КОНСПЕКТ ЛЕКЦИЙ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

В данном конспекте в краткой форме излагается материал лекций по теории вероятностей, прочитанных на 2 потоке 3 курса Физического факультета МГУ в 2005 году. (Разбиение на лекции условное). Из-за краткости и схематичности изложения этот материал не претендует на учебник. Изложение, за исключением некоторых вопросов, соответствует учебнику Ю.П.Пытьева и И.А.Шишмарева «Курс теории вероятностей и математической статистики для физиков», МГУ, 1983. Автор заранее приносит свои извинения за возможные опечатки.

Теория вероятностей есть математический анализ понятия случайного эксперимента. Событие и вероятность являются основными понятиями этой теории (аксиоматика Колмогорова, 1929, альтернативный, эмпирически-статистический подход — Р.Фишер, Р.Мизес.) Лекц. 1

Предполагается, что результат или *исход* случайного эксперимента не может быть определен заранее. Однако сам эксперимент должен обладать свойством статистической устойчивости или устойчивости частот ($\frac{N_A}{N}$).

Пусть $\Omega = \{\omega\}$ — множество всех исходов и A есть некоторое событие, связанное с экспериментом. Естественно считать, что по исходу эксперимента можно сказать, осуществилось событие A или нет. Для наших дальнейших целей событие A можно отождествить с некоторой совокупностью исходов, то есть считать, что A есть подмножество элементов из Ω. Сами элементы ω мы будем называть элементарными событиями (предполагается, что их нельзя «разбить» на более мелкие).

Вероятность — числовая характеристика класса событий. Она имеет свойства, аналогичные относительной частоте события ($\frac{N_A}{N}$)[1], но не сводится к ней (не равна ей).

Примеры.

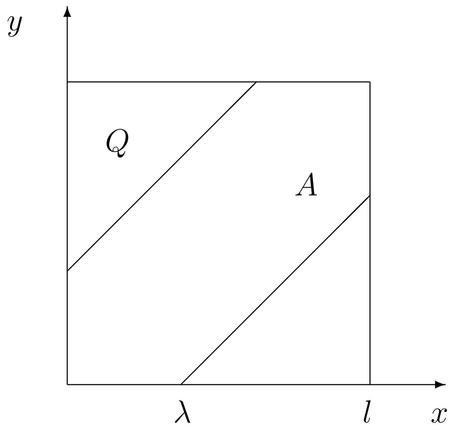
1. Бросание игральной кости — однородного куба. Выпадение граней ω_1, ω_2,..., ω_6 — элементарные события. Вероятность $P(\omega_i) = \frac{1}{6}$, $i = 1,\dots,6$. $A_1 = \{\omega_1, \omega_2\,\omega_3\}$, $A_2 = \{\omega_4, \omega_5\}$, $A_3 = \{\omega_6\}$ — события, вероятности которых легко подсчитать: $P(A_1) = \frac{1}{2}$, $P(A_2) = \frac{1}{3}$, $P(A_3) = \frac{1}{6}$.

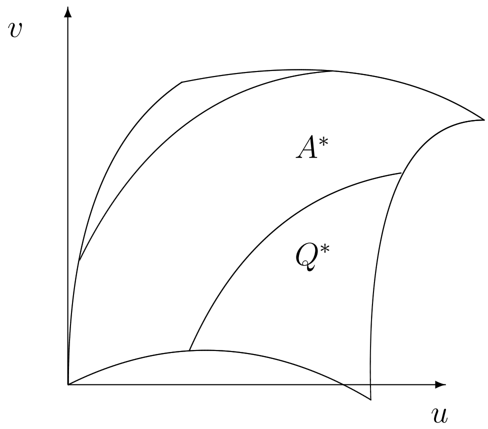
2. Геометрические вероятности. Вероятность того, что точка, наудачу брошенная на отрезок $[a, b]$ попадет в отрезок $[\alpha, \beta]$, $a \leqslant \alpha \leqslant \beta \leqslant b$, $a < b$, равна $\frac{\beta-\alpha}{b-a} = \int_\alpha^\beta \frac{1}{b-a}dx = \int_\alpha^\beta p(x)dx$, где $p(x) = \frac{1}{b-a}$.

Вероятность того, что точка, наудачу брошенная в квадрат Q размером $l \times l$, попадет в заданную область A с площадью S_A равна $\frac{S_A}{l^2}$, в частности, если область A определяется условием $|\xi - \eta| < \lambda$, где ξ и η — координаты точки, то $P(A) = \frac{l^2-(l-\lambda)^2}{l^2} = \int\limits_A \frac{1}{l^2}dxdy = \int\limits_{A^*} p(u, v)dudv$.

[1] Отношение числа исходов, приведших к событию A к числу опытов N, посчитанное *после* проведения эксперимента.

Здесь $(x, y) \to (u, v)$ — невырожденное преобразование, $p(u, v)$ — плотность вероятности.





В этих примерах элементарные события ω — точки, множество Ω — либо отрезок, либо квадрат (образ квадрата).

Алгебра событий. [2]

Рассмотрим отношения между событиями, операции над событиями и некоторые специальные события[3].

1. $B \subset A$ — B влечет A (B — подмножество A).
2. $A = B$, если $B \subset A$ и $A \subset B$.
3. $\overline{A}$ — противоположное событие (A не происходит).
4. $\varnothing$ — невозможное и Ω — достоверное события, $\overline{\varnothing} = \Omega$.
5. $A \cap B$ — произведение (пересечение множеств) — события происходят одновременно.
6. $A \cap B = \varnothing$ события несовместны (не могут произойти одновременно). Например, $A \cap \overline{A} = \varnothing$.
7. $A \cup B$ объединение: происходит либо A, либо B (либо оба). Если они несовместны, то условимся писать $A + B$ (вместо $\bigcup\limits_{i=1}^{n}$ писать $\sum\limits_{i=1}^{n}$).
8. $A \setminus B = \{\omega \in \Omega : \ \omega \in A, \ \omega \notin B\}$ — вычитание.
9. $A \triangle B = (A \setminus B) \cup (B \setminus A)$ — симметрическая разность.

Некоторые свойства операций над событиями.

а) $\overline{\overline{A}} = A, \quad \overline{A} = \Omega \setminus A, \quad A \setminus B = A \cap \overline{B} = \overline{B} \setminus \overline{A}.$

б) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ — взаимная дистрибутивность.

---

[2] Алгебра множеств — это кольцо множеств с единицей. Кольцо множеств — это класс множеств, замкнутый относительно операций $A \cup B$ (объединение) и $A \setminus B$ (вычитание).

[3] Их удобно иллюстрировать с помощью диаграмм Эйлера-Венна.

в) $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$ — принцип двойственности, верен также для любого числа элементов[4].

г) $A \subset B \iff \overline{B} \subset \overline{A}$.

Алгеброй событий называется класс $\mathcal{F}$ событий (система множеств), замкнутый относительно операций $A \cup B$ и $\overline{A}$, то есть

1) из $A, B \in \mathcal{F}$ следует $A \cup B \in \mathcal{F}$,

2) из $A \in \mathcal{F}$ следует $\overline{A} \in \mathcal{F}$.

Пример конечномерного множества Ω. Пусть Ω состоит из n несовместных событий (не считая $\varnothing$ и Ω). Тогда $\mathcal{F}$ — множество всех подмножеств, считая $\varnothing$ и Ω и всего их 2^n.

Классическая вероятность.

Пусть дана полная группа $\{A_i\}$, $i = 1, 2, ..., n$ попарно несовместных равновероятных событий. И пусть некоторое событие A имеет разложение $A = A_{i_1} + ... + A_{i_k}$. В этом случае (по определению) $P(A) = \frac{k}{n}$ (*Докажите корректность определения*).

Иначе: классическая вероятность равна отношению числа благоприятных (для данного события) исходов к числу всех возможных исходов эксперимента. (Речь идет об определении числа исходов *до эксперимента*).

Свойства классической вероятности.

1. $0 \leqslant P(A) \leqslant 1$.

2. $P(\Omega) = 1$, $P(\varnothing) = 0$.

3. $P(A + B) = P(A) + P(B)$. (См. парадокс Р.Мизеса о теннисисте![5])

4. $P(\overline{A}) = 1 - P(A)$.

