cap B_n\} \leq \mathbf{P}(A_n) \leq \mathbf{P}\{\xi_n < x + \varepsilon\} + \mathbf{P}(\bar{B}_n). \quad (2.27)$$

Из (2.27) и неравенства

$$\mathbf{P}\{(\xi_n < x - \varepsilon) \cap B_n\} \geq \mathbf{P}\{\xi_n < x - \varepsilon\} - \mathbf{P}(\bar{B}_n)$$

следует, что

$$F_{\xi_n}(x - \varepsilon) - \mathbf{P}(\bar{B}_n) \leq \mathbf{P}(A_n) \leq F_n(x + \varepsilon) + \mathbf{P}(\bar{B}_n).$$

По условию теоремы $\mathbf{P}(\bar{B}_n) \rightarrow 0$ при $n \rightarrow \infty$, поэтому

$$F(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} \mathbf{P}\{\zeta_n \leq x\} \leq \limsup_{n \rightarrow \infty} \mathbf{P}\{\zeta_n \leq x\} \leq F(x + \varepsilon).$$

Отсюда при $\varepsilon \rightarrow 0$ получаем утверждение теоремы, так как x – точка непрерывности функции $F(x)$. \square

Закон больших чисел в форме Хинчина

Теорема 2.43. Пусть $\{\xi_n\}_{n=1}^\infty$ – последовательность независимых одинаково распределенных случайных величин, имеющих конечное математическое ожидание $M\xi_n = a$. Тогда

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{P} a \text{ при } n \rightarrow \infty.$$

Доказательство. Через $f(t)$ обозначим характеристическую функцию случайной величины ξ_1 , а через S_n – сумму $\xi_1 + \dots + \xi_n$. Достаточно доказать, что для каждого t

$$f_{\frac{S_n}{n}}(t) \rightarrow e^{iat}.$$

В некоторой окрестности нуля выполняется неравенство

$$|f(t) - 1| < \frac{1}{2}.$$

Для таких t определена функция $\ell(t) = \log f(t)$. Поэтому для любого t и достаточно больших n имеет место представление

$$f_{\frac{S_n}{n}}(t) = f^n\left(\frac{t}{n}\right) = e^{\ell\left(\frac{t}{n}\right)n}.$$

Так как

$$\ell'(0) = \frac{f'(0)}{f(0)} = ia, \quad \ell(0) = 0,$$

то

$$e^{\ell\left(\frac{t}{n}\right)n} = \exp\left\{t \frac{\ell\left(\frac{t}{n}\right) - \ell(0)}{\frac{t}{n}}\right\} \rightarrow e^{\ell'(0)t} = e^{ita}. \quad \square$$

Усиленный закон больших чисел

Сформулируем здесь две теоремы. Первая теорема об усиленном законе больших чисел для независимых равно распределенных случайных величин.

Теорема 2.44. Пусть ξ_1, ξ_2, \dots независимы, $M\xi_k = 0$ и $D\xi_k = \sigma_k^2$. Если

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{n^2} \leq \infty,$$

тогда

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{n.н.} 0.$$

Доказательство этой теоремы опирается на неравенство Колмогорова, усиливающего неравенство Чебышева.

Теорема 2.45 (неравенство Колмогорова). Пусть ξ_1, ξ_2, \dots независимы и имеют конечные математическое ожидание и дисперсию. Тогда

$$\mathbf{P}\left\{\max_{1 \leq k \leq n} |\zeta_k - M\zeta_k| \geq x\right\} \leq \frac{D\zeta_n}{x^2},$$

где $\zeta_k = \xi_1 + \dots + \xi_k$.

Для независимых одинаково распределенных случайных величин справедлив критерий сходимости почти всюду.

Теорема 2.46 (усиленный закон больших чисел Колмогорова). Пусть ξ_1, ξ_2, \dots независимы и одинаково распределены. Для того, чтобы

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{n.н.} a.$$

необходимо и достаточно, чтобы существовало конечное $M\xi_n = a$.

Следствие. В схеме Бернулли для числа успехов имеет место не только закон больших чисел

$$\frac{\mu_n}{n} \xrightarrow{P} p,$$

но и усиленный закон больших чисел

$$\frac{\mu_n}{n} \xrightarrow{n.н.} p.$$

Следствие вытекает из критерия Колмогорова, так как $\mu_n = \xi_1 + \dots + \xi_n$, где ξ_1, \dots, ξ_n независимы и $\mathbf{P}\{\xi_k = 1\} = p$, $\mathbf{P}\{\xi_k = 0\} = 1 - p$.

Глава 3

Математическая статистика

3.1. Предмет математической статистики

Теория вероятностей изучает математические модели случайных явлений. Математическая статистика решает обратные задачи: разрабатывает различные методы, которые позволяют по статистическим данным, которые носят случайный характер, подобрать подходящую теоретико – вероятностную модель.

Множество объектов, подлежащих контролю называется генеральной совокупностью. Множество отобранных объектов называется выборкой. Число элементов выборки называется объемом выборки. Повторная выборка — это выборка с возвращением. Бесповторная выборка — выборка без возвращения. Выборка должна быть представительной и случайной. Объекты генеральной совокупности описываются одним или несколькими числовыми параметрами. Допустим, что объекты генеральной совокупности описываются случайной величиной ξ , принимающей на каждом объекте некоторое числовое значение.

Сделав выборку объема n , мы определяем ряд значений ξ . По этой последовательности нам следует получить представление о функции распределения, оценить математическое ожидание, дисперсию и, возможно, другие параметры случайной величины. Наблюдаемые в эксперименте значения x_k случайной величины ξ носят случайный характер. Эти x_k принимают те же значения как и ξ и распределены как ξ . На них следует смотреть как на n случайных величин, распределенных как ξ , в частности, $Mx_k = M\xi = m$ и $Dx_k = D\xi = \sigma^2$.

Случайный вектор (x_1, \dots, x_n) , где x_k независимые и одинаково распределенные случайные величины с функцией распределения $\mathbf{P}\{x_k \leq x\} = F_\xi(x) = F(x)$, является математической моделью независимых измерений, проводимых в равных условиях. Случайная величина ξ является подлежащей определению случайной величиной.

Мы будем рассматривать только повторные выборки. Если выборка составляет незначительную часть генеральной совокупности, то различие между повторной и бесповторной выборкой несущественно, а в предельном случае это различие исчезает.

3.2. Эмпирическая функция распределения

Эмпирическая функция распределения F_n^* определяется соотношением

$$F_n^*(x) = \frac{\mu_n(x)}{n},$$

где μ_n — число x_k , не превосходящих x . Заметим, что F_n^* является случайной величиной.

Теорема 3.1. Для каждого x и любого $\varepsilon > 0$ выполняется равенство

$$\lim_{n \rightarrow \infty} \mathbf{P}\{|F_n^*(x) - F(x)| < \varepsilon\} = 1.$$

Доказательство. Назовем случайное событие $\{x_k \leq x\}$ успехом. Тогда $F(x) = \mathbf{P}\{x_k \leq x\} = p$ — вероятность успеха, μ_n — число успехов и $F_n^*(x)$ — относительная частота успехов. Из теоремы Бернулли следует утверждение теоремы. \square

Разобьем точками $-\infty = z_0 < z_1 < \dots < z_m < z_{m+1} = \infty$ числовую ось на промежутки $\bigcup_{k=0}^m [z_k, z_{k+1})$. Тогда относительная частота попадания x_i в промежуток $(z_k, z_{k+1}]$ равна $p_k^* = F_n^*(z_{k+1}) - F_n^*(z_k)$. При большом объеме выборки p_k^* близки к вероятностям

$$p_k = F(z_{k+1}) - F(z_k) = \int_{z_k}^{z_{k+1}} p(x) dx,$$

где $p(x)$ – плотность распределения случайной величины ξ . Над каждым промежутком $(z_k, z_{k+1}]$ строится прямоугольник, площадь которого пропорциональна p_k^* . Построенная фигура называется гистограммой. Гистограмма дает представление о плотности распределения $p(x)$, но чувствительна к разбиению.

Оценка неизвестной функции распределения может быть получена на основании теоремы Колмогорова.

Теорема 3.2. Если $F(x)$ непрерывна, то при $n \rightarrow \infty$

$$\begin{aligned} \mathbf{P}\{\sqrt{n}(\sup_{-\infty < x < \infty} |F_n^*(x) - F(x)|) < z\} &\rightarrow K(z) = \\ &= \begin{cases} 0, & \text{при } z \leq 0 \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2}, & \text{при } z > 0 \end{cases} \end{aligned}$$

Если \hat{z}_α – решение уравнения $1 - K(z) = \alpha$, то с приближительной вероятностью $1 - \alpha$ для всех x выполняется неравенство

$$F_n^*(x) - \frac{\hat{z}_\alpha}{\sqrt{n}} < F(x) < F_n^*(x) + \frac{\hat{z}_\alpha}{\sqrt{n}}.$$

Следующая теорема утверждает, что с вероятностью 1 эмпирическая функция распределения F_n^* равномерно сходится к F .

Теорема 3.3 (Гливенко). Пусть $F(x)$ функция распределения случайной величины ξ и $F_n^*(x)$ эмпирическая функция распределения результатов n независимых наблюдений над случайной величиной ξ . Тогда при $n \rightarrow \infty$

$$\mathbf{P}\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)| = 0\} = 1.$$

3.3. Выборочный метод

Выборочное среднее определяется как среднее арифметическое элементов выборки

$$\bar{x} = m_1^* = \frac{1}{n} \sum_{k=1}^n x_k,$$

выборочная дисперсия вводится следующим образом:

$$s^2 = \alpha_2^* = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Выборочные моменты порядка j определяются соотношениями

$$m_j^* = \frac{1}{n} \sum_{k=1}^n x_k^j \quad \text{и} \quad \alpha_j^* = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^j,$$

Выборочные моменты и центральные выборочные моменты являются случайными величинами.

Теоретические моменты $M\xi^k$ порядка k будем обозначать через m_k , а центральный момент $M(\xi - M\xi)^k$ порядка k будем обозначать через α_k . В частности, α_2 есть дисперсия, обозначаемая также через σ^2 .

Теорема 3.4. *Выборочные моменты сходятся по вероятности к соответствующим теоретическим моментам при $n \rightarrow \infty$.*

Доказательство. Положим

$$\xi_1 = x_1^k, \dots, \xi_n = x_n^k.$$

Случайные величины ξ_i независимы и одинаково распределены, $M\xi_i = M\xi^k = m_k$. Согласно закону больших чисел

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - M\xi_i \right| > \varepsilon \right\} = 0.$$

Из соотношений

$$m_k^* = \frac{\xi_1 + \dots + \xi_n}{n}, \quad M\xi_i = m_k,$$

следует, что

$$m_k^* \xrightarrow{P} m_k. \quad \square$$

Теорема 3.5. *Выборочные центральные моменты сходятся по вероятности к соответствующим теоретическим моментам.*

Предварительно докажем лемму.

Лемма 3.1. *Пусть $f(t) = f(t_1, \dots, t_k)$ – непрерывная функция по совокупности переменных в точке (m_1, \dots, m_k) и последовательность случайных векторов $(\xi_1^{(n)}, \dots, \xi_k^{(n)})$ сходится при $n \rightarrow \infty$ по вероятности к вектору (m_1, \dots, m_k) . Тогда последовательность случайных величин $f(\xi_1^{(n)}, \dots, \xi_k^{(n)})$ сходится по вероятности к $f(m_1, \dots, m_k)$.*

Доказательство. По определению непрерывности

$$(\forall \varepsilon > 0)(\exists \delta > 0)(|t_i - m_i| < \delta, \quad i = \overline{1, k} \Rightarrow |f(t_1, \dots, t_k) - f(m_1, \dots, m_k)| < \varepsilon).$$

Положим

$$A_i^{(n)} = \{|\xi_i^{(n)} - m_i| < \delta\} \quad \text{и} \quad B^{(n)} = \{|f(\xi_1^{(n)}, \dots, \xi_k^{(n)}) - f(m_1, \dots, m_k)| < \varepsilon\}.$$

Тогда

$$\bigcap_{i=1}^k A_i^{(n)} \subset B^{(n)} \Rightarrow \overline{B^{(n)}} \subset \bigcup_{i=1}^k \overline{A_i^{(n)}}.$$

Отсюда следует, что

$$\mathbf{P}\{|f(\xi_1^{(n)}, \dots, \xi_k^{(n)}) - f(m_1, \dots, m_k)| > \varepsilon\} \leq \sum_{i=1}^k \mathbf{P}\{|\xi_i^{(n)} - m_i| \geq \delta\} \rightarrow 0. \quad \square$$

Доказательство теоремы. Введем функцию

$$f(y) = f(y_0, \dots, y_k) = \sum_{j=0}^k (-1)^{k-j} C_k^j y_1^{k-j} y_j.$$

Заметим, что

$$\begin{aligned} \alpha_k &= M(\xi - M\xi)^k = M\left(\sum_{j=0}^k (-1)^{k-j} C_k^j (M\xi)^{k-j} \xi^j\right) = \\ &= \sum_{j=0}^k (-1)^{k-j} C_k^j (M\xi)^{k-j} M\xi^j = \sum_{j=0}^k (-1)^{k-j} C_k^j m_1^{k-j} m_j = f(1, m_1, \dots, m_k) \end{aligned}$$

и

$$\begin{aligned}\alpha_k^* &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k (-1)^{k-j} C_k^j x_i^j \bar{x}^{k-j} = \\ &= \sum_{j=1}^k (-1)^{k-j} C_k^j \bar{x}^{k-j} \left(\frac{1}{n} \sum_{i=1}^n x_i^j \right) = \sum_{j=1}^k (-1)^{k-j} C_k^j (m_1^*)^{k-j} m_j^* = f(1, m_1^*, \dots, m_k^*)\end{aligned}$$

Так как $m_i^* \xrightarrow{P} m_i$, из леммы следует, что

$$f(1, m_1^*, \dots, m_k^*) \xrightarrow{P} f(1, m_1, \dots, m_k), \quad \text{то есть } \alpha_k^* \xrightarrow{P} \alpha_k. \quad \square$$

3.4. Понятие оценки

Пусть \mathcal{F} – множество распределений и θ – числовая характеристика, однозначно определяемая по распределению. В таком случае θ называется параметром распределения. Например, $\theta = \int_{\mathbf{R}} x dF(x)$. Соответствие не является взаимно однозначным. Задача оценки параметра состоит в нахождении функции $\theta^* = \theta(x_1, \dots, x_n)$ от элементов выборки, которая в каком-либо смысле близка к параметру $\theta = \theta(F_\xi)$. Так как x_k – независимые случайные величины, то θ^* – случайная величина, определяемая распределением F_ξ . Любая функция от элементов выборки называется статистикой. Таким образом, оценка параметра является статистикой. Близость статистики к оцениваемому параметру обеспечивается условиями несмещенности, состоятельности и эффективности.

1. Несмещенность означает отсутствие систематической ошибки, то есть

$$M\theta^* = \theta.$$

В качестве примера рассмотрим выборочное среднее и выборочную дисперсию.

Теорема 3.6. Пусть (x_1, \dots, x_n) – выборка из распределения $F_\xi(x)$ с математическим ожиданием m и дисперсией σ^2 . Тогда

$$M\bar{x} = m, \quad D\bar{x} = \frac{\sigma^2}{n}, \quad Ms^2 = \frac{n-1}{n}\sigma^2$$

и

$$Ds^2 = \frac{\alpha_4 - \alpha_2^2}{n} - 2\frac{\alpha_4 - 2\alpha_2^2}{n^2} + \frac{\alpha_4 - 3\alpha_2^2}{n^3} = O\left(\frac{1}{n}\right),$$

если конечен момент α_4 .

Доказательство. Первые два утверждения легко проверяются. Остановимся на доказательстве двух последних равенств. Имеем

$$\begin{aligned}s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m + m - \bar{x})^2 = \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 + (\bar{x} - m)^2 - \frac{2}{n} (\bar{x} - m) \sum_{i=1}^n (x_i - m) = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 - (\bar{x} - m)^2.\end{aligned}$$

Поэтому

$$Ms^2 = \frac{1}{n} \sum_{i=1}^n M(x_i - m)^2 - M(\bar{x} - m)^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Введем $y_k = x_k - m$. Тогда

$$\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k = \frac{1}{n} \sum_{k=1}^n x_k - m.$$

Поэтому

$$s^2 = \alpha_2^* = \frac{1}{n} \sum_{k=1}^n (y_k - \bar{y})^2 = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2$$

и

$$(\alpha_2^*)^2 = \frac{1}{n^2} \left(\sum_{k=1}^n y_k^2 \right)^2 - \frac{2}{n} \bar{y}^2 \left(\sum_{k=1}^n y_k^2 \right) + \bar{y}^4.$$

Заметим, что $M y_k = 0$ и y_k независимы. Поэтому в сумме

$$M \bar{y}^4 = \frac{1}{n^4} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=1}^n M(y_{\ell_1} y_{\ell_2} y_{\ell_3} y_{\ell_4})$$

отличны от нуля только n слагаемых вида $M y_\ell^4 = \alpha_4$ и $C_n^2 C_4^2 = 3n(n-1)$ слагаемых вида $M y_\ell^2 y_k^2 = \alpha_2^2$, $k \neq \ell$. В результате, имеем $M \bar{y}^4 = (\alpha_4 + 3(n-1)\alpha_2^2)/n^3$. Далее,

$$M \left[\left(\sum_{k=1}^n y_k^2 \right)^2 / n \right] = M \left(\frac{1}{n} \left[\sum_{k=1}^n y_k^4 + 2 \sum_{k \neq \ell} y_k^2 y_\ell^2 \right] \right) = \alpha_4 + 2 \frac{1}{n} \frac{n(n-1)}{2} \alpha_2^2 = \alpha_4 + (n-1)\alpha_2^2.$$

Представим \bar{y}^2 в виде

$$\bar{y}^2 = \frac{1}{n^2} \left[\sum_{k=1}^n y_k^2 + 2 \sum_{k \neq \ell} y_k y_\ell \right].$$

Тогда

$$\frac{1}{n} \bar{y}^2 \left(\sum_{k=1}^n y_k^2 \right) = \frac{1}{n^3} \left(\sum_{k=1}^n y_k^2 \right)^2 + 2 \frac{1}{n^3} \left(\sum_{k \neq \ell} y_k y_\ell \right) \left(\sum_{k=1}^n y_k^2 \right)$$

Поэтому

$$n^2 M \left[\frac{1}{n} \bar{y}^2 \left(\sum_{k=1}^n y_k^2 \right) \right] = M \left[\frac{1}{n} \left(\sum_{k=1}^n y_k^2 \right)^2 \right] = \alpha_4 + (n-1)\alpha_2^2,$$

так как $M(y_k y_\ell^3) = M(y_k y_\ell y_r^2) = 0$ при $k \neq \ell \neq r$. Легко теперь убедиться, что,

$$D s^2 = D \alpha_2^* = M(\alpha_2^*)^2 - (M \alpha_2^*)^2 = \frac{\alpha_4 - \alpha_2^2}{n} - 2 \frac{\alpha_4 - 2\alpha_2^2}{n^2} + \frac{\alpha_4 - 3\alpha_2^2}{n^3}. \quad \square$$

Согласно доказанной теореме выборочное среднее являются несмещенной оценкой математического ожидания, а выборочная дисперсия является смещенной оценкой. Несмещенной оценкой дисперсии является статистика

$$\hat{s}^2 = \frac{n}{n-1} s^2,$$

так называемая исправленная выборочная дисперсия.

В общем случае, можно показать, что

$$M \alpha_k^* = \alpha_k + O\left(\frac{1}{n}\right) \text{ и } M(\alpha_k^* - \alpha_k)^2 = O\left(\frac{1}{n}\right).$$

Вторая из этих формул верна, если конечен момент α_{2k} . Отсюда следует, что

$$D \alpha_k^* = O\left(\frac{1}{n}\right).$$

2. Состоятельность оценки. Естественно потребовать, чтобы оценка θ_n^* при $n \rightarrow \infty$ приближалась к оцениваемому параметру. Оценка θ_n^* называется состоятельной, если

$$\theta_n^* \xrightarrow{P} \theta.$$

Теорема 3.7. Если при $n \rightarrow \infty$ $M\theta_n^* \rightarrow \theta$ и $D\theta_n^* \rightarrow 0$, тогда оценка θ_n^* состоятельна.

Доказательство. По неравенству Чебышева

$$\mathbf{P}\{|\theta_n^* - M\theta_n^*| > \varepsilon\} \leq \frac{D\theta_n^*}{\varepsilon^2} \rightarrow 0.$$

Воспользуемся неравенством $|\theta_n^* - \theta| \leq |\theta_n^* - M\theta_n^*| + |M\theta_n^* - \theta|$. Так как

$$(\forall \varepsilon > 0)(\exists n_0)(\forall n \geq n_0 \Rightarrow |M\theta_n^* - \theta| \leq \varepsilon/2),$$

из $|\theta_n^* - \theta| > \varepsilon$ следует, что $|\theta_n^* - M\theta_n^*| > \varepsilon/2$. Поэтому

$$\mathbf{P}\{|\theta_n^* - \theta| > \varepsilon\} \leq \mathbf{P}\{|\theta_n^* - M\theta_n^*| > \varepsilon/2\} \rightarrow 0,$$

то есть $\theta_n^* \xrightarrow{P} \theta$. \square

Согласно теоремам 3.4 и 3.5 имеет место следующее утверждение:

Теорема 3.8. Начальные и центральные моменты t_k^* и α_k^* , $k = 1, 2, \dots$, являются состоятельными оценками начальных и центральных моментов t_k и α_k .

3. Эффективность. Мерой рассеяния является дисперсия. Естественно вопрос о нахождении оценок с наименьшей дисперсией.

Пусть имеется семейство абсолютно непрерывных одномерных распределений

$$\mathcal{F} = \{p(t, \theta)\},$$

зависящих от одного параметра θ и пусть x_1, \dots, x_n – выборка из распределения с плотностью $p(t, \theta)$ с неизвестным параметром θ .

Введем совместную плотность распределения

$$f(u, \theta) = \prod_{i=1}^n p(u_i, \theta), \quad u = (u_1, \dots, u_n).$$

Ясно, что

$$\int_{\mathbf{R}^n} f(u, \theta) du = 1. \quad (*)$$

Пусть $\theta_n^* = \varphi(x_1, \dots, x_n)$ – оценка параметра θ . Положим

$$g(\theta) \stackrel{\text{def}}{=} M\theta_n^* = \int_{\mathbf{R}^n} \varphi(u) f(u, \theta) du. \quad (**)$$

Определение 42. Информацией Фишера семейства плотностей $\{f(u, \theta)\}$ называется величина

$$J(\theta) = \int_{\mathbf{R}^n} \left(\frac{\partial \ln f(u, \theta)}{\partial \theta} \right)^2 f(u, \theta) du.$$

Теорема 3.9 (неравенство Рао – Крамера). Если семейство плотностей $\{f(u, \theta)\}$, $u = (u_1, \dots, u_n)$, и оценка $\theta_n^* = \varphi(x_1, \dots, x_n)$ таковы, что интегралы в равенствах (*) и (**) можно дифференцировать по параметру θ , то имеет место неравенство

$$D\theta_n^* \geq \frac{[g'(\theta)]^2}{J(\theta)}.$$

Знак равенства при фиксированном θ достигается тогда и только тогда, когда существует число K , не зависящее от u_1, \dots, u_n , но может быть зависящее от θ такое, что

$$\frac{\partial \ln f}{\partial \theta} = K(\varphi - g(\theta)).$$

Доказательство. Согласно предположениям теоремы

$$\int_{\mathbf{R}^n} \frac{\partial f(u, \theta)}{\partial \theta} du = 0 \quad \text{и} \quad g'(\theta) = \int_{\mathbf{R}^n} \varphi(u) \frac{\partial f(u, \theta)}{\partial \theta} du.$$

Отсюда следует, что

$$\int_{\mathbf{R}^n} f(u, \theta) \frac{\partial \ln f(u, \theta)}{\partial \theta} du = 0 \quad \text{и} \quad g'(\theta) = \int_{\mathbf{R}^n} \varphi(u) f(u, \theta) \frac{\partial \ln f(u, \theta)}{\partial \theta} du.$$

Умножим первое равенство на $g(\theta)$ и вычтем результат из второго равенства

$$g'(\theta) = \int_{\mathbf{R}^n} [\varphi(u) - g(\theta)] f(u, \theta) \frac{\partial \ln f(u, \theta)}{\partial \theta} du.$$

Положим

$$\varphi_1 = (\varphi - g)\sqrt{f} \quad \text{и} \quad \varphi_2 = \frac{\partial \ln f}{\partial \theta} \sqrt{f}.$$

С помощью неравенства Коши–Буняковского получим неравенство Рао–Крамера

$$[g'(\theta)]^2 = \left(\int_{\mathbf{R}^n} (\varphi - g) \frac{\partial \ln f}{\partial \theta} f du \right)^2 \leq \int_{\mathbf{R}^n} \left(\frac{\partial \ln f}{\partial \theta} \right)^2 f du \int_{\mathbf{R}^n} (\varphi - g)^2 f du = J(\theta) D\theta_n^*.$$

Знак равенства в неравенстве Коши–Буняковского достигается тогда и только тогда, когда φ_1 и φ_2 пропорциональны

$$\frac{\partial \ln f}{\partial \theta} \sqrt{f} = K(\varphi - g(\theta)) \sqrt{f}. \quad \square$$

Замечания.

1. Информация Фишера есть не что иное как

$$J(\theta) = M\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2, \quad \text{здесь } x = (x_1, \dots, x_n) \text{ — случайная выборка.}$$

2. Так как

$$M\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right) = \int_{\mathbf{R}^n} \frac{\partial \ln f}{\partial \theta} f du = 0,$$

то

$$J(\theta) = D\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right).$$

3. Так как

$$f(x, \theta) = \prod_{i=1}^n p(x_i, \theta)$$

и случайные величины

$$\frac{\partial \ln p(x_i, \theta)}{\partial \theta}$$

независимы и одинаково распределены, поэтому

$$J(\theta) = D\left(\sum_{i=1}^n \frac{\partial \ln p(x_i, \theta)}{\partial \theta}\right) = \sum_{i=1}^n D\left(\frac{\partial \ln p(x_i, \theta)}{\partial \theta}\right) = nJ_1(\theta),$$

где

$$J_1(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln p(t, \theta)}{\partial \theta}\right)^2 p(t, \theta) dt$$

есть информация Фишера одного наблюдения.

Замечания. Для дискретных случайных величин интеграл в информации Фишера заменяется суммой.

Отношение

$$e(\theta_n^*) = \frac{[g'(\theta)]^2}{D(\theta_n^*)J(\theta)}$$

называется эффективностью оценки. Оценка эффективна в рассматриваемых предположениях, когда $e(\theta_n^*) = 1$. Если $e(\theta_n^*) < 1$ и $e(\theta_n^*) \rightarrow 1$, то говорят об асимптотической эффективности. Если дисперсия оценки ведет себя как $o(1/n)$, то оценка называется "сверхэффективной".