5. $A \subset B$, $P(A) \leqslant P(B)$, $P(B \backslash A) = P(B) - P(A)$. Следует из п. 3, если учесть $B = A + B \backslash A$.

6. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$, $A_1 \cup A_2 = A_1 + A_2 \backslash (A_1 \cap A_2)$.

Доказать самим формулу для n слагаемых:

$$P(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + ... + (-1)^{n-1} P(A_1 \cap ... \cap A_n).$$

Классическая вероятность применяется в основном в комбинаторных задачах, где подсчитывается число способов выбора (размещения) m объектов из n с или без учета порядка (различимых или неразличимых) и с возвращением (повторением) или без.

Примеры.

а. «Выборка с возвращением, упорядоченная выборка». Найти вероятность того, что хотя бы у одной пары в группе из n человек совпадают дни рождения.

Ответ: $P(A) = 1 - \frac{r(r-1)...(r-n+1)}{r^n}$. Если считать $r = 365$, то[6]

n	5	30	40	50	60
P	0.027	0.706	0.891	0.879	0.994

[4] Формулы де Моргана.

[5] Вероятность выигрыша теннисиста в Лондоне 0.6, в Москве 0.8, причем эти события несовместны. На самом деле эти вероятности нельзя складывать, т.к. события относятся к разным вероятностным пространствам.

[6] В числителе число $A_r^n = \frac{r!}{(r-n)!}$ размещений n дней рождения по r дням года.

б. «Выборка без возвращения, неупорядоченная выборка». Имеется n объектов, среди них k отмеченных. Выбирается (без возвращения) n_1 объектов. Найти вероятность того, что среди них k_1 отмеченных.

$$P(A) = \frac{C_k^{k_1} C_{n-k}^{n_1-k_1}}{C_n^{n_1}}.$$

Это — гипергеометрическое распределение.

в. Еще три примера (из физики). Распределение r частиц по n ячейкам.

Статистика **Максвелла-Больцмана** — все частицы разные, запретов никаких нет. Число состояний n^r, вероятность каждого состояния $p = n^{-r}$.

Статистика **Бозе-Эйнштейна** — частицы тождественны (неразличимы), в каждой ячейке может быть сколько угодно частиц. Число состояний C_{n+r-1}^r, вероятность каждого $p = \frac{1}{C_{n+r-1}^r}$.

Статистика **Ферми-Дирака** — дополнительно к предыдущему действует принцип запрета: в каждой ячейке может быть только одна частица. Число состояний C_n^r, вероятность $p = \frac{1}{C_n^r}$.

. 2 **Аксиомы теории вероятностей.**

Пусть Ω — пространство элементарных событий, $\mathcal{F}$ — алгебра событий (подмножеств Ω). Следующие условия являются аксиомами теории вероятностей:

1. $\mathcal{F}$ является σ-алгеброй, т.е. алгеброй, замкнутой относительно операции счетного объединения событий, $A_n \in \mathcal{F}$, $n = 1, 2, \ldots, \Rightarrow \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$. Пример алгебры, не являющейся σ-алгеброй: полуинтервалы $(a, b]$, $0 < a < b \leqslant 1$ и их конечные системы на $(0, 1]$.

2. На σ-алгебре $\mathcal{F}$ определена функция $P(\cdot)$, принимающая числовые значения, $P(A) \geqslant 0$, $A \in \mathcal{F}$, называемая вероятностью.

3. $P(A + B) = P(A) + P(B)$ — аксиома сложения (верная также в случае конечного разложения, $P(\sum\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$).

4. $P(\sum\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$ — σ-аддитивность.

5. $P(\Omega) = 1$.

Тройка $(\Omega, \mathcal{F}, P(\cdot))$ называется вероятностным пространством (вероятностной моделью).

Аксиомы $1 - 5$ непротиворечивы, поскольку существует пример — классическая вероятность, но не полны, поскольку не задана явно вероятность $P(\cdot)$.

Задание вероятности на σ-алгебре $\mathcal{F}$ элементарно лишь для простейших случаев, например, в случае, когда Ω конечно или счетно. Этот случай мы рассмотрим чуть позже.

Примерами σ-алгебр являются $\mathcal{F}_* = \{\varnothing, \Omega\}$ и $\mathcal{F}^*$ — множество всех подмножеств Ω. Первая из них тривиальна, а вторая — слишком обширна, чтобы на ней можно непротиворечиво задать вероятность. На каких же σ-алгебрах можно задать вероятность и как это сделать? Решение этого вопроса лежит на следующем пути. Пусть $\mathcal{A}$ — произвольная система множеств A_i, $\bigcup A_i = \Omega \in \mathcal{A}$. Ясно, что любые σ-алгебры $\mathcal{F}_1$ и $\mathcal{F}_2$, для которых $\mathcal{A} \subset \mathcal{F}_1, \mathcal{F}_2$, удовлетворяют условию

$\mathcal{F}_* \subset \mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}^*$, причем $\mathcal{F}_1 \cap \mathcal{F}_2$ — тоже σ-алгебра. Пусть $\mathcal{F}_\mathcal{A} = \bigcap\limits_{\alpha:\mathcal{A}\subset\mathcal{F}_\alpha} \mathcal{F}_\alpha$, тогда $\mathcal{F}_\mathcal{A}$ — минимальная σ-алгебра, содержащая $\mathcal{A}$.

Если, например, в качестве $\mathcal{A}$ взять систему интервалов на прямой, то минимальная σ-алгебра, содержащая $\mathcal{A}$, даст σ-алгебру *борелевских множеств* на прямой. Оказывается, что к этой же σ-алгебре приводит система открытых множеств прямой или замкнутых множеств прямой и т.д. Сам процесс построения σ-алгебры называется борелевским замыканием класса $\mathcal{A}$. Дальнейшая идея заключается в том, чтобы задать вероятность на более простом классе множеств $\mathcal{A}$ (например, на полуалгебре)[7], а затем воспользоваться теоремой о единственном продолжении меры на минимальную σ-алгебру, содержащую $\mathcal{A}$. (Мера — функция множества — обобщение понятия длины; вероятность — мера, нормированная на единицу).

Дискретное вероятностное пространство. $\Omega = \{\omega_i\}$ — конечно или счетно, $\mathcal{F}$ — множество всех подмножеств Ω, $P(\cdot)$ достаточно определить для каждого элементарного события $P(\{\omega_i\}) = p_i$, лишь бы $\sum\limits_{i=1}^{\infty} p_i = 1$.

Тогда вероятность для любого события $A \in \mathcal{F}$ равна $P(A) = \sum\limits_{\omega_i \in A} p_i$.

Корректность следует из *Леммы о суммировании по блокам*:

Пусть ряд $\sum\limits_{i=1} c_i = S$ сходится абсолютно и пусть $I = I_1 + I_2 + \ldots$ — разбиение множества натуральных чисел I. Обозначим $S_k = \sum\limits_{i\in I_k} c_i$. Тогда ряд $\sum\limits_{k=1}^{\infty} S_k$ сходится абсолютно и равен S.

Доказательство. Абсолютная сходимость следует из $\sum\limits_{k=1}^{n} |S_k| \leqslant \sum\limits_{k=1}^{n} \sum\limits_{i\in I_k} |c_i|$

Имеем $\forall \varepsilon > 0 \ \exists N_1 = N_1(\varepsilon)$, что при $n \geqslant N_1$: $|\sum\limits_{i=1}^{n} c_i - S| < \varepsilon/2$ и $\sum\limits_{i=n+1}^{\infty} |c_i| < \varepsilon/2$.

Далее выберем N_2 столь большим, чтобы в сумму $\sum\limits_{k=1}^{N_2} S_k = \sum\limits_{k=1}^{N_2} \sum\limits_{i\in I_k} c_i$ входили все числа c_i с номерами $i \leqslant N_1$ и положим $N_0 = \max(N_1, N_2)$.