Пример 1. Пусть (x_1, \dots, x_n) – выборка из нормально распределенной генеральной совокупности с известной дисперсией σ^2 и с оцениваемым математическим ожиданием m ,

$$p(u, m) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-m)^2}{2\sigma^2}}$$

$$J_1(m) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln p(u, m)}{\partial m}\right)^2 p(u, m) du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{(u-m)^2}{\sigma^4} e^{-\frac{(u-m)^2}{2\sigma^2}} du = \frac{1}{\sigma^2}.$$

Для несмещенной оценки математического ожидания неравенство Рао – Крамера принимает вид

$$Dm^* \geq \frac{\sigma^2}{n}.$$

Если в качестве оценки m^* взять выборочное среднее $m^* = \bar{x}$, то $Mm^* = m$, и $Dm^* = \frac{\sigma^2}{n}$. Поэтому оценка $m^* = \bar{x}$ является эффективной.

Пример 2. Пусть (x_1, \dots, x_n) – выборка из нормально распределенной генеральной совокупности с известным математическим ожиданием m и с оцениваемой дисперсией θ , имеющая плотность

$$p(u, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(u-m)^2}{2\theta}}$$

$$J_1(\theta) = \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{\infty} \left(\frac{(u-m)^2}{2\theta^2} - \frac{1}{2\theta}\right)^2 e^{-\frac{(u-m)^2}{2\theta}} du = \frac{1}{2\theta^2}.$$

Поэтому для несмещенной оценки имеет место неравенство

$$D\theta^* \geq \frac{2\theta^2}{n}.$$

Рассмотрим три оценки

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \widehat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2.$$

Для нормально распределенной случайной величины имеем

$$\alpha_4 = 3\sigma^4.$$

Поэтому

$$Ms^2 = \frac{n-1}{n}\sigma^2, \quad Ds^2 = \frac{2(n-1)}{n^2}\sigma^4 = \frac{2(n-1)}{n^2}\theta^2 \quad (\text{см. теорему 3.6}).$$

Так что

$$e(s^2) = \frac{n-1}{n} < 1,$$

значит, оценка $\theta^* = s^2$ не является эффективной, но является асимптотически эффективной.

Оценка $\theta^* = \widehat{s}^2$ является несмещенной оценкой и

$$D\widehat{s}^2 = \frac{2\theta^2}{n-1}.$$

Поэтому эффективность этой оценки меньше единицы

$$e(\widehat{s}^2) = \frac{1}{D\widehat{s}^2 n J_1(\theta)} = \frac{n-1}{n} < 1.$$

Для оценки s_0^2 выполняются равенства

$$Ms_0^2 = \theta \quad \text{и} \quad Ds_0^2 = \frac{2\theta^2}{n},$$

то есть s_0^2 является несмещенной и эффективной оценкой дисперсии θ .

Пример 3. Пусть выборка (x_1, \dots, x_n) – биномиально распределена

$$k \rightarrow P_N(k) = C_N^k p^k (1-p)^{N-k}$$

с известным N и параметром $\theta = p$.

$$J_1(p) = \sum_{i=0}^N \left(\frac{\partial \ln P_N(i)}{\partial p} \right)^2 P_N(i) = \sum_{i=0}^N \left(\frac{i}{p} - \frac{N-i}{1-p} \right)^2 C_N^i p^i (1-p)^{N-i} = \frac{N}{p(1-p)}.$$

Для несмещенной оценки неравенство Рао – Крамера принимает вид

$$Dp^* \geq \frac{p(1-p)}{nN}.$$

Положим

$$p^* = \frac{\bar{x}}{N} = \frac{1}{nN} \sum_{i=1}^n x_i.$$

Тогда

$$Mp^* = p \quad \text{и} \quad Dp^* = \frac{p(1-p)}{nN},$$

то есть p^* – эффективная оценка.

Пример 4. Пусть выборка (x_1, \dots, x_n) имеет распределение Пуассона с параметром λ

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Тогда

$$J_1(\lambda) = \sum_{i=0}^{\infty} \left(\frac{\partial \ln P(i)}{\partial \lambda} \right)^2 P(i) = \sum_{i=0}^{\infty} \left(\frac{i}{\lambda} - 1 \right)^2 \frac{\lambda^i}{i!} e^{-\lambda} = \frac{1}{\lambda}.$$

Для несмещенной оценки λ^* неравенство Рао – Крамера принимает вид

$$D\lambda^* \geq \frac{\lambda}{n}.$$

Пусть $\lambda^* = \bar{x}$. Тогда $M\lambda^* = Mx_i = \lambda$ и $D\lambda^* = \frac{\lambda}{n}$. Поэтому оценка $\lambda^* = \bar{x}$ является эффективной.

3.5. Асимптотическая нормальность выборочных моментов

Определение 43. Если последовательность случайных величин ζ_n такова, что при некоторых A_n и B_n

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{\zeta_n - A_n}{B_n} \leq x \right\} = \Phi(x),$$

то говорят, что последовательность $\{\zeta_n\}$ асимптотически нормальна с параметрами (A_n, B_n) или просто (A_n, B_n) -асимптотически нормальна.

Теорема 3.10. Если конечен теоретический момент m_{2k} , то при $n \rightarrow \infty$ выборочный момент m_k^* асимптотически нормален с параметрами $(m_k, \sqrt{\frac{m_{2k} - m_k^2}{n}})$.

Доказательство. Случайная величина

$$nm_k^* = \sum_{i=1}^n x_i^k$$

является суммой независимых одинаково распределенных случайных величин. Если конечен теоретический момент $m_{2k} = Mx_i^{2k}$, то к этой сумме можно применить центральную предельную теорему. Так как $Mx_i^k = m_k$ и $Dx_i^k = m_{2k} - m_k^2$, то случайная величина

$$\frac{nm_k^* - nm_k}{\sqrt{(m_{2k} - m_k^2)n}} = \frac{m_k^* - m_k}{\sqrt{\frac{m_{2k} - m_k^2}{n}}}$$

асимптотически нормальна с параметрами $(0, 1)$. \square

Асимптотическая нормальность центральных выборочных моментов α_k^* доказывается сложнее. Ограничимся случаем $k = 2$.

Теорема 3.11. Если конечен теоретический момент α_4 , то при $n \rightarrow \infty$ выборочная дисперсия $\alpha_2^* = s^2$ асимптотически нормальна с параметрами $(\alpha_2, \sqrt{\frac{\alpha_4 - \alpha_2^2}{n}})$.

Доказательство. Положим $y_k = x_k - Mx_k$, $k = 1, \dots, n$. Перепишем формулу для s^2 в виде

$$s^2 = \frac{1}{n} \sum_{k=1}^n (y_k - \bar{y})^2 = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2,$$

где $\bar{y} = (y_1 + \dots, y_n)/n$. Случайную величину

$$\zeta_n = \sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2 - My_1^2 \right) = \sqrt{n}(s^2 - \alpha_2)$$

представим в виде $\xi_n + \eta_n$, где

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n (y_k^2 - My_k^2), \quad \eta_n = -\bar{y}^2 \sqrt{n}.$$

Напомним, что y_k^2 , $k = 1, \dots, n$ – независимые одинаково распределенные случайные величины. Так как

$$M|\eta_n| = \sqrt{n}M\bar{y}^2 = \frac{Dy_1}{\sqrt{n}} = \frac{Dx_1}{\sqrt{n}} = \frac{\alpha_2}{\sqrt{n}},$$

из неравенства Чебышева следует, что для любого $\varepsilon > 0$

$$\mathbf{P}\{|\eta_n| > \varepsilon\} \leq \frac{M|\eta_n|}{\varepsilon} = \frac{\alpha_2}{\varepsilon\sqrt{n}} \rightarrow 0.$$

Отсюда и из теоремы 2.42 следует, что предельное распределение ζ_n совпадает с предельным распределением ξ_n . Согласно центральной предельной теореме случайная величина ξ_n асимптотически нормально распределена с параметрами $(0, \sqrt{\alpha_4 - \alpha_2^2})$. Поэтому центральный выборочный момент $\alpha_2^* = s^2 = (1/\sqrt{n})\zeta_n + \alpha_2$ при $n \rightarrow \infty$ асимптотически нормально распределен с параметрами $(\alpha_2, \sqrt{\frac{\alpha_4 - \alpha_2^2}{n}})$. \square

Теорема 3.12. *Допустим, что $\{\xi_n\}_{n=1}^\infty$ – асимптотически нормально распределенная последовательность случайных величин с параметрами $(a_n, \sigma_n/\sqrt{n})$, где $(a_n - a)/\sqrt{n} \rightarrow 0$, $\sigma_n \rightarrow \sigma$ при $n \rightarrow \infty$. Пусть функция $g(x)$ ограничена, имеет в окрестности точки a непрерывные частные производные второго порядка и $g'(a) \neq 0$. Тогда последовательность случайных величин $\eta_n = g(\xi_n)$ асимптотически нормальна с параметрами $(g(a), |g'(a)|\sigma/\sqrt{n})$, то есть при $n \rightarrow \infty$*

$$\mathbf{P}\left\{ \frac{(\eta_n - g(a))}{|g'(a)|\sigma} \sqrt{n} \leq x \right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

Доказательство. Пусть

$$g(x) = g(a_n) + g'(a_n)(x - a_n) + r(x).$$

Покажем, что предельное распределение $\sqrt{n}(\eta_n - g(a_n))$ совпадает с предельным распределением $\zeta_n = \sqrt{n}g'(a_n)(\xi_n - a_n)$. Введем случайное событие

$$A_n = \{\omega : |\xi_n(\omega) - a_n| < n^{-3/8}\}.$$

По неравенству Чебышева

$$\mathbf{P}(\overline{A_n}) \leq n^{3/4} \frac{\sigma_n}{n} \rightarrow 0 \quad \text{при } n \rightarrow \infty. \quad (3.1)$$

На множестве A_n случайную величину $r(\xi_n)$ можно представить в форме Лагранжа

$$r(\xi_n) = \frac{1}{2} g''(a_n + \theta(\xi_n - a_n)) (\xi_n - a_n)^2, \quad 0 < \theta < 1.$$

Так как $g''(a_n + \theta(\xi_n - a_n))$ ограничена на множестве A_n некоторой константой K и на A_n выполняется неравенство $|\xi_n - a_n| \leq 1/n^{3/8}$, то

$$|\sqrt{n}r(\xi_n)| \leq \sqrt{n}K \frac{1}{n^{3/4}} = \frac{K}{n^{1/4}}, \quad \omega \in A_n, \quad (3.2)$$

где K — некоторая константа.

Для произвольного $\varepsilon > 0$ имеем

$$\mathbf{P}\{|\sqrt{nr}(\xi_n)| > \varepsilon\} = \mathbf{P}\{(|\sqrt{nr}(\xi_n)| > \varepsilon) \cap A_n\} + \mathbf{P}\{(|\sqrt{nr}(\xi_n)| > \varepsilon) \cap \overline{A_n}\}. \quad (3.3)$$

Так как $\varepsilon > 0$ фиксировано и $|\sqrt{nr}(\xi_n)| \rightarrow 0$ в силу (3.2) при $n \rightarrow \infty$ равномерно по $\omega \in A_n$, то первое слагаемое в (3.3), начиная с некоторого n , равняется нулю.

Второе слагаемое в (3.3) оценивается вероятностью (3.1). Таким образом, $\mathbf{P}\{|\sqrt{n}R_n| > \varepsilon\} \rightarrow 0$ при $n \rightarrow \infty$. Следовательно, согласно теореме 2.42 предельное распределение случайной величины $(\eta_n - f(a_n))\sqrt{n}$ совпадает с предельным распределением случайной величины ζ_n . Случайная величина ζ_n распределена асимптотически нормально с параметрами $(0, |g'(a_n)\sigma_n|)$. Нетрудно проверить, что предельное распределение случайной величины $\sqrt{n}(\eta_n - g(a_n))/|g'(a_n)|\sigma_n$ совпадает с предельным распределением случайной величины $\sqrt{n}(\eta_n - g(a))/|g'(a)|\sigma$. Поэтому η_n является асимптотически нормально распределенной случайной величиной с параметрами $(g(a), |g'(a)|\sigma/\sqrt{n})$. \square

Приведем здесь формулировку схожего утверждения для последовательности асимптотически нормальных векторов. Асимптотическая нормальность здесь означает, что $\mathbf{P}\{(\xi_n - m)/\sqrt{n} \leq x\}$ сходится к нормальному распределению с ковариационной матрицей $\|\sigma_{\ell k}\|$ и нулевым математическим ожиданием.

Теорема 3.13. Пусть $\xi_n = (\xi_{n1}, \dots, \xi_{nr})$, $n = 1, 2, \dots$ — последовательность асимптотически нормальных векторов с моментами

$$M\xi_{n\ell} = m_{n\ell} \rightarrow m_\ell \quad (\ell = 1, \dots, r),$$

$$\text{cov}(\xi_{n\ell}, \xi_{nk}) = \frac{\sigma_{\ell k}(n)}{n} \quad (k, \ell = 1, \dots, r),$$

где $(m_{n\ell} - m_\ell)\sqrt{n} \rightarrow 0$ и $\sigma_{\ell k}(n) \rightarrow \sigma_{\ell k}$ при $n \rightarrow \infty$ и пусть функция $g(x) = g(x_1, \dots, x_r)$ определена в окрестности точки $m = (m_1, \dots, m_r)$ и имеет в этой окрестности непрерывные частные производные второго порядка. Если $\eta_n = g(\xi_{n1}, \dots, \xi_{nr})$ и величина $\sigma^2 = \sum_{k, \ell=1}^r g_k g_\ell \sigma_{k\ell}$, где $g_i = \frac{\partial g}{\partial x_i}(m)$, положительна, то при $n \rightarrow \infty$

$$\mathbf{P}\left\{\frac{\sqrt{n}(\eta_n - g(m))}{\sigma} \leq x\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

3.6. Методы нахождения оценок

Метод моментов

Пусть (x_1, \dots, x_n) — выборка из семейства распределений $\{F(u, \theta)\}$, $u \in \mathbf{R}$, зависящего от неизвестного r -мерного параметра $\theta = (\theta_1, \dots, \theta_r)$ из множества $\Theta \subset \mathbf{R}^r$. Пусть вектор-функция $g(x) = (g_1(x), \dots, g_r(x))$ такова, что уравнение

$$m_g(\theta) = t,$$

где $t = (t_1, \dots, t_r) \in m_g(\Theta)$, $m_g(\theta) = (m_{g_1}(\theta), \dots, m_{g_r}(\theta))$ и $m_{g_j}(\theta) = Mg_j(\xi) = \int_{-\infty}^{\infty} g_j(u) dF(u, \theta)$, однозначно и непрерывно разрешимо относительно θ в области значений $\theta \in \Theta$, то есть $\theta = m_g^{-1}(t)$. Допустим, что вектор

$$\bar{g} = (\bar{g}_1, \dots, \bar{g}_r) = \left(\frac{1}{n} \sum_{i=1}^n g_1(x_i), \dots, \frac{1}{n} \sum_{i=1}^n g_r(x_i) \right)$$

принадлежит области $m_g(\Theta)$.

Определение 44. Оценка $\theta^* = m_g^{-1}(\bar{g})$ называется оценкой, полученной по методу моментов.

Подобного рода оценки θ^* состоятельны.

Теорема 3.14. Оценка $\theta^* = (\theta_1^*, \dots, \theta_r^*)$, получаемая как решение уравнения $m_g(\theta) = \bar{g}$, является состоятельной.

Доказательство. Имеем

$$\theta_k^* = (m_g^{-1})_k(\bar{g}_1, \dots, \bar{g}_r),$$

где $\bar{g}_k \xrightarrow{P} m_{gk}$, $k = 1, \dots, r$, согласно теореме 3.4 о состоятельности эмпирических моментов, и m_g^{-1} является непрерывной векторнозначной функцией. По лемме 3.1

$$\theta_k^* = (m_g^{-1})_k(\bar{g}_1, \dots, \bar{g}_r) \xrightarrow{P} (m_g^{-1})_k(m_{g1}, \dots, m_{gr}) = \theta_k. \quad \square$$

Если в качестве функций $g_k(x)$, $k = 1, \dots, r$, выбрать функции $g_k(x) = x^k$, то наше векторное уравнение превращается в систему уравнений для обычных моментов. Свое название "метод моментов" получил потому, что суть его состоит в приравнивании "теоретических" и эмпирических моментов.

Для того, чтобы имела место асимптотическая нормальность оценки θ^* достаточно потребовать дополнительно, чтобы отображение m_g было непрерывно дифференцируемым и чтобы были конечны интегралы $\int g_i(x) dF(x, \theta)$, $i = 1, \dots, r$.

Пример 5. Рассмотрим семейство показательных распределений с плотностью вероятности $\{p(u, \lambda)\} = \lambda e^{-\lambda u}$, $u \geq 0$, с неизвестным параметром λ . Построим оценку для λ сначала с функцией $g(x) = x$, а затем с функцией $g(x) = x^2$. Справедливо следующее равенство

$$m_1(\lambda) = Mx_1 = \int_0^{\infty} u \lambda e^{-\lambda u} du = \frac{1}{\lambda}.$$

Из уравнения $m_1(\lambda) = \bar{x}$ найдем оценку по методу моментов

$$\lambda_1^* = \frac{1}{\bar{x}}.$$

Для второй функции $g(x) = x^2$ имеем

$$m_2(\lambda) = Mx_1^2 = \int_0^{\infty} u^2 \lambda e^{-\lambda u} du = \frac{2}{\lambda^2},$$

Решением уравнения $m_2(\lambda) = \frac{1}{n} \sum_{i=1}^n x_i^2$ является оценка

$$\lambda_2^* = \left(\frac{1}{2n} \sum_{i=1}^n x_i^2 \right)^{-1/2}.$$

Из приведенной выше теоремы 3.14 следует, что обе оценки λ_1^* и λ_2^* являются состоятельными.

Пример 6. Пусть (x_1, \dots, x_n) – нормально распределенная выборка с неизвестными параметрами m и σ^2

$$p(u, m, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Рассмотрим $g_1(x) = x$, $g_2(x) = x^2$. получим следующие уравнения для метода моментов

$$m = \bar{x}, \quad \sigma^2 + m^2 = \frac{1}{n} \sum_{i=1}^n x_i^2,$$

из которых находим оценки

$$m^* = \bar{x}, \quad (\sigma^2)^* = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2 = s^2.$$

Метод наибольшего правдоподобия

Пусть (x_1, \dots, x_n) — выборка из распределения с плотностью $p(u, \theta)$, $u \in \mathbf{R}$, зависящая, вообще говоря, от векторного параметра θ . Функция

$$L(x_1, \dots, x_n, \theta) = \prod_{k=1}^n p(x_k, \theta)$$

как функция θ называется функцией правдоподобия. Оценкой наибольшего правдоподобия называется статистика $\theta_n^* = \theta^*(x_1, \dots, x_n)$ для которой

$$L(x_1, \dots, x_n, \theta_n^*) = \max_{\theta} L(x_1, \dots, x_n, \theta). \quad (3.4)$$

Оценка наибольшего правдоподобия определяется тем условием, что вероятность получить конкретную выборку (x_1, \dots, x_n) должна быть наибольшей. На практике вместо $\max L$ ищут максимум $\log L$.

Теорема 3.15. *В условиях теоремы Рао–Крамера, если для параметра θ существует несмещенная и эффективная оценка θ^* , то уравнение правдоподобия имеет единственное решение θ^* .*

Доказательство. Если θ^* эффективная несмещенная оценка, то неравенство Рао–Крамера обращается в равенство. Что возможно только при выполнении условия

$$\frac{\partial \log L(x_1, \dots, x_n, \theta)}{\partial \theta} = k(\theta^* - \theta),$$

где константа k зависит от θ . Уравнение правдоподобия эквивалентно уравнению

$$\theta^* - \theta = 0,$$

которое имеет единственное решение, а именно, статистику $\theta^*(x_1, \dots, x_n)$, удовлетворяющую (3.4). \square

Пример 7. Нормальное распределение имеет плотность

$$p(u, m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-m)^2}{2\sigma^2}}, \quad -\infty < m < \infty, \quad \sigma > 0.$$

Положим $\theta = (m, \sigma)$. Функция правдоподобия запишется в виде

$$L(x_1, \dots, x_n, m, \sigma) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 \right\}.$$

Так как L и $\log L$ достигают максимума при одних и тех же θ , для нахождения точки максимума решаем систему уравнений

$$\begin{aligned} \frac{\partial \log L}{\partial m} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m) = 0, \\ \frac{\partial \log L}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - m)^2 = 0. \end{aligned}$$

Получим

$$m^* = \bar{x}, \quad (\sigma^2)^* = s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Так как второй дифференциал функции $\log L$ в точке (\bar{x}, s) отрицательно определен, то в этой точке действительно достигается максимум этой функции.

Пример 8. Для биномиально распределенной выборки объема n имеем

$$L(x_1, \dots, x_n, p) = p^\nu (1-p)^{n-\nu},$$

где $p = \mathbf{P}(x_i = 1)$, $\mathbf{P}(x_i = 0) = 1 - p$, и ν – число элементов выборки, равных единице. Поэтому

$$\log L(x_1, \dots, x_n, p) = \nu \log p + (n - \nu) \log(1 - p),$$

и

$$\frac{\partial L}{\partial p} = \frac{\nu}{p} - \frac{n - \nu}{1 - p}.$$

Решением уравнения является оценка

$$p^* = \frac{\nu}{n}.$$

Нетрудно проверить, что в этой точке действительно достигается максимум функции L .

Пример 9. Пусть выборка объема n равномерно распределена на отрезке $[0, \theta]$. В этом случае

$$p(u, \theta) = \begin{cases} \theta^{-1}, & z \in [0, \theta], \\ 0, & z \notin [0, \theta]. \end{cases}$$

Поэтому

$$L(x_1, \dots, x_n, \theta) = \begin{cases} \theta^{-n}, & \text{если } x_i \in [0, \theta] \text{ при всех } i = 1, \dots, n. \\ 0, & \text{в противном случае.} \end{cases}$$

Обозначим через $x_{(n)} = \max x_i$. Тогда $L(x_1, \dots, x_n, \theta) = 0$ при $\theta \in [0, x_{(n)})$, и $L(x_1, \dots, x_n, \theta) = \theta^{-n}$ при $\theta \in (x_{(n)}, \infty)$. Поэтому максимум функции L достигается в точке $\theta^* = x_{(n)}$.

Пусть x_1, \dots, x_n – выборка из распределения с плотностью $p(u, \theta)$, зависящей от параметра θ . Приведем без доказательства теорему об асимптотической нормальности оценки наибольшего правдоподобия. Будем предполагать, что

- 1) Истинное значение параметра θ_0 лежит внутри интервала (θ_1, θ_2) и в этом интервале существуют производные $\frac{\partial p}{\partial \theta}$, $\frac{\partial^2 p}{\partial \theta^2}$ и $\frac{\partial^3 p}{\partial \theta^3}$.
- 2) Интеграл $\int p(u, \theta) du$ можно два раза дифференцировать под знаком интеграла.
- 3) Информация Фишера одного наблюдения $J_1(\theta) = \int \left(\frac{\partial \log p}{\partial \theta} \right)^2 p(u) du$ в точке θ_0 положительна и существует мажоранта $H(u)$ семейства функций $\frac{\partial^3 \log p(u, \theta)}{\partial \theta^3}$

$$\left| \frac{\partial^3 \log p(u, \theta)}{\partial \theta^3} \right| \leq H(u),$$

удовлетворяющая соотношению

$$M_\theta H(\xi) = \int H(x) p(x, \theta) dx \leq M,$$

где M не зависит от θ .

Все рассмотренные ранее распределения удовлетворяют условиям 1, 2 и 3.

Теорема 3.16. При выполнении условий 1, 2 и 3 уравнение правдоподобия $\frac{\partial L(x_1, \dots, x_n, \theta)}{\partial \theta} = 0$ имеет решение θ_n^* , которое при $n \rightarrow \infty$ сходится по вероятности к θ_0 . Эта оценка наибольшего правдоподобия асимптотически нормальна с параметрами $(\theta_0, \frac{1}{\sqrt{J_1(\theta_0)n}})$ и асимптотически эффективна. Точнее, $\theta_n^* \xrightarrow{P} \theta_0$, $e(\theta_n^*) \rightarrow 1$ и

$$\lim_{n \rightarrow \infty} \mathbf{P}\{(\theta_n^* - \theta_0)\sqrt{J_1(\theta_0)n} \leq x\} = \Phi(x).$$

Пример 10. Рассмотрим Γ -распределение с плотностью

$$p(u, \alpha) = \frac{\alpha^\lambda}{\Gamma(\lambda)} u^{\lambda-1} e^{-\alpha u}, \quad u \geq 0, \quad \alpha > 0,$$

в случае когда параметр λ известен. Имеем

$$L(x_1, \dots, x_n, \alpha) = \lambda n \log \alpha - n \log \Gamma(\lambda) + (\lambda - 1) \sum_{i=1}^n \log x_i - \alpha \sum_{i=1}^n x_i,$$

$$\frac{\partial L}{\partial \alpha} = \frac{\lambda n}{\alpha} - \bar{x}n, \quad \alpha_n^* = \frac{\lambda}{\bar{x}}.$$

Согласно доказанной теореме полученная оценка α_n^* является состоятельной, асимптотически эффективной и асимптотически нормальной.

3.7. Доверительные интервалы

Пусть x_1, \dots, x_n – выборка из распределения с плотностью $p(u, \theta)$, зависящего от параметра θ . Интервал (θ_1, θ_2) со случайными границами $\theta_i(x_1, \dots, x_n)$ называется *доверительным интервалом* для параметра θ с *надежностью* $\gamma = 1 - \alpha$, где α – *уровень значимости*, если $\mathbf{P}\{\theta_1 < \theta < \theta_2\} = \gamma$.

Пусть $X(x_1, \dots, x_n)$ – некоторая статистика с функцией распределения $F(x, \theta) = \mathbf{P}\{X \leq x\}$. Решение $\hat{x}_\alpha = \hat{x}_\alpha(\theta)$ уравнения $F(x, \theta) = 1 - \alpha$ называется *квантилью порядка α* . Пусть α_1 и α_2 – *разбиение α* , то есть $\alpha_1, \alpha_2 > 0$, $\alpha_1 + \alpha_2 = \alpha$, и пусть \hat{x}_{α_1} , $\hat{x}_{1-\alpha_2}$ квантили порядка α_1 и $1 - \alpha_2$. Тогда неравенство

$$\hat{x}_{1-\alpha_2} \leq X \leq \hat{x}_{\alpha_1} \tag{3.5}$$

выполняется с вероятностью γ , так как

$$\mathbf{P}\{\hat{x}_{1-\alpha_2} \leq X \leq \hat{x}_{\alpha_1}\} = F(\hat{x}_{\alpha_1}, \theta) - F(\hat{x}_{1-\alpha_2}, \theta) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha = \gamma.$$

Предположим, что $F(x, \theta)$ есть убывающая (возрастающая) функция от θ , тогда квантиль \hat{x}_α является возрастающей (убывающей) функцией параметра θ . Монотонность по θ квантили $\hat{x}_\alpha(\theta)$ позволяет разрешить неравенство (3.5) относительно θ :

$$\theta_1(x_1, \dots, x_n, \alpha_1) \leq \theta \leq \theta_2(x_1, \dots, x_n, 1 - \alpha_2).$$

Таким образом, (θ_1, θ_2) является доверительным интервалом для параметра θ с надежностью γ .