Тогда при $n \geqslant N_0$ получаем

$$|\sum_{k=1}^{n} S_k - S| \leqslant |\sum_{i=1}^{n} c_i - S| + |\sum_{k=1}^{n} S_k - \sum_{i=1}^{n} c_i| \leqslant |\sum_{i=1}^{n} c_i - S| + |\sum_{i=K_1+1}^{n} |c_i| \leqslant \varepsilon/2 + \varepsilon/2 = \varepsilon. \quad \square$$

Обратное, вообще говоря, неверно, но если $c_i \geqslant 0$, то верно и обратное утверждение.

Пример. Бросание монеты до первого выпадения герба (герб — единица, аверс — ноль). Элементарные события $\omega_1 = \{1\}$, $\omega_2 = \{01\}$, ..., $\omega_\infty = \{00\ldots\}$.

---

[7] Полуалгебра — полукольцо с единицей; полукольцо — непустое множество, замкнутое относительно пересечения, и в котором каждая разность допускает конечное разложение $A \backslash B = \sum\limits_{j=1}^{n} A_j$.

Вероятности: $P(\omega_1) = p_1 = \frac{1}{2}$, $P(\omega_2) = p_2 = \frac{1}{4}$, ..., $P(\omega_\infty) = p_\infty = 0$, $\sum\limits_{i=1}^{\infty} p_i = \sum\limits_{i=1}^{\infty} \frac{1}{2^i} = 1$. Событие ω_∞ возможное, но невероятное!

Аксиоматически определенная вероятность обладает всеми свойствами, которые мы отметили для классической вероятности, поскольку первые три фактически повторяют аксиомы, а остальные выводятся из них.

1. $0 \leqslant P(A) \leqslant 1$.
2. $P(\Omega) = 1, \quad P(\varnothing) = 0$.
3. $P(A + B) = P(A) + P(B)$.
4. $P(\overline{A}) = 1 - P(A)$.
5. $A \subset B \Rightarrow P(A) \leqslant P(B), \quad P(B \backslash A) = P(A) - P(B)$.
6. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$, т.к. $A_1 \cup A_2 = A_1 + (A_2 \backslash (A_1 \cap A_2))$.

Однако, есть еще одно свойство, которое вытекает из σ-аддитивности и называется непрерывностью вероятности, точнее, непрерывностью относительно предельного перехода.

Сначала определим понятие предела последовательности событий (множеств) $\{A_k\}$. Также, как для числовых последовательностей, можно это сделать через верхний и нижний пределы.

$A^* = \varlimsup\limits_{n\to\infty} A_n = \limsup A_n = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$ — событие, заключающееся в том, что произошло бесконечно много событий из $\{A_k\}$.

$A_* = \varliminf\limits_{n\to\infty} A_n = \liminf A_n = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$ — событие, заключающееся в том, что произошли все события из $\{A_k\}$ за исключением, быть может, конечного их числа[8].

Очевидно, что $\varliminf A_n \subset \varlimsup A_n$. Если $\varlimsup A_n = \varliminf A_n$, то говорят, что последовательность событий имеет предел. Для монотонных последовательностей событий: $A_1 \subset A_2 ... \subset A_n ...$ или $B_1 \supset B_2 ... \supset B_n ...$ предел $\lim \uparrow A_n$ ($\lim \downarrow B_n$) всегда существует.

7. $P(\lim A_n) = \lim P(A_n)$ — следствие 4 аксиомы, которую можно переформулировать как непрерывность вероятности относительно монотонных предельных переходов, в частности, $\lim P(A_n) = 0$ если $\lim \downarrow A_n = \varnothing$.

Условная вероятность.

Пусть $(\Omega, \mathcal{F}, P(\cdot))$ — вероятностное пространство. Пусть известно, что в ходе эксперимента произошло событие B ($P(B) > 0$). Естественно после этого сузить множество исходов до $\Omega_B = B$, а вместо любого события A рассматривать $A_B = A \cap B$. Таким образом, мы временно переходим от $(\Omega, \mathcal{F}, P(\cdot))$ к $(\Omega_B, \mathcal{F}_B, P_B(\cdot))$, где $\Omega_B = \Omega \cap B$, $\mathcal{F}_B = \{A \cap B, A \in \mathcal{F}\}$,

$$P_B(A_B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

(Покажите, что $\mathcal{F}_B$ — σ-алгебра, а $P_B(\cdot)$ удовлетворяет аксиомам.)

Можно вернуться к старому вероятностному пространству $(\Omega, \mathcal{F}, P(\cdot))$ и рассматривать вероятность $P_B(\cdot)$ на $\mathcal{F}$. В этом случае ее называют условной

---

[8]Полезное представление $\chi_{A^*} = \varlimsup \chi_{A_n}$, $\chi_{A_*} = \varliminf \chi_{A_n}$, где χ_A — характеристическая функция множества A.

вероятностью и обозначают $P(\cdot|B)$. Событие B можно рассматривать как параметр. Формула (1) легко интерпретируется в терминах классической вероятности, а в общем случае она является определением.

Свойства.

$P(\Omega|B) = 1.\quad P(A_1 + A_2|B) = P(A_1|B) + P(A_2|B).$

$P(\overline{A}|B) = 1 - P(A|B).\quad P(\sum\limits_{i=1}^{\infty} A_i|B) = \sum\limits_{i=1}^{\infty} P(A_i|B.)$

Эти свойства полезны для решения задач. **Пример**: Вероятность аварии ракеты 0.1, причем на старте 0.09. Какова вероятность аварии в случае успешного старта? Пусть событие A — авария, B — авария на старте, $B \subset A$, $\overline{A} \subset \overline{B}$. Искомая вероятность $P(A|\overline{B}) = 1 - P(\overline{A}|\overline{B}) = 1 - \frac{P(\overline{A}\cap\overline{B})}{P(\overline{B})} = 1 - \frac{P(\overline{A})}{P(\overline{B})} = 1 - \frac{1-0.1}{1-0.09} = \frac{1}{91}$.

Теорема умножения вероятностей.

$P(A\cap B) = P(A|B)P(B)$, $P(B) > 0$.

Для трех событий

$P(A\cap B\cap C) = P(A|B\cap C)P(B|C)P(C) = P(A\cap B|C)P(C)$, отсюда

$P(A\cap B|C) = P(A|B\cap C)P(B|C)$.

Независимость. *События A и B независимы*[9]*, если* $P(A\cap B) = P(A)\cdot P(B)$.

Если $P(A) = 0$ (или $P(A) = 1$), то A и B независимы.

Если $P(A) > 0$, то $P(B|A) = P(B)$ означает независимость A и B, но в общем случае это не эквивалентно определению независимости.

Свойства.

1. A и Ω независимы.

2. A и B, если $P(A) = 0$, независимы.

3. Если A и B независимы, то A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$ — также независимы (доказать самостоятельно).

4. Если A и B_i попарно независимы $i = 1, 2, \ldots, n$, то независимы A и $\sum B_i$ (но не $\bigcup B_i$).

Формула полной вероятности. Формулы Байеса.

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, P(\cdot))$.

Пусть $\{A_i,\ \ i = 1.2.\ldots\}$ — полная группа (необязательно конечная) попарно несовместных (необязательно равновероятных) событий, $A_i\cap A_j = \varnothing$, $i \neq j$ и пусть $B \in \mathcal{F}$ — некоторое событие, $P(B) > 0$. Тогда

$B = B\cap(\sum\limits_i A_i)$ и $P(B) = \sum\limits_i P(B|A_i)P(A_i)$ — формула полной вероятности.

Из $P(B\cap A_k) = P(A_k|B)P(B) = P(B|A_k)P(A_k)$ следует

$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum\limits_i P(B|A_i)P(A_i)}$ — формулы Байеса. □

Независимость в совокупности. События $\{A_i\}$ независимы в совокупности, если для любых m и любых наборов различных индексов $i_1, i_2, \ldots, i_m$ имеет место равенство $P(\bigcap\limits_{k=1}^{m} A_{i_k}) = \prod\limits_{k=1}^{m} P(A_{i_k})$. Лекц. 3

[9] стохастически независимы.

Задача (Пример Бернштейна). Пусть правильный тетраэдр раскрашен так, что на трех его гранях красный, синий и зеленый цвет соответственно, а на четвертой — все три цвета. Проверьте, что события «выпадение разных цветов» попарно независимы, но не независимы в совокупности.

Пример построения независимых событий.