Разбиение α_1, α_2 можно варьировать. Средняя длина $M[\theta_2 - \theta_1]$ доверительного интервала при этом, вообще говоря, меняется. Желательно среди всех доверительных интервалов с надежностью γ выбрать тот, который имеет наименьшую среднюю длину. Для симметричных распределений, то есть для распределений с четной плотностью $p(u, \theta)$, оптимальным разбиением, как будет видно ниже, является $\alpha_1 = \alpha_2 = \alpha/2$. Для χ^2 -распределенной случайной величины имеются таблицы выбора разбиения α наилучшим образом.

**Доверительные интервалы для параметров
нормального распределения**

Пусть x_1, \dots, x_n – выборка из нормального распределения с параметрами (m, σ) .

Теорема 3.17. а) Выборочное среднее \bar{x} и исправленная выборочная дисперсия \hat{s}^2 взаимно независимы; \bar{x} распределена нормально с параметрами $(m, \sigma/\sqrt{n})$ и $(n-1)\hat{s}^2/\sigma^2$ имеет распределение χ^2 с $(n-1)$ -й степенью свободы.

б) Случайная величина $\frac{\bar{x}-m}{\hat{s}}\sqrt{n}$ подчиняется распределению Стьюдента с $(n-1)$ степенью свободы.

Доказательство. Случайные величины $y_i = (x_i - m)/\sigma$ независимы и распределены нормально с параметрами $(0, 1)$. Введем $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ и $s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$. Достаточно показать, что случайная величина $(n-1)s_1^2 = (n-1)\hat{s}^2/\sigma^2$ имеет распределение χ^2 с $(n-1)$ степенью свободы и случайные величины \bar{y} и s_1^2 независимы. Пусть $C = (c_{ij})$ – ортогональная матрица такая, что $c_{1j} = \frac{1}{\sqrt{n}}$, $j = 1, \dots, n$, и пусть $(z_1, \dots, z_n)^t = C(y_1, \dots, y_n)^t$ – ортогональное преобразование заданное соотношениями

$$z_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^n y_j = \sqrt{n}\bar{y}, \quad z_i = \sum_{j=1}^n c_{ij}y_j, \quad i = 2, \dots, n.$$

Так как $\sum_{j=1}^n c_{ij}c_{jk} = \delta_{ik}$, имеем

$$\sum_i z_i^2 = \sum_j y_j^2$$

и

$$Mz_i = \sum_{j=1}^n c_{ij}My_j = 0, \quad M[z_iz_k] = \sum_{j=1}^n c_{ij}c_{jk}My_j^2 = \sum_{j=1}^n c_{ij}c_{jk} = 0, \quad \text{если } i \neq k,$$

$$Dz_i = D\left(\sum_{j=1}^n c_{ij}y_j\right) = \sum_{j=1}^n c_{ij}^2 Dy_j = \sum_{j=1}^n c_{ij}^2 = 1.$$

Таким образом, z_1, \dots, z_n распределены по нормальному закону с параметрами $(0, 1)$ и некоррелированы, а значит, z_1, \dots, z_n независимы. Учитывая, что $z_1 = \sqrt{n}\bar{y}$, случайная величина

$$(n-1)s_1^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{j=1}^n z_j^2 - z_1^2 = \sum_{j=2}^n z_j^2$$

имеет распределение χ^2 с $(n-1)$ степенью свободы как и $(n-1)\hat{s}^2/\sigma^2$. Отсюда и выше сказанного следует, что z_1 и $z_2^2 + \dots + z_n^2$ независимы (см. замечание на стр. 34). Поэтому $\bar{y} = z_1/\sqrt{n}$ и $s_1^2 = \sum_{j=2}^n z_j^2/(n-1)$ независимы. Для того, чтобы завершить доказательство теоремы заметим, что случайная величина

$$Y = \frac{\bar{x} - m}{\hat{s}}\sqrt{n} = \frac{\sigma\bar{y}\sqrt{n}}{\sigma s_1} = \frac{\bar{y}\sqrt{n}}{s_1} = z_1 \sqrt{\frac{n-1}{\sum_{j=2}^n z_j^2}}$$

является отношением Стьюдента с $(n-1)$ степенью свободы. \square

Пусть независимая выборка x_1, \dots, x_n взята из нормального распределения с параметрами (m, σ) .

а) Доверительный интервал для параметра m при известном σ . В качестве оценки возьмем среднее арифметическое \bar{x} . Статистика \bar{x} отвечает всем требованиям, предъявляемым к оценкам и имеет нормальное распределение с параметрами $(m, \frac{\sigma}{\sqrt{n}})$. Обозначим через \hat{u}_α квантиль нормального распределения порядка α , то есть решение уравнения

$$1 - \Phi(u) = \alpha.$$

Так как

$$\Phi(\hat{u}_{1-\alpha}) = 1 - \Phi(\hat{u}_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\hat{u}_\alpha}^{\infty} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\hat{u}_\alpha} e^{-\frac{u^2}{2}} du = \Phi(-\hat{u}_\alpha),$$

имеем $\hat{u}_{1-\alpha} = -\hat{u}_\alpha$. Пусть $\alpha = \alpha_1 + \alpha_2$. Статистика $Y = \frac{\bar{x}-m}{\sigma} \sqrt{n}$ имеет нормальное распределение с параметрами $(0, 1)$. Поэтому неравенство $-\hat{u}_{\alpha_2} \leq Y \leq \hat{u}_{\alpha_1}$, то есть

$$m - \hat{u}_{\alpha_2} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq m + \hat{u}_{\alpha_1} \frac{\sigma}{\sqrt{n}}, \quad (3.6)$$

выполняется с вероятностью $1 - \alpha$. Из (3.6) находим доверительный интервал для параметра m

$$\bar{x} - \hat{u}_{\alpha_1} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x} + \hat{u}_{\alpha_2} \frac{\sigma}{\sqrt{n}}.$$

Длина доверительного интервала равна $(\hat{u}_{\alpha_1} + \hat{u}_{\alpha_2}) \frac{\sigma}{\sqrt{n}}$. Покажем, что эта длина будет наименьшей, если $\alpha_1 = \alpha_2 = \alpha/2$. Пусть, например, $\alpha_2 > \alpha_1$ и выберем $\Delta > 0$ так, чтобы $\alpha_2 - \Delta > \alpha_1 + \Delta$; тогда $\hat{u}_{\alpha_1} > \hat{u}_{\alpha_1+\Delta} > \hat{u}_{\alpha_2-\Delta} > \hat{u}_{\alpha_2}$. Из неравенств

$$\begin{aligned} \Delta &= 1 - \Phi(\hat{u}_{\alpha_2}) - (1 - \Phi(\hat{u}_{\alpha_2-\Delta})) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{\hat{u}_{\alpha_2}}^{\hat{u}_{\alpha_2-\Delta}} e^{-\frac{x^2}{2}} dx \geq \frac{1}{\sqrt{2\pi}} e^{-\frac{\hat{u}_{\alpha_2-\Delta}^2}{2}} (\hat{u}_{\alpha_2-\Delta} - \hat{u}_{\alpha_2}), \end{aligned}$$

$$\begin{aligned} \Delta &= 1 - \Phi(\hat{u}_{\alpha_1+\Delta}) - (1 - \Phi(\hat{u}_{\alpha_1})) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{\hat{u}_{\alpha_1+\Delta}}^{\hat{u}_{\alpha_1}} e^{-\frac{x^2}{2}} dx \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{\hat{u}_{\alpha_1+\Delta}^2}{2}} (\hat{u}_{\alpha_1} - \hat{u}_{\alpha_1+\Delta}), \end{aligned}$$

следует, что

$$\hat{u}_{\alpha_2-\Delta} - \hat{u}_{\alpha_2} \leq \Delta \sqrt{2\pi} e^{\frac{\hat{u}_{\alpha_2-\Delta}^2}{2}} < \Delta \sqrt{2\pi} e^{\frac{\hat{u}_{\alpha_1+\Delta}^2}{2}} \leq \hat{u}_{\alpha_1} - \hat{u}_{\alpha_1+\Delta},$$

то есть

$$\hat{u}_{\alpha_2-\Delta} + \hat{u}_{\alpha_1+\Delta} < \hat{u}_{\alpha_1} + \hat{u}_{\alpha_2}.$$

Таким образом, при $\alpha_1 \neq \alpha_2$ доверительный интервал не имеет наименьшей длины.

б) Доверительный интервал для параметра m при неизвестном σ . Для построения доверительного интервала воспользуемся отношением Стьюдента. Пусть $S_n(t)$ – функция распределения Стьюдента с n степенями свободы. Обозначим через $\hat{t}_\alpha(n)$ квантиль распределения $S_n(t)$, то есть решение уравнения

$$S_n(t) = 1 - \alpha.$$

Так как плотность распределения Стьюдента четная функция, то $t_{1-\alpha}(n) = -t_\alpha(n)$ и при построении доверительного интервала как и в пункте а) следует выбрать разбиение $\alpha_1 = \alpha_2 = \alpha/2$. По теореме 3.17 неравенство

$$-\frac{\hat{s}}{\sqrt{n}} \hat{t}_{\alpha/2}(n-1) \leq \bar{x} - m \leq \frac{\hat{s}}{\sqrt{n}} \hat{t}_{\alpha/2}(n-1)$$

выполняется с вероятностью $1 - \alpha$. Отсюда находим доверительный интервал для параметра m

$$\bar{x} - \frac{\hat{s}}{\sqrt{n}} \hat{t}_{\alpha/2}(n-1) \leq m \leq \bar{x} + \frac{\hat{s}}{\sqrt{n}} \hat{t}_{\alpha/2}(n-1)$$

в) Доверительный интервал для параметра σ при известном m . Воспользуемся статистикой

$$n \frac{s_0^2}{\sigma^2} = \sum_{k=1}^n \left(\frac{x_k - m}{\sigma} \right)^2,$$

имеющей χ^2 -распределение с n степенями свободы. Обозначим через $K_n(x)$ функцию распределения случайной величины ns_0^2/σ^2 и через $\widehat{k}_\alpha(n)$ – квантиль $K_n(x)$, то есть корень уравнения

$$K_n(x) = 1 - \alpha.$$

Пусть $\alpha = \alpha_1 + \alpha_2$. Тогда неравенство

$$\widehat{k}_{1-\alpha_2}(n) \leq n \frac{s_0^2}{\sigma^2} \leq \widehat{k}_{\alpha_1}(n)$$

выполняется с вероятностью $1 - \alpha$. Разрешая это неравенство относительно σ находим доверительный интервал для параметра σ

$$\sqrt{\frac{ns_0^2}{\widehat{k}_{\alpha_1}(n)}} \leq \sigma \leq \sqrt{\frac{ns_0^2}{\widehat{k}_{1-\alpha_2}(n)}}$$

Можно доказать, что этот интервал имеет наименьшую среднюю длину среди всех доверительных интервалов этого вида с доверительной вероятностью $1 - \alpha$, если α_1 и α_2 выбраны так, что плотность $k_n(x) = K'_n(x)$ удовлетворяет соотношению

$$k_n(\widehat{k}_{1-\alpha_2}(n)) = k_n(\widehat{k}_{\alpha_1}(n)).$$

г) **Доверительный интервал для параметра σ при неизвестном m .** Воспользуемся теоремой 3.17, согласно которой статистика $\frac{\widehat{\sigma}^2}{\sigma^2}(n-1)$ имеет χ^2 -распределение с $(n-1)$ -й степенью свободы. Как и в пункте в) это приводит к доверительному интервалу

$$\widehat{s} \sqrt{\frac{n-1}{\widehat{k}_{\alpha_1}(n-1)}} \leq \sigma \leq \widehat{s} \sqrt{\frac{n-1}{\widehat{k}_{1-\alpha_2}(n-1)}}$$

с доверительной вероятностью $1 - \alpha$ ($\alpha = \alpha_1 + \alpha_2$).

Общий подход к доверительному оцениванию

При большом объеме выборки оценки метода моментов и наибольшего правдоподобия, как правило, асимптотически нормальны. Остановимся подробно на рассмотрении оценки, полученной методом наибольшего правдоподобия. В условиях теоремы 3.16 оценка θ_n^* асимптотически нормальна с параметрами $(\theta, \sqrt{\frac{1}{n J_1(\theta)}})$, то есть

$$\mathbf{P}\{(\theta_n^* - \theta_0) \sqrt{J_1(\theta_0)n} \leq x\} \rightarrow \Phi(x).$$

В случае, когда $\sigma^2(\theta) = \frac{1}{J_1(\theta)}$ не зависит от θ и равна σ^2 , будем считать, что θ_n^* нормально распределенной с параметрами $(\theta, \sigma/\sqrt{n})$. При этом допущении доверительный интервал имеет вид

$$\theta_n^* - \widehat{u}_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \theta \leq \theta_n^* + \widehat{u}_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

с приближительной надежностью $\gamma = 1 - \alpha$, где \widehat{u}_α – квантиль порядка α нормального распределения с параметрами $(0, 1)$.

Если $\sigma^2(\theta)$ зависит от θ , будем искать дифференцируемую функцию $\eta = g(\theta)$ такую, что

- 1) оценка $\eta_n^* = g(\theta_n^*)$ параметра $g(\theta)$ асимптотически нормальна с параметрами $(g(\theta), \frac{|g'(\theta)\sigma(\theta)|}{\sqrt{n}})$, причем $g'(\theta)\sigma(\theta)$ не зависит от неизвестного параметра θ ;
- 2) функция g строго монотонна на множестве допустимых значений параметра θ .

Пусть $g(\theta)$ такова, что $g'(\theta)\sigma(\theta) = 1$, то есть функция $g(\theta)$ является одной из первообразных неопределенного интеграла

$$\int \frac{d\theta}{\sigma(\theta)}.$$

В таком случае $g(\theta_n^*)$ асимптотически нормальна с параметрами $(g(\theta), 1/\sqrt{n})$. Поэтому неравенство

$$\eta_n^* - \hat{u}_{\alpha/2} \frac{1}{\sqrt{n}} \leq g(\theta) \leq \eta_n^* + \hat{u}_{\alpha/2} \frac{1}{\sqrt{n}}$$

выполняется с приближительной надежностью $\gamma = 1 - \alpha$. В результате, интервал

$$g^{-1}\left(\eta_n^* - \hat{u}_{\alpha/2} \frac{1}{\sqrt{n}}\right) \leq \theta \leq g^{-1}\left(\eta_n^* + \hat{u}_{\alpha/2} \frac{1}{\sqrt{n}}\right)$$

накрывает параметр θ с приближительной вероятностью γ .