Пусть $(\Omega_1, \mathcal{F}_1, P_1(\cdot))$, $(\Omega_2, \mathcal{F}_2, P_2(\cdot))$ — дискретные (для простоты) вероятностные пространства, $\omega_i^1 \in \Omega_1$, $\omega_j^2 \in \Omega_2$, $P_1(\omega_i^1) = p_i^1$, $P_2(\omega_j^2) = p_j^2$. Рассмотрим множество упорядоченных пар $\{\omega_i^1\omega_j^2\}$, обозначим его $\Omega = \Omega_1 \times \Omega_2$ — прямое произведение. Множество всевозможных подмножеств Ω, которое, очевидно, σ-алгебра, обозначим $\mathcal{F}$. Наконец, введем вероятность: $P(\{\omega_i^1\omega_j^2\}) = p_{ij} = P_1(\omega_i^1)P_2(\omega_j^2) = p_i^1 p_j^2$ (проверить корректность!). Полученное вероятностное пространство назовем прямым произведением: $(\Omega, \mathcal{F}, P(\cdot)) = (\Omega_1, \mathcal{F}_1, P_1(\cdot)) \times (\Omega_2, \mathcal{F}_2, P_2(\cdot))$.

Пусть в этой схеме $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$,

$$A = \sum_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in \Omega}} \{\omega_i^1\omega_j^2\} \text{ и } B = \sum_{\substack{i:\omega_i^1 \in \Omega \\ j:\omega_j^2 \in A_2}} \{\omega_i^1\omega_j^2\}, \quad A, B \in \mathcal{F}.$$

Найдем вероятность

$$P(A) = \sum_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in \Omega}} p_{ij} = \sum_{i:\omega_i^1 \in A_1} p_i^1 \sum_{j:\omega_j^2 \in \Omega} p_j^2 = \sum_{i:\omega_i^1 \in A_1} p_i^1.$$

Аналогично, $P(B) = \sum\limits_{j:\omega_j^2 \in A_2} p_j^2$.

Наконец, $P(A \cap B) = \sum\limits_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in A_2}} p_{ij} = \sum\limits_{i:\omega_i^1 \in A_1} p_i^1 \sum\limits_{j:\omega_j^2 \in A_2} p_j^2 = P(A)P(B)$,

откуда следует их независимость. □

Определение 1. Пусть $(\Omega, \mathcal{F}, P(\cdot))$ — дискретное вероятностное пространство. Последовательностью независимых испытаний называется вероятностное пространство $(\Omega_n, \mathcal{F}_n, P_n(\cdot))$, которое является прямым произведением n одинаковых пространств (n-й степенью): $(\Omega, \mathcal{F}, P(\cdot))$, т.е. $(\Omega_n, \mathcal{F}_n, P_n(\cdot)) = \overset{n}{\times} (\Omega, \mathcal{F}, P(\cdot))$. Подробнее: Ω_n состоит из цепочек $(\omega_{i_1}\omega_{i_2}\ldots\omega_{i_n})$ длины n с необязательно различными индексами, $\mathcal{F}_n$ — алгебра подмножеств Ω_n, $P_n(\omega_{i_1}\omega_{i_2}\ldots\omega_{i_n}) = p_{i_1}p_{i_2}\ldots p_{i_n}$.

Определение 2. Схемой Бернулли называется последовательность n независимых испытаний, полученная на основе вероятностного пространства, в котором содержится лишь два элементарных события (исхода): ω_1 — успех и ω_2 — неудача (1 и 0 соответственно). Обозначив $P(\omega_1) = p$, $P(\omega_2) = q = 1 - p$, получим $P_n(\omega_{i_1}\omega_{i_2}\ldots\omega_{i_n}) = p^k q^{n-k}$, где k — число успехов в серии из n испытаний Бернулли.

Какова вероятность при n испытаниях получить k успехов? Это $p_n(k) = C_n^k p^k q^{n-k}$ — биномиальное распределение, $\sum\limits_{k=0}^{n} p_n(k) = 1$.

На это распределение очень много задач и вы с ними познакомитесь на семинарских занятиях.

Отрицательное биномиальное распределение (распределение Паскаля):

Вероятность того, что для достижения n успехов в схеме Бернулли потребуется $n+k$ испытаний равна $p(n,n+k) = C_{n+k-1}^k p^n q^k$.

$$\sum_{k=0}^{\infty} p(n,n+k) = \sum_{k=0}^{\infty} \frac{(n+k-1)(n+k-2)\dots(n)}{k!} p^n q^k =$$
$$= \sum_{k=0}^{\infty} \frac{(-n)(-n-1)\dots(-n-k+1)}{k!} p^n (-q)^k = p^n (1-q)^{-n} = 1.$$

Пусть $\Omega = \{\omega_1 \omega_2 \dots \omega_r\}$. Тогда

$$p_n(s_1, s_2, \dots, s_r) = \frac{n!}{s_1! s_2! \dots s_r!} p_1^{s_1} p_2^{s_2} \dots p_r^{s_r}, \quad \sum_i^r s_i = n$$

— полиномиальное распределение (общий член разложения полинома $(p_1 + p_2 + \dots + p_r)^n$).

Предельные теоремы (для биномиального распределения).

Теорема Пуассона.

Рассмотрим последовательность биномиальных распределений. Пусть $p = \frac{\lambda}{n}$ (т.е. $\lambda = pn = Const$).

Тогда при $n \to \infty$: $p_n(k) \Longrightarrow p_\infty(k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

Доказательство.

$$p_n(k) = C_n^k p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} =$$
$$= \frac{1(1-\frac{1}{n})\dots(1-\frac{k-1}{n})}{(1-\frac{\lambda}{n})^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \to \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square$$

Замечание. Теорема, очевидно, справедлива и в случае, когда величина pn не постоянна, но стремится к λ.

Теорему практически можно применять[10] при $n \geqslant 100$, $\lambda \leqslant 10$.

Теорема Муавра-Лапласа. (В этом случае npq велико, так что $p > 0$ и $q > 0$ — фиксированны).

Локальная.

$$\lim_{n\to\infty} \frac{\sqrt{npq}\ \ p_n(m)}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}}} = 1, \quad p_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x_m^2}{2}}, \quad x_m = \frac{m-np}{\sqrt{npq}}.$$

Интегральная. Эта теорема будет получена как следствие из более общей, так называемой Центральной Предельной Теоремы.

$$P\left\{a \leqslant \frac{m-np}{\sqrt{npq}} < b\right\} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

А локальную можно вывести из интегральной следующим образом.

[10] Аккуратную оценку погрешности можно посмотреть в книге *Боровков А.А.* Теория вероятностей. — М.: Наука, 1976.

Положим $a = \frac{k-np}{\sqrt{npq}}$, $b = \frac{k+1-np}{\sqrt{npq}}$, так что $b - a = \frac{1}{\sqrt{npq}}$.
Тогда, используя теорему о среднем, получим

$$p_n(k) = P\{m = k\} = P\left\{a \leqslant \frac{m-np}{\sqrt{npq}} < b\right\} \approx$$
$$\approx \frac{1}{\sqrt{2\pi}}(b-a)e^{-\frac{x^2}{2}} \,|_{x\in(a,b)} \approx \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{npq}}e^{-\frac{x_k^2}{2}}. \quad \square$$

Замечание. Если $n^k p \to \lambda$, где $0 < k < 1$, то можно применять обе приближенные формулы.

Пример. Сколько раз нужно бросить монету, чтобы с вероятностью 0.99 частота появления герба отличалась от $\frac{1}{2}$ не более, чем на 0.01?

Решение. Согласно интегральной теореме Муавра-Лапласа

$$P\left(\left|\frac{m}{n} - p\right| < 0.01\right) = P\left(\left|\frac{m}{n} - p\right| < \sqrt{\frac{pq}{n}}\varepsilon\right) = P\{\left|\frac{m-np}{\sqrt{npq}}\right| < \varepsilon\} =$$
$$\approx \frac{1}{\sqrt{2\pi}}\int\limits_{-\varepsilon}^{\varepsilon} e^{-\frac{x^2}{2}}dx = 1 - 2\Phi(-\varepsilon) = 0.99.$$

Здесь $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$ — интеграл ошибок (табулирован).