Пример 11. Доверительный интервал для вероятности успеха в схеме Бернулли.

Пусть μ – число успехов в серии n независимых испытаний с неизвестной вероятностью p успеха в одном испытании. Оценка $p^* = \frac{\mu}{n}$ параметра p получена методом наибольшего правдоподобия. Оценка p^* асимптотически нормальна по теореме Муавра–Лапласа с параметрами $(p, \sqrt{\frac{p(1-p)}{n}})$. Выберем функцию g , равную

$$g(p) = \int \frac{dp}{\sqrt{p(1-p)}} = 2 \arcsin \sqrt{p}.$$

Тогда оценка $2 \arcsin \sqrt{\frac{\mu}{n}}$ асимптотически нормальна с параметрами $(2 \arcsin \sqrt{p}, \frac{1}{\sqrt{n}})$. Неравенство

$$2 \arcsin \sqrt{\frac{\mu}{n} - \frac{\hat{u}_{\alpha/2}}{\sqrt{n}}} \leq 2 \arcsin \sqrt{p} \leq 2 \arcsin \sqrt{\frac{\mu}{n} + \frac{\hat{u}_{\alpha/2}}{\sqrt{n}}}$$

выполняется с приближительной вероятностью $\gamma = 1 - \alpha$. Отсюда получаем приближенный доверительный интервал для параметра p

$$\sin^2 \left(\arcsin \sqrt{p^*} - \frac{\hat{u}_{\alpha/2}}{2\sqrt{n}} \right) \leq p \leq \sin^2 \left(\arcsin \sqrt{p^*} + \frac{\hat{u}_{\alpha/2}}{2\sqrt{n}} \right).$$

3.8. Непараметрические критерии проверки гипотез. Критерий значимости

Предположим, что гипотетическое распределение полностью задано заранее. И необходимо проверить статистическую гипотезу, что наша выборка получена из генеральной совокупности и этим распределением. Если гипотеза верна, то эмпирическая функция распределения $F_n^*(x)$ является приближением в каком-то смысле к функции $F(x)$ при $n \rightarrow \infty$. Предположим, что построена мера отклонения $F_n^*(x)$ от $F(x)$ – статистика S , которая является функцией от выборки и при справедливости основной гипотезы распределение этой статистики известно точно или асимптотически точно. Выберем число S_0 так, чтобы $\mathbf{P}\{S > S_0\} = \alpha$. Если α мало, то можем считать *практически достоверным фактом*, что событие с вероятностью α не произойдет в единичном опыте. По выборочным значениям вычислим величину S . Если окажется, что $S < S_0$, то гипотезу признаем разумной интерпретацией имеющихся данных. В противном случае гипотезу отвергаем. Описанный критерий называется *критерием значимости или согласия*. Параметр α называется уровнем значимости. Подчеркнем особо, что проверка гипотезы не есть доказательство гипотезы.

Одним из наиболее известных критериев согласия является χ^2 -критерий Пирсона.

**Критерий χ^2 в случае
полностью определенного гипотетического распределения
(χ^2 -критерий Пирсона)**

Рассмотрим гипотезу H , состоящую в том, что данные x_1, \dots, x_n образуют выборку из n значений случайной величины с данной функцией распределения $F = F_\xi$. Определим меру расхождения теоретического и гипотетического распределений. Множество значений случайной величины ξ разобьем на r промежутков I_1, I_2, \dots, I_r . По известной функции F вычислим вероятности событий $\mathbf{P}\{\xi \in I_i\} = p_i$. Через ν_i обозначим количество выборочных значений, попадающих в промежуток I_i . Ясно, что $\sum_{i=1}^r \nu_i = n$. Относительная частота попадания значения выборки в промежуток I_i равна $\frac{\nu_i}{n}$. За меру расхождения между выборочным распределением и гипотетическим примем

$$\sum_{i=1}^r c_i \left(\frac{\nu_i}{n} - p_i \right)^2.$$

Пирсон предложил c_i принять равным n/p_i . Тогда статистика, на основе которой строится критерий

$$\hat{\chi}_n^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} = \sum_{i=1}^r \frac{\nu_i^2}{np_i} - n,$$

просто выражается через эмпирические частоты ν_i и гипотетические частоты np_i .

Теорема 3.18. *При справедливости гипотезы H распределение $\hat{\chi}_n^2$ при $n \rightarrow \infty$ слабо сходится к χ^2 -распределению с $(r-1)$ -й степенью свободы.*

Доказательство. Если гипотеза H верна, то случайные величины ν_1, \dots, ν_r имеют полиномиальное распределение

$$\mathbf{P}\{\nu_i = n_i, i = 1, \dots, r\} = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}, \quad n_1 + \dots + n_r = n.$$

Из теоремы 2.38 следует, что случайный вектор $\eta_n = (\eta_{n1}, \dots, \eta_{nr})$ с компонентами

$$\eta_{nk} = \frac{\nu_k - np_k}{\sqrt{np_k}}$$

сходится при $n \rightarrow \infty$ к нормальному распределению с нулевыми средними и матрицей ковариации

$$\| \text{cov}(\eta_k, \eta_\ell) \| = \| \delta_{k\ell} - \sqrt{p_k p_\ell} \|.$$

Пусть η_k обозначают компоненты случайные величины с предельным нормальным распределением. Характеристическая функция предельного распределения F равна

$$e^{-\frac{1}{2} B_\eta(t)},$$

где

$$B_\eta(t) = \sum_{k=1}^r t_k^2 - \sum_{k,\ell} t_k t_\ell \sqrt{p_k p_\ell} = \sum_{k=1}^r t_k^2 - \left(\sum_{k=1}^r \sqrt{p_k} t_k \right)^2.$$

Пусть C – ортогональная матрица вида

$$\left\| \begin{array}{cccc} \sqrt{p_1} & c_{21} & \cdots & c_{r1} \\ \sqrt{p_2} & c_{22} & \cdots & c_{r2} \\ \cdots & \cdots & \cdots & \cdots \\ \sqrt{p_r} & c_{2r} & \cdots & c_{rr} \end{array} \right\|$$

Положим $(\xi_1, \dots, \xi_r) = \xi = \eta C = (\eta_1, \dots, \eta_r) C$ и $(u_1, \dots, u_r) = u = t C = (t_1, \dots, t_r) C$. Характеристические функции f_ξ и f_η связаны соотношением $f_\xi(u) = f_\eta(u C')$. Поэтому $f_\xi(u) = e^{-R_\xi(u)/2}$, где

$$B_\xi(u) = B_\eta(u C') = \sum_{k=1}^r t_k^2 - u_1^2 = \sum_{k=1}^r u_k^2 - u_1^2 = \sum_{k=2}^r u_k^2,$$

т.е. $D\xi_1 = 0$ и случайные величины ξ_2, \dots, ξ_r распределены нормально с параметрами $(0, 1)$ и независимы. Из предельной теоремы 3.19 теории меры и интеграла (формулировка которой приведена ниже) следует, что последовательность распределений случайных величин $\widehat{\chi}_n^2 = \eta_{n1}^2 + \dots + \eta_{nr}^2$ сходится слабо к распределению случайной величины $\chi^2 = \eta_1^2 + \dots + \eta_r^2$

$$\mathbf{P}\{\widehat{\chi}_n^2 \leq x\} = \int_{u_1^2 + \dots + u_r^2 \leq c} dF_{\eta_{n1}, \dots, \eta_{nr}}(u_1, \dots, u_r) \longrightarrow \int_{u_1^2 + \dots + u_r^2 \leq c} dF_{\eta_1, \dots, \eta_r}(u_1, \dots, u_r) = \mathbf{P}\{\chi^2 \leq x\}.$$

В силу ортогональности матрицы C имеет место равенство $\eta_1^2 + \dots + \eta_r^2 = \xi_1^2 + \dots + \xi_r^2 = \xi_2^2 + \dots + \xi_r^2$, учитывая, что $\xi_1 = 0$ п.н. Утверждение теоремы следует теперь из определения χ^2 -распределения. \square

На практике задаются уровнем значимости α . Согласно теореме 3.18 при больших n с вероятностью, приближенно равной α , выполняется неравенство

$$\widehat{\chi}_n^2 = \sum_{k=1}^r \frac{(\nu_k - np_k)^2}{np_k} \geq \widehat{k}_\alpha(r-1),$$

где $k_\alpha(r-1)$ — α -квантиль χ^2 -распределения с $(r-1)$ -й степенью свободы. Если гипотеза H верна, то практически невозможно встретить в единичной выборке значение $\widehat{\chi}_n^2$, превышающее $k_\alpha(r-1)$. Поэтому основная гипотеза считается принятой или выборка совместима с гипотезой, если $\widehat{\chi}_n^2 < k_\alpha(r-1)$ и отвергнутой, если выполняется противоположное неравенство.

Выбор точек z_1, \dots, z_{r-1} на практике осуществляется таким образом, чтобы удовлетворить неравенствам

$$np_k \geq 10, \quad \nu_k \geq 10.$$

Пример 12. В последовательности n независимых испытаний событие A произошло ν раз. Совместимо ли это с гипотезой, что событие A имеет заданную вероятность?

Каждую компоненту выборки будем рассматривать как индикатор события A . Гипотеза H формулируется следующим образом: вероятности событий A и \bar{A} равны p и q соответственно. Построим статистику Пирсона

$$\widehat{\chi}_n^2 = \frac{(\nu - np)^2}{np} + \frac{(n - \nu - nq)^2}{nq} = \frac{\nu^2}{np} + \frac{(n - \nu)^2}{nq} - n = \frac{(\nu - np)^2}{npq}.$$

Зададимся уровнем значимости α и по таблице значений χ^2 -распределения с одной степенью свободы найдем квантиль $k_\alpha(1)$. Сравним значение статистики $\widehat{\chi}_n^2$ с квантилью $k_\alpha(1)$. Если $\widehat{\chi}_n^2 < k_\alpha(1)$, то будем считать выборку совместимой с гипотезой, в противном случае отвергнем ее.

Пусть $F_n(x)$, $x = (x_1, \dots, x_m)$ — функция распределения случайного вектора $\xi_n = (\xi_{n1}, \dots, \xi_{nm})$ и $F(x)$ — функция распределения случайного вектора $\xi = (\xi_1, \dots, \xi_m)$. Последовательность $F_n(x)$ слабо сходится к $F(x)$ ($F_n(x) \Rightarrow F(x)$), если $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности предельной функции.

Через \mathbf{P}_{ξ_n} и \mathbf{P}_ξ обозначим вероятностные меры, порожденные функциями распределения $F_n(x)$ и $F(x)$.

Теорема 3.19. Если $F_n(x) \Rightarrow F(x)$, то $\lim_{n \rightarrow \infty} \mathbf{P}_{\xi_n}(A) = \mathbf{P}_\xi(A)$ для любого борелевского множества A , граница которого имеет P -меру нуль.

Эта теорема доказывается с помощью предельной теоремы Хелли.

Теорема 3.20 (Хелли). Если $g(x)$ – непрерывная ограниченная функция на пространстве \mathbf{R}^m и $F_n(x) \Rightarrow F(x)$, то

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}^m} g(x) dF_n(x) = \int_{\mathbf{R}^m} g(x) dF(x).$$

**Критерий χ^2 в случае,
когда по выборке оцениваются некоторые параметры**

Предположим теперь, что функция распределения $F(x, \theta_1, \dots, \theta_s)$ зависит от s параметров. Гипотеза H , подлежащая проверке, состоит в том, что выборка x_1, \dots, x_n извлечена из генеральной совокупности с функцией распределения F при некоторых значениях параметра.

Как и в предыдущем пункте будем считать, что выборка разбита на r групп, соответствующих r непересекающимся промежуткам I_1, \dots, I_r . Пусть ν_i количество выборочных значений, попавших в I_i . При истинных значениях параметров $\theta_1, \dots, \theta_s$ и $p_i(\theta_1, \dots, \theta_s) = \mathbf{P}\{x_k \in I_i\}$ – вероятности попадания случайной величины x_k в промежуток I_i . Параметры θ_i неизвестны и должны быть оценены по выборке. Но если θ_i заменить их оценками, то тогда p_i станут случайными величинами и мы не сможем применить теорему Пирсона. Однако оказывается, если в качестве оценок параметров θ_i взять оценки максимального правдоподобия θ_{in}^* , то как показал Р.Фишер статистика $\hat{\chi}_n^2$ имеет слабым пределом при $n \rightarrow \infty$ распределение χ^2 с $r - s - 1$ степенями свободы.

Теорема 3.22 (Р.Фишер). *Статистика*

$$\hat{\chi}_n^2 = \sum_{i=1}^r \frac{[\nu_i - np_i(\theta_{1n}^*, \dots, \theta_{sn}^*)]^2}{np_i(\theta_{1n}^*, \dots, \theta_{sn}^*)}$$

слабо сходится к χ^2 распределению с $r - s - 1$ степенями свободы.

Критерий Колмогорова

Другим непараметрическим критерием является критерий Колмогорова. Этот критерий основан на статистике

$$D_n = \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)|,$$

где $F_n^*(x)$ – эмпирическая функция распределения

$$F_n^*(x) = \frac{1}{n} \sum_{k=1}^n I_{\{x_k \leq x\}}.$$

Теорема 3.23. Если распределение $F(x)$ непрерывно, то распределение статистики D_n не зависит от $F(x)$.