$\varepsilon = \sqrt{\frac{n}{pq}}0.01 = 2.58$ (из таблиц), откуда $n = (129)^2 = 16641$. $\square$

. 4 **Случайная величина.**

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, P(\cdot))$.

1. Случайной величиной называется вещественная функция $\xi(\omega)$ (отображение $\Omega \mapsto \mathcal{R}_1$), такая, что для любого вещественного x множество $\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{F}$, т.е. событие.

2. Вероятность $P(\{\xi(\omega) < x\}) = F_\xi(x)$ называется функцией распределения случайной величины ξ.

Свойства функции распределения.

1. Для $x_1 < x_2$: $\{\xi < x_2\} = \{\xi < x_1\} + \{x_1 \leqslant \xi < x_2\} \Rightarrow P(\xi < x_2) = P(\xi < x_1) + P(x_1 \leqslant \xi < x_2) \Rightarrow F(x_1) \leqslant F(x_2)$, $P(x_1 \leqslant \xi < x_2) = F(X_2) - F(x_1)$.

2. $0 \leqslant F(x) \leqslant 1$, $x \in \mathcal{R}_1$.

3. $F(x)$ непрерывна слева, $\lim\limits_{x_k \uparrow x} F(x_k) = F(x)$. Пусть $x_1 < x_2 < \cdots < x_n < \cdots < x$. Доказательство следует из того, что $\{\xi < x\} = \bigcup\limits_k \{\xi < x_k\}$ и $F(x) = \lim\limits_{x_k \uparrow x} F(x_k)$ в силу непрерывности вероятности для монотонных последовательностей событий.

4. $P\{\xi \leqslant x\} = \lim\limits_{x_k \downarrow x} F(x_k) = F(x+0)$. В этом случае для $x_1 > x_2 > \cdots > x_n > \cdots > x$ имеем $\{\xi \leqslant x\} = \bigcap\limits_k \{\xi < x_k\}$ и далее используем непрерывность вероятности для монотонных последовательностей событий.

5. Выпишем соотношения для $x_1 \leqslant x_2$:

$P\{x_1 \leqslant \xi < x_2\} = F(x_2) - F(x_1)$,

$P\{x_1 \leqslant \xi \leqslant x_2\} = F(x_2+0) - F(x_1)$, в частности, при $x_1 = x_2 = x$: $P\{\xi = x\} = F(x+0) - F(x)$,

$P\{x_1 < \xi \leqslant x_2\} = F(x_2+0) - F(x_1+0)$,

$P\{x_1 < \xi < x_2\} = F(x_2) - F(x_1+0)$.

6. $F(-\infty) = \lim\limits_{n\to\infty} F(-n) = 0$, $F(+\infty) = \lim\limits_{n\to\infty} F(+n) = 1$.

Существование пределов следует из монотонности и ограниченности. Равенства — из соотношения $1 = P\{-\infty < \xi < \infty\} = P(+\infty) - P(-\infty)$.

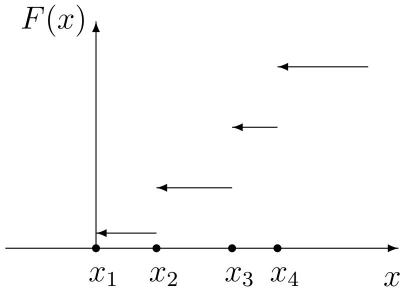
Отметим еще, что функция распределения имеет не более, чем счетное число скачков. (Число $N(\frac{1}{k})$ скачков размером не более $\frac{1}{k}$ ограничено величиной k).

Функция, удовлетворяющая условиям $1-6$ является функцией распределения (некоторой случайной величины). Однако, из равенства $F_\xi(x) = F_\eta(x)$ не следует совпадения ξ и η; они могут различаться с вероятностью единица.

I. Дискретные случайные величины.

Множество значений конечно или счетно. $P\{\xi = x_k\} = p_k$, $\sum p_k = 1$. Для любого подмножества A значений $\{x_k\}$ имеем $p\{\xi \in A\} = \sum\limits_{k:\, x_k\in A} p_k$. Корректность следует из Леммы о суммировании по блокам.

Функция распределения $F(x) = \sum\limits_{k:\, x<x_k} p_k$ полностью определяет распределение: значения x_k — точки разрыва , вероятности p_k — величины скачков.



Функция распределения

II. Случайная величина ξ называется непрерывной (абсолютно непрерывной), если $F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(x)dx$, где $p_\xi(x)$ — неотрицательная кусочно-непрерывная функция, которая называется плотностью вероятности (плотностью распределения вероятности) случайной величины ξ, $\int\limits_{-\infty}^{\infty} p_\xi(x)dx = 1$.

В точках непрерывности $p_\xi(x) = \frac{dF_\xi(x)}{dx}$.

$P\{x_1 \leqslant \xi < x_2\} = F(x_2) - F(x_1) = \int\limits_{x_1}^{x_2} p_\xi(x)dx$.

Замечания.

1. Интеграл понимается в смысле Лебега (в простейших случаях он совпадает с Римановским).

2. Для любого Борелевского (измеримого) множества A имеем $P(A) = \int\limits_{A} dF(x)$

3. Если $p(x)$ непрерывна на $[x, x+\Delta x]$, то по теореме о среднем $P\{x \leqslant \xi < x + \Delta x\} = p(x) + o(\Delta x)$.

4. $P\{\xi = x\} = F(x+0) - F(x) = 0$ для непрерывных случайных величин, поэтому для них неравенства $\leqslant, \geqslant$ можно заменить на $<, >$ и наоборот.

III. Сингулярные случайные величины. Для них множество точек роста функции распределения имеет (Лебегову) меру нуль. Пример — Канторовская лестница. Общая длина отрезков стационарности стремится к $\sum\limits_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1$.

Показано, что случайные величины исчерпываются этими тремя типами, но существуют смешанные типы: $F(x) = \alpha F_{\xi_1}(x) + (1-\alpha)F_{\xi_2}(x)$, $0 < \alpha < 1$.

Многомерные случайные величины. Пусть задано $(\Omega, \mathcal{F}, P(\cdot))$. Векторной случайной величиной называется векторная функция $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega) \ldots, \xi_n(\omega))$, такая, что для любых $x_1, x_2, \ldots, x_n$: $\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \ldots, \xi_n(\omega) < x_n\} \in \mathcal{F}$. $P\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\} = F(x_1, x_2, \ldots, x_n)$ — n-мерная функция распределения.

Далее рассмотрим случай $n = 2$.

$F(x, y) = P\{\xi < x, \eta < y\}$.

Свойства.

1. $F(x, y)$ не убывает по x и по y.
2. $F(x, y)$ непрерывна слева по каждому аргументу.
3. $F(+\infty, +\infty) = 1$, $F(-\infty, y) = F(x, -\infty) = 0$.
4. $P\{x_1 \leqslant \xi < x_2, y_1 \leqslant \eta < y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + +F(x_1, y_1)$.
5. $F_\xi(x) = F(x, +\infty)$, $F_\eta(y) = F(+\infty, y)$ — *маргинальные распределения*.

$$\{\xi < x\} = \sum_{k=-\infty}^{\infty} \{\xi < x, k \leqslant \eta < k+1\} \quad \Rightarrow$$

$$P\{\xi < x\} = \sum_{k=-\infty}^{\infty} [F(x, k+1) - F(x, k)] = F(x, \infty).$$

Для абсолютно непрерывных случайных величин $F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(x, y) dx dy$ и далее:

1. Плотность $p(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$ в точках непрерывности.

2. $P(\xi \in D) = \int\limits_D p(x) dx$ ($\xi, x \in \mathcal{R}_n$, $D \subset \mathcal{R}_n$).

3. $P\{x \leqslant \xi < x + \Delta x, y \leqslant \eta < y + \Delta y\} = p(x, y)\Delta x \Delta y + o(\Delta x \Delta y)$ в окрестности точки непрерывности.

4. Маргинальные плотности $p_\xi(x) = \int\limits_{-\infty}^{\infty} p(x, y) dy, \quad p_\eta(y) = \int\limits_{-\infty}^{\infty} p(x, y) dx$.

Условные распределения.

Пусть — событие, $P\{B\} \neq 0$. Тогда $P\{\xi < x | B\} = F_\xi(x|B) = \int\limits_{-\infty}^{x} p_\xi(x|B) dx$.

Пусть задана $p(x,y)$ — плотность распределения сл. величины (ξ,η). Рассмотрим $F(\xi<x|B)$, где $B=\{y\leqslant\eta<y+\Delta y\}$:

$$F(\xi<x,B)=\frac{P\{\xi<x,y\leqslant\eta<y+\Delta y\}}{P\{y\leqslant\eta<y+\Delta y\}}=\frac{\int\limits_{-\infty}^{x}dx\int\limits_{y}^{y+\Delta y}p(x,y)dy}{\int\limits_{-\infty}^{\infty}dx\int\limits_{y}^{y+\Delta y}p(x,y)dy}=$$

$$=\frac{\int\limits_{-\infty}^{x}p(x,y)dx\Delta y+o(\Delta y)}{p_\eta(y)\Delta y+o(\Delta y)}.$$

Разделив на Δy и переходя к пределу при $\Delta y\to 0$, получаем

$$F_{\xi|\eta}(x|y)=\frac{\int\limits_{-\infty}^{x}p(x,y)dx}{p_\eta(y)}\qquad\text{и}\qquad p_{\xi|\eta}(x|y)=\frac{p(x,y)}{p_\eta(y)}.$$

Наконец, на основе этой формулы можно записать

$$p_\xi(x)=\int\limits_{-\infty}^{\infty}p_{\xi|\eta}(x|y)p_\eta(y)dy.$$

(аналог формулы полной вероятности).

Назависимость случайных величин. Лекц. 5

Случайные величины $\xi_1,...,\xi_n$ называются независимыми в совокупности (попарно), если $\forall x_1,...,x_n$ события $\{\xi_1<x_1\}$, ..., $\{\xi_n,<x_n\}$ независимы в совокупности (попарно). Пусть теперь $n=2$.

$$F(x,y)=P\{\xi<x,\eta<y\}=P\{\xi<x\}\ P\{\eta<y\}=F_\xi(x)\ F_\eta(y). \tag{2}$$

$$\begin{aligned}&P\{x_1\leqslant\xi<x_2,y_1\leqslant\eta<y_2\}=\\&=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)=\\&=F_\xi(x_2)F_\eta(y_2)-F_\xi(x_1)F_\eta(y_2)-F_\xi(x_2)F_\eta(y_1)+F_\xi(x_1)F_\eta(y_1)=\\&=(F_\xi(x_2)-F_\xi(x_1))\ (F_\eta(y_2)-F_\eta(y_1))=\\&=P\{x_1\leqslant\xi<x_2\}\ P\{y_1\leqslant\eta<y_2\}.\end{aligned}\tag{3}$$

Формулы (2) и (3) эквивалентны $(x_1,y_1\to\infty)$. Для дискретных сл. величин $P(\xi=x_k,\eta=y_j)=P(\xi=x_k)P(\eta=y_j)$. Впрочем индексы можно опустить.

$p_{\xi\eta}(x,y)=p_\xi(x)p_\eta(y)$ в точках непрерывности этих функций.

Вернемся к (2). Верно ли обратное?

Теорема. Пусть $F_{\xi\eta}(x,y)=F_1(x)\cdot F_2(y)$, причем $F_1(+\infty)=1$. Тогда $F_1(x)=F_\xi(x)$, $F_2(y)=F_\eta(y)$.

Доказательство следует из свойств двумерной функции распределения.

Случайные величины ξ, η называются эквивалентными, если $P(\xi\neq\eta)=0$ (совпадают почти всюду).

Функции от случайных величин.

Пусть $\eta = f(\xi)$, где $f(\cdot)$ — измеримая функция, $\mathcal{R}_n \mapsto \mathcal{R}_m$. Обычно требуется найти распределение η, если известно распределение ξ.

Пример. Пусть $y = f(x)$ — невырожденное преобразование (биекция). Пусть дана плотность $p_\xi(x)$. Найти плотность $p_\eta(y)$, если $\eta = f(\xi)$.

Для произвольной области имеем:

$$\int\limits_D p_\eta(y)dy = P(\eta \in D) = P(f(\xi) \in D) = P(\xi \in f^{-1}(D)) =$$

$$= \int\limits_{f^{-1}(D)} p_\xi(x)dx = \int\limits_D p_\xi(f^{-1}(y)) \left|\frac{\partial(x)}{\partial(y)}\right| dy.$$

Отсюда в силу произвольности D получаем, что

$p_\eta(y) = p_\xi(f^{-1}(y))|\frac{\partial(x)}{\partial(y)}|dy$.

Пример. Задана плотность $p_\xi(x_1, x_2)$. Найти распределение $\xi_1 + \xi_2$.

Пусть

$$\begin{array}{llllll} \eta_1 = & \xi_1+\xi_2, & y_1 = & x_1+x_2, & x_1 = & y_1 - y_2, \\ \eta_2 = & \xi_2, & y_2 = & x_2, & x_2 = & y_2, \end{array} \qquad |\tfrac{\partial(x)}{\partial(y)}| = 1.$$

Тогда $p_\eta(y) = p_\xi(y_1 - y_2, y_2)$ и

$p_{\eta_1}(y_1) = \int\limits_{-\infty}^{\infty} p_\xi(y_1 - y_2, y_2)dy_2$.

Если ξ и η независимы, то $p_{\eta_1}(y_1) = \int\limits_{-\infty}^{\infty} p_{\xi_1}(y_1 - y_2)p_{\xi_2}(y_2)dy_2$ (свертка!).

Независимость функций от независимых сл. величин. Рассмотрим дискретный случай. Пусть ξ и η — независимые сл. величины, тогда $f_1(\xi)$ и $f_2(\eta)$ (где f_1 и f_2 — измеримые функции) также независимые сл. величины.

Действительно, пусть ξ принимает значения x_i, а η — значения y_j. Для любых s и t имеем:

$$P\{f_1(\xi) = s, f_2(\eta) = t\} = \sum_{\substack{i:f_1(x_i)=s \\ j:f_2(y_j)=t}} P\{\xi = x_i, \eta = y_j\} =$$

$$= \sum_{i:f_1(x_i)=s} P\{\xi = x_i\} \cdot \sum_{j:f_2(y_j)=t} P\{\eta = y_j\} = P\{f_1(\xi) = s\} \cdot P\{f_2(\eta) = t\}. \quad \square$$

Числовые характеристики случйных величин. Моменты.

Начальный момент порядка k:

$\mathsf{M}\xi^k = \int\limits_{-\infty}^{\infty} x^k p_\xi(x)dx$, если $\int\limits_{-\infty}^{\infty} |x|^k p_\xi(x)dx < +\infty$ или

$\mathsf{M}\xi^k = \sum\limits_{i=1}^{\infty} x_i^k p_i$, если $\sum\limits_{i=1}^{\infty} |x_i|^k p_i < +\infty$.

Математическое ожидание или среднее значение: $\mathsf{M}\xi = \int\limits_{-\infty}^{\infty} xdF_\xi(x)$.

Теорема. Пусть сл. величина $\eta = f(\xi)$. Тогда[11] $\mathsf{M}\eta = \int\limits_{-\infty}^{\infty} f(x)dF_\xi(x)$.

[11] если момент существует

Доказательство (для дискретных случайных величин)

$$\mathsf{M}\eta = \sum_{k=1}^{\infty} y_k P(\eta = y_k) = \sum_k y_k \sum_{i:f(x_i)=y_k} p_i = \sum_i f(x_i)p_i.$$

Следствие: $\mathsf{M}(\sum C_k\xi_k) = \sum C_k \mathsf{M}\xi_k$.

Центральный момент k-го порядка $\mathsf{M}(\xi - \mathsf{M}\xi)^k = \int\limits_{-\infty}^{\infty} (x - \mathsf{M}\xi)^k dF_\xi(x)$.

Центральный момент второго порядка $\mathsf{M}(\xi - \mathsf{M}\xi)^2 = \int\limits_{-\infty}^{\infty} (x - \mathsf{M}\xi)^2 dF_\xi(x) = \mathsf{D}\xi$ носит название дисперсии, $\sqrt{\mathsf{D}\xi}$ — стандартное отклонение.