Доказательство. Покажем, что статистика D_n для любой непрерывной $F(x)$ имеет такое же распределение, как и в случае, когда $F(x)$ определяет равномерное распределение на $[0, 1]$. Пусть x_1, \dots, x_n – выборка с распределением $F(x)$. Предположим, что $F(a) = 0$, $F(b) = 1$ и $F(x)$ строго возрастает на (a, b) , причем a и b могут быть бесконечными. Введем случайные величины $y_k = F(x_k)$. Так как x_1, \dots, x_n независимы, то и y_1, \dots, y_n независимы.

Покажем, что y_1, \dots, y_n равномерно распределены на $[0, 1]$. При любом $y \in (0, 1)$ имеем

$$\mathbf{P}\{y_k \leq y\} = \mathbf{P}\{x_k \leq F^{-1}(y)\} = F(F^{-1}(y)) = y$$

Для $y = F(x)$ случайные события $\{y_k \leq y\}$ и $\{x_k \leq x\}$ равносильны. Поэтому

$$F_n^*(x, x_1, \dots, x_n) = \frac{1}{n} \sum_{k=1}^n I_{\{x_k \leq x\}} = \frac{1}{n} \sum_{k=1}^n I_{\{y_k \leq y\}} = F_n^*(y, y_1, \dots, y_n).$$

Поэтому

$$D_n = \sup_{a < x < b} |F_n^*(x, x_1, \dots, x_n) - F(x)| = \sup_{0 < y < 1} |F_n^*(y, y_1, \dots, y_n) - y|.$$

В общем случае в качестве $F^{-1}(y)$ следует взять $\sup\{x : F(x) = y\}$. \square

А.Н. Колмогоров доказал, что

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\sqrt{n}D_n \leq x\} = K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 x^2}, \quad x > 0.$$

На основе этого предельного равенства строится критерий Колмогорова. Пусть \widehat{k}_α – α -квантиль предельного распределения

$$1 - K(\widehat{k}_\alpha) = \alpha.$$

Тогда гипотеза о том, что выборка взята из распределения $F(x)$, принимается, если $\sqrt{n}D_n \leq \widehat{k}_\alpha$, и отвергается, если $\sqrt{n}D_n > \widehat{k}_\alpha$. Уровень значимости этого критерия приблизительно равен α .

Критерий Смирнова

Пусть x_1, \dots, x_n и y_1, \dots, y_m – две независимые выборки. Первая имеет распределение $F(x)$, вторая выборка – $G(x)$. Введем статистику

$$D_{n,m} = \sup_{-\infty < x < \infty} |F_n^*(x, x_1, \dots, x_n) - F_m^*(x, y_1, \dots, y_m)|.$$

Н.В. Смирнов доказал, что если $F(x) = G(x)$ и непрерывны, то при $n, m \rightarrow \infty$ так, что $\frac{n}{m} \rightarrow \tau$, $0 < \tau < \infty$, случайная величина

$$\sqrt{\frac{nm}{n+m}} D_{n,m}$$

стремится к распределению $K(x)$. Эта предельная теорема позволяет строить критерий проверки гипотезы, что выборки x_1, \dots, x_n и y_1, \dots, y_m взяты из одного и того же распределения.

Критерий значимости для параметров

1) Пусть x_1, \dots, x_n и y_1, \dots, y_m выборки из нормальной совокупности с неизвестными средними. Для проверки гипотезы $H : \sigma_x^2 = \sigma_y^2$ используется статистика

$$F_{n,m} = \frac{\widehat{s}_x^2}{\widehat{s}_y^2} = \frac{m-1}{n-1} \frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2}{\sum_{i=1}^m \left(\frac{y_i - \bar{y}}{\sigma_y}\right)^2},$$

которая имеет распределение Фишера с $(n-1, m-1)$ степенями свободы. Пусть $\widehat{F}_\alpha(n, m)$ обозначим квантиль распределения Фишера, то есть решение уравнения

$$\mathbf{P}\{F_{n,m} > \widehat{F}_\alpha(n, m)\} = \alpha.$$

Заметим, что для квантилей $\widehat{F}_\alpha(n, m)$ выполняется соотношение

$$\widehat{F}_\alpha(n, m) = \frac{1}{\widehat{F}_{1-\alpha}(m, n)}.$$

Имеем

$$\begin{aligned} \mathbf{P}\{\widehat{F}_{1-\alpha/2}(n, m) < F_{n, m} \leq \widehat{F}_{\alpha/2}(n, m)\} &= \mathbf{P}\{F_{n, m} \leq \widehat{F}_{\alpha/2}(n, m)\} - \mathbf{P}\{F_{n, m} \leq \widehat{F}_{1-\alpha/2}(n, m)\} \\ &= 1 - \alpha/2 - (1 - (1 - \alpha/2)) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha. \end{aligned}$$

Положим

$$C_1 = \widehat{F}_{1-\alpha/2}(n, m) \text{ и } C_2 = \widehat{F}_{\alpha/2}(n, m).$$

Если для полученных выборок выполняется неравенство

$$C_1 \leq \frac{\widehat{s}_x^2}{\widehat{s}_y^2} \leq C_2$$

то гипотеза принимается, в противном случае гипотеза отклоняется.

2) Для проверки гипотезы о равенстве дисперсий ($H : \sigma_x^2 = \sigma_y^2$) для выборок x_1, \dots, x_n и y_1, \dots, y_m из нормальной совокупности с известными средними используется статистика

$$F_{n, m} = \frac{s_{0x}^2}{s_{0y}^2},$$

которая имеет распределение Фишера с (n, m) степенями свободы.

3) Гипотеза о равенстве средних ($H : m_x = m_y$) при известных дисперсиях σ_x^2 и σ_y^2 проверяется с использованием статистики

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{D(\bar{x} - \bar{y})}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

Статистика Z имеет нормальное распределение с параметрами $(0, 1)$. Гипотеза принимается, если $|Z| < \widehat{u}_{\alpha/2}$ и отклоняется в противном случае. Здесь $\widehat{u}_{\alpha/2}$ – квантиль нормального распределения с параметрами $(0, 1)$.

4) Обсудим здесь гипотезу о равенстве средних при неизвестной дисперсии в предположении равенства дисперсий $\sigma_x^2 = \sigma_y^2 = \sigma^2$. В общем случае получено только приближенное решение задачи. Статистики

$$(n-1) \frac{\widehat{s}_x^2}{\sigma_x^2} = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right)^2 \text{ и } (m-1) \frac{\widehat{s}_y^2}{\sigma_y^2} = \sum_{i=1}^m \left(\frac{y_i - \bar{y}}{\sigma_y} \right)^2$$

распределены по закону χ^2 с $(n-1)$ и $(m-1)$ степенями свободы. Тогда статистика

$$(n-1) \frac{\widehat{s}_x^2}{\sigma_x^2} + (m-1) \frac{\widehat{s}_y^2}{\sigma_y^2}$$

распределена по закону χ^2 с $(n+m-2)$ степенями свободы.

Заметим, что статистика

$$\frac{(n-1)\widehat{s}_x^2 + (m-1)\widehat{s}_y^2}{n+m-2}$$

является несмещенной оценкой дисперсии, так как

$$M \left[\frac{(n-1)\widehat{s}_x^2 + (m-1)\widehat{s}_y^2}{n+m-2} \right] = \frac{n-1}{n+m-2} \sigma^2 + \frac{m-1}{n+m-2} \sigma^2 = \sigma^2.$$

Далее, математическое ожидание случайной величины $\bar{x} - \bar{y}$ равно $m_x - m_y$ и дисперсия при наших допущениях равна $\sigma^2 \left(\frac{1}{n} + \frac{1}{m} \right)$. Поэтому случайная величина

$$\frac{(\bar{x} - \bar{y}) - (m_x - m_y)}{\sigma \sqrt{\left(\frac{1}{n} + \frac{1}{m} \right)}}$$

имеет нормальное распределение с параметрами $(0, 1)$. Случайная величина

$$\frac{(\bar{x} - \bar{y}) - (m_x - m_y)}{\sqrt{((n-1)\hat{s}_x^2 + (m-1)\hat{s}_y^2)}} \sqrt{n+m-2} \sqrt{\frac{mn}{n+m}}$$

не зависит от m_x , m_y и σ^2 и распределена по Стьюденту с $n+m-2$ степенями свободы. Для проверки гипотезы $H : m_x = m_y$ используется статистика

$$\tau_{n+m-2} = \frac{(\bar{x} - \bar{y})}{\sqrt{((n-1)\hat{s}_x^2 + (m-1)\hat{s}_y^2)}} \sqrt{n+m-2} \sqrt{\frac{mn}{n+m}},$$

распределенная по Стьюденту с $(n+m-2)$ степенями свободы.

Оценка коэффициента корреляции

Пусть $(x_1, y_1), \dots, (x_n, y_n)$ – независимые наблюдения над двумерным случайным вектором (X, Y) . Введем статистику

$$r^* = \frac{m_{11}}{s_x s_y},$$

где

$$m_{11} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{и} \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Как показал Фишер статистика

$$\frac{1}{2} \log \frac{1+r^*}{1-r^*}$$

асимптотически нормально распределена с параметрами

$$\left(\frac{1}{2} \log \frac{1+r}{1-r}, \frac{1}{\sqrt{n-3}} \right),$$

где r – коэффициент корреляции случайных величин X и Y . Напомним, что функция

$$\frac{1}{2} \log \frac{1+r}{1-r}$$

является обратной к гиперболическому тангенсу. Если \hat{x}_α – квантиль нормального распределения с параметрами $(0, 1)$, то доверительный интервал для коэффициента корреляции r с приближенной надежностью $1 - \alpha$ имеет вид

$$\tanh \left(\frac{1}{2} \log \frac{1+r^*}{1-r^*} - \hat{x}_{\alpha/2} \frac{1}{\sqrt{n-3}} \right) < r < \tanh \left(\frac{1}{2} \log \frac{1+r^*}{1-r^*} + \hat{x}_{\alpha/2} \frac{1}{\sqrt{n-3}} \right).$$

3.9. Статистические гипотезы.

Критерий Неймана – Пирсона

Пусть ξ – случайная величина из семейства распределений с плотностью $p(t, \theta)$, описывающая генеральную совокупность и x_1, \dots, x_n – независимая выборка из генеральной совокупности с совместной плотностью $p(u, \theta) = \prod_{k=1}^n p(u_k, \theta)$, $u = (u_1, \dots, u_n)$.

Предположения о параметре θ будем называть статистическими гипотезами. Ставится задача: проверить, что выборка согласуется с гипотезой о параметре θ или нет. Один из возможных подходов – строится критическое множество S в пространстве всех значений принимаемых выборкой: гипотеза отвергается, если конкретная выборка попадает в S и принимается в противном случае. Подобного

рода статистический критерий называется S -критерием. Относительно параметра θ имеются основная (проверяемая) гипотеза H_0 и часто конкурирующая (альтернативная) гипотеза H_1 . Гипотезы бывают простыми, $\theta = \theta_0$, и сложными, $\theta \in \Theta_0$.

Пример 13.

H_0 : выборка из нормального распределения с параметрами (a_0, σ_0) – простая гипотеза;

H_1 : выборка из нормального распределения с параметрами (a, σ_0) , $a \neq a_0$ – сложная гипотеза.

Рассмотрим случай двух простых гипотез. Пусть $H_0 : \theta = \theta_0$ и $H_1 : \theta = \theta_1$ – две простые гипотезы. С каждым S -критерием связаны ошибки двух родов.

Ошибка первого рода – отвергается гипотеза H_0 , когда она верна.

Ошибка второго рода – принимается гипотеза H_0 , когда верна гипотеза H_1 .

Если $\mathbf{P}_{H_i}(B) = \mathbf{P}_i(B) = \int_B p(u, \theta_i) du$, то $\alpha = \mathbf{P}_0(S)$ – вероятность ошибки первого рода и $\beta = \mathbf{P}_1(\bar{S}) = 1 - \mathbf{P}_1(S)$ – вероятность ошибки второго рода, а $1 - \beta = \mathbf{P}_1(S)$ – вероятность отвергнуть гипотезу H_0 , когда верна гипотеза H_1 .

Вероятность α ошибки первого рода называется уровнем значимости, а вероятность $\mathbf{P}_1(S)$, равная $1 - \beta$, обозначается через $W(S, \theta_1)$ и называется мощностью критерия. Среди всех S -критериев с уровнем значимости α ищется наиболее мощный критерий S^* . Если \mathcal{L}_α – множество всех S -критериев с уровнем значимости α , то

$$W(S^*, \theta_1) = \int_{S^*} p(u, \theta_1) du = \max_{S \in \mathcal{L}_\alpha} \int_S p(u, \theta_1) du.$$

Отметим, что оптимальный критерий не всегда существует.

Теорема Неймана-Пирсона

Предположим, что по выборке x_1, \dots, x_n нужно различить две простые гипотезы H_0 и H_1 , согласно которым выборка имеет плотность распределения $p_0(u)$ или $p_1(u)$, $u = (u_1, \dots, u_n)$. Введем критическое множество $S_c = \{u = (u_1, \dots, u_n) : p_1(u) \geq c p_0(u)\}$. Ограничимся рассмотрением случая, когда для любого α существует постоянная c такая, что $\mathbf{P}_{H_0}(S_c) = \alpha$.