$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2$.

Свойства математического ожидания и дисперсии.

1. $\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta =$

$$= \int\limits_{-\infty}^{\infty} zp_{\xi+\eta}(z)dz = \int\limits_{-\infty}^{\infty} (x+y) \int\limits_{-\infty}^{\infty} p(x,y)dxdy = \iint\limits_{-\infty}^{\infty} xp(x,y)dxdy +$$

$$+ \iint\limits_{-\infty}^{\infty} yp(x,y)dxdy = \int\limits_{-\infty}^{\infty} xp_\xi(x)dx + \int\limits_{-\infty}^{\infty} yp_\eta(y)dy.$$

(Аналогично для дискретных сл. величин.)

2. $\mathsf{M}(C_1\xi + C_2\eta) = C_1\mathsf{M}\xi + C_2\mathsf{M}\eta$ и $\mathsf{M}A\xi = A\mathsf{M}\xi$ (A — матрица, ξ — сл. вектор).

3. Если ξ и η независимы, то $\mathsf{M}(\xi\eta) = \mathsf{M}\xi\mathsf{M}\eta$. (При $\mathsf{M}\xi < \infty$, $\mathsf{M}\eta < \infty$. Заметим, что $\mathsf{M}(\xi\eta)$ в этом случае существует.)

Обратное неверно! Пусть, например, ξ и η — независимые сл. величины и $\mathsf{M}\eta = \mathsf{M}\xi = 0$. Тогда $\mathsf{M}(\xi\eta)\xi = \mathsf{M}\xi^2\eta = 0$. но отсюда не следует, что $\xi\eta$ и η независимы. В самом деле,

$\eta\backslash\xi$	-1	2
-1	4/9	2/9
2	2/9	1/9

,

но $P\{\xi\eta = 1, \xi = -1\} = \frac{4}{9} \neq P\{\xi\eta = 1\}P\{\xi = -1\} = \frac{8}{27}$.

Примеры моментов

Пусть ξ_i — число успехов в i-м испытании Бернулли. $\mathsf{M}\xi_i = 1 \cdot p + 0 \cdot q = p$, $\mathsf{M}\xi^2 = 1^2 \cdot p + 0^2 \cdot q = p$, $\mathsf{D}\xi_i = p - p^2 - pq$.

Распределение Пуассона

$M\xi = \sum\limits_{k=0}^{\infty} k\frac{\lambda^k}{n!}e^{-\lambda} = \lambda \sum\limits_{k=1}^{\infty} k\frac{\lambda^{k-1}}{n!}e^{-\lambda} = \lambda e^{-\lambda}\frac{d}{d\lambda}e^\lambda = \lambda$. Аналогичным приемом получаем $\mathsf{M}\xi^2 = \lambda + \lambda^2$ и $\mathsf{D}\xi = \lambda$.

Нормальное распределение $\mathcal{N}(\mu, \sigma^2)$:

$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$.

$$\mathsf{M}\xi = \int\limits_{-\infty}^{\infty} xp(x)dx = \int\limits_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}dx =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} (y+\mu)\exp\{-\frac{1}{2\sigma^2}y^2\}dy = \mu.$$

Аналогично

$$\mathsf{D}\xi = \int\limits_{-\infty}^{\infty} (x-\mu)^2\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}dx = \frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} t^2e^{-\frac{t^2}{2}}dt = \sigma^2.$$

4. Неравенства.

а/ $\xi \geqslant \eta \Rightarrow \mathsf{M}\xi \geqslant \mathsf{M}\eta$ (Следствие 2.)

б/ Нер-во Коши-Буняковского

$\mathsf{M}(\xi-\lambda\eta)^2 = \mathsf{M}\xi^2 - 2\lambda\mathsf{M}\xi\eta + \lambda^2\mathsf{M}\eta^2 \geqslant 0 \Rightarrow (\mathsf{M}\xi\eta)^2 \leqslant \mathsf{M}\xi^2\mathsf{M}\eta^2$.

Если равенство, то $\exists\lambda$, $\mathsf{M}(\xi-\lambda\eta)^2 = 0$.

в/ Неравенство Маркова.

$$P\{|\xi| > \varepsilon\} \leqslant \frac{\mathsf{M}|\xi|^k}{\varepsilon^k},\ k = 1, 2, ...$$

Пусть

$$\eta = \begin{cases} 0, & |\xi| \leqslant \varepsilon, \\ \varepsilon, & |\xi| > \varepsilon. \end{cases}$$

Тогда $|\eta|^k \leqslant |\xi|^k \Rightarrow \varepsilon^k P\{|\xi| > \varepsilon\} \leqslant \mathsf{M}|\xi|^k$.

При $k = 2$ — неравенство Чебышёва[12].

5. $\mathsf{D}C = 0$. Обратно: если $\mathsf{D}\xi = 0$, то $P\{\xi = \mathrm{const}\} = 1$. Доказательство. $\forall\varepsilon > 0$ $P\{|\xi| > \varepsilon\} = 0$. Рассмотрим последовательность $A_k = \{|\xi - \mathsf{M}\xi| > \frac{1}{k}\}$. $A_k \uparrow A = \{|\xi - \mathsf{M}\xi| > 0\} \Rightarrow P\{|\xi - \mathsf{M}\xi| > 0\} = \lim\limits_{k\to\infty} P\{|\xi - \mathsf{M}\xi| > \frac{1}{k}\} = 0$.

Возвращаясь к неравенству Коши: если равенство, то $\exists\lambda$, такое, что $\xi - \lambda\eta = const$ с вероятностью 1 (почти наверное).

6. $\mathsf{D}C\xi = C^2\mathsf{D}\xi$.

7. $\mathsf{D}\sum\limits_{i=1}^{n}\xi_i = \mathsf{M}\left[\sum\limits_{i=1}^{n}(\xi_i - \mathsf{M}\xi_i)\right]^2 = \mathsf{M}\sum\limits_{i=1}^{n}(\xi_i - \mathsf{M}\xi_i)^2 + \sum\limits_{i,j=1, i\neq j}^{n}\mathsf{M}(\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j) = \sum\limits_{i=1}^{n}\mathsf{D}\xi_i + \sum\limits_{i,j=1, i\neq j}^{n}\mathrm{cov}\xi_i\xi_j$

($=\sum\limits_{i=1}^{n}\mathsf{D}\xi_i$ если ξ_i, ξ_j попарно независимы — формула Бьенемэ.)

Моменты векторных случайных величин.

$\mathsf{M}\xi = (\mathsf{M}\xi_1, \mathsf{M}\xi_2, ..., \mathsf{M}\xi_n)$, $\mathsf{M}A\xi = A\mathsf{M}\xi$,

$\mathsf{D}\xi = (\mathsf{D}\xi_1, \mathsf{D}\xi_2, ..., \mathsf{D}\xi_n)$

[12] Пафнутий Львович Чебышёв 1821–1894 — русский математик, создатель Петербургской научной школы.

$\|\mathrm{cov}\xi_i\xi_j\|$ — ковариационная матрица. Введем $r_{ij} = \frac{\mathrm{cov}\xi_i\xi_j}{\sqrt{\mathsf{D}\xi_i\mathsf{D}\xi_j}}$, $|r_{ij}| \leqslant 1$.

Каков смысл r_{ij}? Пусть $|r_{ij}| = 1$ и $\varepsilon = \pm 1$, тогда $\mathsf{M}\left(\frac{\xi_i}{\sqrt{\mathsf{D}\xi_i}} + \varepsilon\frac{\xi_j}{\sqrt{\mathsf{D}\xi_j}}\right)^2 = 2(1+\varepsilon r_{ij}) = 0$ и величины ξ_i и ξ_j линейно зависимы с вероятностью 1.

Условное математическое ожидание

$$\mathsf{M}(\xi|\eta = y) = \int\limits_{-\infty}^{\infty} x p_{\xi|\eta}(x|y)dx.$$

Это случайная величина, функция от η. Она обладает всеми свойствами математического ожидания с вероятностью 1.