Теорема 3.24. *Среди всех критериев, различающих гипотезы H_0 и H_1 с заданной ошибкой 1-го рода α , наиболее мощный критерий определяется критическим множеством S_c .*

Доказательство. Пусть S_c и S – критические множества с $\mathbf{P}_{H_0}(S_c) = \alpha$ и $\mathbf{P}_{H_0}(S) = \alpha$. Имеем

$$\mathbf{P}_{H_0}(S_c \setminus S_c S) = \alpha - \mathbf{P}_{H_0}(S_c S) = \mathbf{P}_{H_0}(S \setminus S S_c).$$

Отсюда находим, что

$$\begin{aligned} \mathbf{P}_{H_1}(S_c \setminus S_c S) &= \int_{S_c \setminus S_c S} p_1(u) du \geq c \int_{S_c \setminus S_c S} p_0(u) du = c \mathbf{P}_{H_0}(S_c \setminus S_c S) = \\ &= c \mathbf{P}_{H_0}(S \setminus S S_c) = c \int_{S \setminus S S_c} p_0(u) du > \int_{S \setminus S S_c} p_1(u) du = \mathbf{P}_{H_1}(S \setminus S S_c), \end{aligned} \quad (3.7)$$

так как на дополнении к S_c выполняется неравенство $p_1(u) < c p_0(u)$. Добавим к обеим частям равенства (3.7) слагаемое $\mathbf{P}_{H_1}(S S_c)$. В результате, получим

$$\mathbf{P}_{H_1}(S_c) \geq \mathbf{P}_{H_1}(S). \quad \square$$

С S -критерием связан индикатор отвергнуть гипотезу H_0 по выборке (x_1, x_1, \dots, x_n) . Введем понятие рандомизированного критерия. Пусть задана функция $0 \leq \varphi(\bar{x}) \leq 1$ для всех x . Значения $\varphi(x_1, \dots, x_n)$ для каждой выборки (x_1, x_1, \dots, x_n) , будем рассматривать как вероятность отвергнуть гипотезу H_0 . Тогда $1 - \varphi(x_1, \dots, x_n)$ – вероятность принять гипотезу. Рандомизированный критерий,

построенный по функции $\varphi(\bar{x})$, будем называть φ – критерием. уровнем значимости φ – критерия (ошибка первого рода) определим как $M_{H_0}(\varphi(\bar{x})) = \int_{R^n} \varphi(\bar{u})p(\bar{u}, \theta_0)d\bar{u}$: $\alpha = M_{H_0}(\varphi(\bar{x}))$. Величину ошибки второго рода определим как $\beta = 1 - M_{H_1}(\varphi(\bar{x}))$.

Рассмотрим множество \mathcal{L}_α всех φ – критериев с фиксированным уровнем значимости α , то есть $\{\varphi \in \mathcal{L}_\alpha; M_{H_0}(\varphi(\bar{x})) = \alpha\}$. Наиболее мощный критерий введем как

$$W(\varphi^*; H_1) = \max_{\varphi \in \mathcal{L}_\alpha} M_{H_1}(\varphi(\bar{x})).$$

Пусть $p_0(x) = p(x, H_0)$ и $p_1(x) = p(x, H_1)$, $M_0\varphi = \int_R \varphi(\bar{x})p_0(\bar{x})d\bar{x}$ и $M_1\varphi = \int_R \varphi(\bar{x})p_1(\bar{x})d\bar{x}$, $p_1(x)/p_0(x)$ – отношение правдоподобия.

Теорема 3.25. (Нейман – Пирсон) Для любого $0 \leq \alpha \leq 1$ существуют такие числа $c \geq 0$ и $0 \leq \varepsilon \leq 1$, что φ^* – критерий с функцией

$$\varphi^*(\bar{x}) = \begin{cases} 1, & \text{если } p_1(\bar{x}) > cp_0(\bar{x}) \\ \varepsilon, & \text{если } p_1(\bar{x}) = cp_0(\bar{x}) \\ 0, & \text{если } p_1(\bar{x}) < cp_0(\bar{x}) \end{cases}$$

определяет оптимальный критерий с уровнем значимости α .

Ограничимся случаем $0 < \alpha < 1$. Положим

$$g(c) = P\{p_1(\bar{x}) > cp_0(\bar{x}) | H_0\}.$$

Функция

$$1 - g(c) = P\left\{\frac{p_1(\bar{x})}{p_0(\bar{x})} \leq c | H_0\right\}$$

является функцией распределения случайной величины $p_1(x)/p_0(x)$. Кроме того, функция $g(c)$ убывает, непрерывной справа, $g(\infty) = 0$ и $g(-0) = 1$.

Определим c_α условием

$$g(c_\alpha) \leq \alpha \leq g(c_\alpha - 0).$$

Если $g(c_\alpha) < g(c_\alpha - 0)$, то выбираем

$$\varepsilon_\alpha = \frac{\alpha - g(c_\alpha)}{g(c_\alpha - 0) - g(c_\alpha)}.$$

Если $g(c_\alpha) = g(c_\alpha - 0)$, то полагаем $\varepsilon_\alpha = 0$. Если $g(c) = \alpha$ для целого отрезка $c_1 \leq c \leq c_2$, то полагаем $\varepsilon_\alpha = c_1$ или $\varepsilon_\alpha = c_2$. Определяя $c = c_\alpha$ и $\varepsilon = \varepsilon_\alpha$, строим критерий φ^* .

Можно показать, что φ^* –критерий имеет уровень значимости α и обладает свойством оптимальности.

Замечание. Теорема справедлива и для дискретных распределений $p_0(\bar{x})$ и $p_1(\bar{x})$. Надо только везде интегралы заменить суммами.

Оптимальный критерий для проверки гипотезы о параметрах нормального распределения

Пусть x_1, \dots, x_n – независимая выборка из нормального распределения с параметрами (a, σ) . Пусть σ известна, а относительно параметра a имеются две гипотезы:

- гипотеза H_0 : $a = a_0$;
- гипотеза H_1 : $a = a_1 > a_0$.

Отношение правдоподобия $p_1(u)/p_0(u)$ имеет вид

$$\frac{p_1(u)}{p_0(u)} = \exp \left\{ \frac{n}{\sigma^2} \bar{u}(a_1 - a_0) - \frac{n}{2\sigma^2} (a_1^2 - a_0^2) \right\}, \quad u = (u_1, \dots, u_n),$$

где $\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$. Ясно, что критическое множество $\{p_1(u)/p_0(u) > c\}$ совпадает с множеством $\{\bar{u} > c_1\}$ при некотором c_1 . Воспользовавшись тем, что выборочное среднее \bar{x} распределено нормально с параметрами $(a, \sigma/\sqrt{n})$, найдем ошибки первого и второго рода:

$$\alpha = \mathbf{P}\{\bar{x} > c_1 \mid H_0\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{c_1 - a_0}{\sigma} \sqrt{n}}^{\infty} e^{-\frac{u^2}{2}} du = 1 - \Phi\left(\frac{c_1 - a_0}{\sigma} \sqrt{n}\right) \quad (3.8)$$

$$\beta = \mathbf{P}\{\bar{x} < c_1 \mid H_1\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{c_1 - a_1}{\sigma} \sqrt{n}} e^{-\frac{u^2}{2}} du = \Phi\left(\frac{c_1 - a_1}{\sigma} \sqrt{n}\right) \quad (3.9)$$

Обозначим через \hat{u}_α квантиль нормального распределения

$$1 - \Phi(\hat{u}_\alpha) = 1 - \alpha.$$

Напомним, что $\hat{u}_\alpha = -\hat{u}_{1-\alpha}$. Тогда из (3.8) вытекает

$$\frac{c_1 - a_0}{\sigma} \sqrt{n} = \hat{u}_\alpha.$$

С помощью (3.9) находим квантиль \hat{u}_β

$$\hat{u}_\beta = -\frac{c_1 - a_1}{\sigma} \sqrt{n},$$

и по таблице значений функции $\Phi(x)$ определяем ошибку второго рода β .

Сравнивая

$$a_0 + \hat{u}_\alpha \frac{\sigma}{\sqrt{n}} = a_1 - \hat{u}_\beta \frac{\sigma}{\sqrt{n}} = c_1,$$

находим объем выборки n , который при оптимальном критерии обеспечивает ошибки первого и второго рода α и β , а именно,

$$n = \frac{\sigma^2(\hat{u}_\alpha + \hat{u}_\beta)^2}{(a_1 - a_0)^2}.$$

Величины α и β выбираются в зависимости от степени нежелательности ошибочных решений.

Пусть теперь параметр a известен. Рассмотрим следующие две гипотезы:

- гипотеза H_0 : $a = 0$, $\sigma = \sigma_0$;
- гипотеза H_1 : $a = 0$, $\sigma = \sigma_1 > \sigma_0$.

Отношение правдоподобия

$$\frac{p_1(u)}{p_0(u)} = \frac{\sigma_0^n}{\sigma_1^n} \exp \left\{ \frac{1}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^n u_i^2 \right\}, \quad u = (u_1, \dots, u_n)$$

приводит к критическому множеству

$$\sum_{i=1}^n u_i^2 \geq c_1.$$

Случайная величина $\sigma^{-2} \sum_{i=1}^n x_i^2$ имеет χ^2 -распределение с n степенями свободы с функцией распределения

$$K_n(x) = \int_0^x \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt, \quad x > 0.$$

Ошибки первого и второго рода определяются из равенств

$$\alpha = \mathbf{P}\left\{\frac{1}{\sigma_0^2} \sum_{i=1}^n x_i^2 \geq \frac{c_1}{\sigma_0^2}\right\} = 1 - K_n\left(\frac{c_1}{\sigma_0^2}\right),$$

$$\beta = \mathbf{P}\left\{\frac{1}{\sigma_1^2} \sum_{i=1}^n x_i^2 < \frac{c_1}{\sigma_1^2}\right\} = K_n\left(\frac{c_1}{\sigma_1^2}\right).$$

Обозначим через $\widehat{k}_\alpha(n)$ – квантиль порядка α χ^2 -распределения с n степенями свободы. По заданному уровню значимости α находим $c_1 = \widehat{k}_\alpha(n)\sigma_0^2$ и по таблице значений функции $K_n(x)$ находим ошибку второго рода.

Оптимальный критерий в схеме Бернулли

Пусть x_1, \dots, x_n – дискретная выборка, имеющая распределение

$$\mathbf{P}\{x_k = 1\} = p, \quad \mathbf{P}\{x_k = 0\} = 1 - p$$

с неизвестным параметром p . Событие $\{x_k = 1\}$ будем называть "успехом". Пусть $\mu_n = \sum_{k=1}^n x_k$ – число успехов в выборке. Предположим, что $0 < p_0 < p_1 < 1$. Рассмотрим две гипотезы:

- $H_0 : \mathbf{P}\{\mu_n = k | H_0\} = p_0(k) = C_n^k p_0^k (1 - p_0)^{n-k}$;
- $H_1 : \mathbf{P}\{\mu_n = k | H_1\} = p_1(k) = C_n^k p_1^k (1 - p_1)^{n-k}$.

Критерий для проверки гипотезы H_0 против H_1 строится исходя из неравенства

$$\frac{p_1(k)}{p_0(k)} = \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right]^k \frac{(1-p_1)^n}{(1-p_0)^n} \geq c,$$

которое равносильно неравенству $k \geq c_1$ при некотором c_1 . Для вычисления ошибок первого и второго рода воспользуемся асимптотической нормальностью случайной величины μ_n с параметрами $(np, \sqrt{np(1-p)})$. Имеем

$$\alpha = \mathbf{P}\{\mu_n \geq c_1 | H_0\} = \mathbf{P}\left\{\frac{\mu_n - np_0}{\sqrt{np_0(1-p_0)}} \geq \frac{c_1 - np_0}{\sqrt{np_0(1-p_0)}} \middle| H_0\right\},$$

$$\beta = \mathbf{P}\{\mu_n < c_1 | H_1\} = \mathbf{P}\left\{\frac{\mu_n - np_1}{\sqrt{np_1(1-p_1)}} < \frac{c_1 - np_1}{\sqrt{np_1(1-p_1)}} \middle| H_1\right\}.$$

Задавшись ошибкой первого рода α , найдем из первого равенства квантиль \widehat{u}_α нормального распределения с параметрами $(0, 1)$ и границу c_1 . Из второго равенства находим ошибку второго рода β .

Для вероятностей ошибок α и β граница c_1 представляется в виде

$$c_1 \approx np_0 + \widehat{u}_\alpha \sqrt{np_0(1-p_0)} \approx np_1 - \widehat{u}_\beta \sqrt{np_1(1-p_1)}.$$

Отсюда находим необходимый приблизительный объем выборки

$$n \approx \frac{(\widehat{u}_\alpha \sqrt{p_0(1-p_0)} + \widehat{u}_\beta \sqrt{p_1(1-p_1)})^2}{(p_0 - p_1)^2},$$

гарантирующий ошибку первого рода α и ошибку второго рода β .

Рекомендуемая литература

ОСНОВНАЯ:

1. Севастьянов Б.А. Курс теории вероятностей и математической статистики. М.: Наука, 1982.
2. Боровков А.А. Теория вероятностей. М.: Эдиториал УРСС, 1999.
3. Чистяков В.П. Курс теории вероятностей. М.: Наука, 1987.
4. Ширяев А.Н. Вероятность. М.: Наука, 1989.
5. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М.: Высшая школа, 1992.
6. Боровков А.А. Математическая статистика. М.: Эдиториал УРСС, 2000.
7. Севастьянов Б.А., Чистяков В.П., Зубков А.М. Сборник по теории вероятностей. М.: Наука, 1980.
8. Ивченко Г.И., Медведев Ю.И., Чистяков А.В. Сборник задач по математической статистике. М.: Высшая школа, 1989.

ДОПОЛНИТЕЛЬНАЯ:

9. Гнеденко Б.В. Курс теории вероятностей. М.: Наука, 1988.
10. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1998.
11. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика. М.: Наука, 1989.
12. Феллер В. Теория вероятностей и ее приложения. т. 1. М.: Мир, 1984.
13. Крамер Г. Математические методы статистики. М.: Мир, 1975.
14. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 1998.
15. Прохоров А.В., Ушаков В.Г., Ушаков Н.Г. Задачи по теории вероятностей: Основные понятия. Предельные теоремы. Случайные процессы. М.: Наука, 1986.