Если учесть, что $p_{\xi|\eta}(x|y) = \frac{p_{\xi\eta}(x,y)}{p_\eta(y)}$, то

$$\mathsf{M}[\mathsf{M}(\xi|\eta = y)] = \int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{\infty} x p_{\xi|\eta}(x|y)dx\right] p_\eta(y)dy =$$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x p_{\xi\eta}(x,y)dxdy = \mathsf{M}\xi.$$

Задача о наилучшем среднеквадратичном приближении.

Пусть ξ и η — сл. величины, наблюдаем ξ, требуется оценить η, т.е. найти такую функцию $f(\cdot)$, что $\mathsf{M}(\eta - f(\xi))^2 \sim \min\limits_{f(\cdot)}$

Теорема. $f(\xi) = \mathsf{M}(\eta|\xi)$ (п.н.).

Доказательство. $\mathsf{M}(\eta - f(\xi))^2 = \mathsf{M}[\mathsf{M}((\eta - f(\xi))^2|\xi)] =$

$$= \mathsf{M}[\mathsf{M}((\eta - \mathsf{M}(\eta|\xi) + \mathsf{M}(\eta|\xi) - f(\xi))^2|\xi)] =$$
$$= \mathsf{M}\{\mathsf{M}((\eta - \mathsf{M}(\eta|\xi))^2|\xi) + 2\mathsf{M}((\eta - \mathsf{M}(\eta|\xi))(\mathsf{M}(\eta|\xi) - f(\xi))|\xi) +$$
$$+ (\mathsf{M}(\eta|\xi) - f(\xi))^2|\xi)\} = \mathsf{M}(\eta - \mathsf{M}(\eta|\xi))^2 + \mathsf{M}(\mathsf{M}(\eta|\xi) - f(\xi))^2 \geqslant$$
$$\geqslant \mathsf{M}(\eta - \mathsf{M}(\eta|\xi))^2, \text{ т.к. } \mathsf{M}((\eta - \mathsf{M}(\eta|\xi))(\mathsf{M}(\eta|\xi) - f(\xi))|\xi) = 0,$$

причем равенство выполняется, только если $f(\xi) = \mathsf{M}(\eta|\xi)$ (п.н.). □

Другие задачи наилучшего приближения.

1) Приближение постоянной. $\mathsf{M}(\xi - C)^2 \sim \min\limits_C \Rightarrow \mathsf{M}\xi^2 - 2C\mathsf{M}\xi + C^2 \sim \min\limits_C$ $\Rightarrow C = \mathsf{M}\xi$ и $\min \mathsf{M}(\xi - C)^2 = \mathsf{D}\xi$.

2) Приближение η по наблюдениям ξ в классе линейных функций.

$$\mathsf{M}(\eta - a\xi - b)^2 = \mathsf{M}(\eta - a\xi)^2 - 2b\mathsf{M}(\eta - a\xi) + b^2 \sim \min_{a,b}. \tag{4}$$

Дифференцируя (4) по (b), находим $b = \mathsf{M}\eta - a\mathsf{M}\xi$, после чего $(*)$ принимает вид $\mathsf{M}(\eta - \mathsf{M}\eta - a(\xi - \mathsf{M}\xi))^2 = \mathsf{D}\eta - 2a\mathrm{cov}\xi\eta + a^2\mathsf{D}\xi$, дифференцируя по (a), получаем $-2\mathrm{cov}\xi\eta + 2a\mathsf{D}\xi = 0 \Rightarrow a = \frac{\mathrm{cov}\xi\eta}{\mathsf{D}\xi}$ и $a\xi + b = a(\xi - \mathsf{M}\xi) + \mathsf{M}\eta = \frac{\mathrm{cov}\xi\eta}{\mathrm{cov}\xi\xi}(\xi - \mathsf{M}\xi) + \mathsf{M}\eta$.

Среднеквадратичная ошибка при этом равна $\mathsf{M}(\eta - a\xi - b)^2 =$

$$= \mathsf{M}(\eta - \mathsf{M}\eta - a(\xi - \mathsf{M}\xi)^2 = \mathsf{D}\eta - 2a\mathrm{cov}\xi\eta + a^2\mathsf{D}\xi = \mathsf{D}\eta - \frac{(\mathrm{cov}\xi\eta)^2}{\mathsf{D}\xi}. \quad \square$$

Задача наилучшего линейного приближения в векторном случае. Пусть ξ — измеренный вектор, $\mathsf{M}\xi = 0$, $\mathsf{M}\eta = 0$. Нужно найти матрицу A, такую, что $\mathsf{M}\|\eta - A\xi\|^2 \sim \min\limits_A$.

$$Y(A) = \mathsf{M}\|\eta - A\xi\|^2 = \mathsf{M}\sum_i(\eta - A\xi)_i^2 =$$
$$= \mathsf{M}\,\mathrm{tr}(\eta - A\xi)(\eta - A\xi)^* = \mathrm{tr}\,\Sigma_{\eta\eta} - \mathrm{tr}\,A\Sigma_{\xi\eta} - \mathrm{tr}\,\Sigma_{\eta\xi}A^* + \mathrm{tr}\,A\Sigma_{\xi\xi}A^*.$$

Здесь $\Sigma_{\eta\eta} = \mathsf{M}\eta\eta^*$, $\Sigma_{\eta\xi} = \mathsf{M}\eta\xi^*$, $\Sigma_{\xi\xi} = \mathsf{M}\xi\xi^*$, а знак $(^*)$ означает сопряжение (транспонирование).

Варьируя Y по A, получим $\delta Y = Y(A + \delta A) - Y(A) =$

$$= -\,\mathrm{tr}\,\delta A\Sigma_{\xi\eta} - \mathrm{tr}\,\Sigma_{\eta\xi}\delta A^* + \mathrm{tr}\,\delta A\Sigma_{\xi\xi}A^* + \mathrm{tr}\,A\Sigma_{\xi\xi}\delta A^* = 0.$$

Отсюда, учитывая, что $\mathrm{tr}\,Q^* = \mathrm{tr}\,Q$, имеем $\mathrm{tr}(\Sigma_{\eta\xi} - A\Sigma_{\xi\xi})\delta A = 0$, откуда в силу произвольности δA следует $\Sigma_{\eta\xi} = A\Sigma_{\xi\xi}$, таким образом, $A = \Sigma_{\eta\xi}\Sigma_{\xi\xi}^{-1}$ и $\hat{\eta} = \Sigma_{\eta\xi}\Sigma_{\xi\xi}^{-1}\xi$. При этом с.к. ошибка (погрешность) линейного приближения равна

$$\mathsf{M}\|\eta - \hat{\eta}\|^2 = \mathrm{tr}(\Sigma_{\eta\eta} - \Sigma_{\eta\xi}\Sigma_{\xi\xi}^{-1}\Sigma_{\xi\eta})$$

и $\mathsf{M}\|\eta - \mathsf{M}\eta\|^2 = \mathrm{tr}\,\Sigma_{\eta\eta}$ — априорная погрешность. $\square$

ц. 7 **Последовательности случайных величин. Сходимость. Виды сходимости.**

Сходимость последовательностей случайных величин требует уточнения. Начнем с наиболее часто используемых типов.

Определение сходимости по вероятности. Говорят, что последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ по вероятности, если $\forall\varepsilon > 0$: $\lim\limits_{n\to\infty} P\{|\xi - \xi_n| > \varepsilon\} = 0$, при этом иногда пишут $\xi_n \xrightarrow[n\to\infty]{P} \xi$.

Закон больших чисел в форме Чебышёва.

Пусть ξ_i попарно независимы, причем $\mathsf{D}\xi_n < C$ (ограничены в совокупности). Тогда

$$\eta_n = \frac{1}{n}\sum_i^n(\xi - \mathsf{M}\xi_i) \xrightarrow[n\to\infty]{P} 0.$$

Доказательство. $\mathsf{D}\eta_n = \frac{\sum \mathsf{D}\xi_i}{n^2} < \frac{C}{n}$.

Из неравенства Чебышёва следует

$$P\{|\eta_n| > \varepsilon\} \leqslant \frac{\mathsf{D}\eta_n}{\varepsilon^2} < \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0. \quad \square$$

Следствия.

1. Пусть $\mathsf{M}\xi_k = \mu$, $\mathsf{D}\xi_n < C$. Тогда $\frac{1}{n}\sum\limits_{i=1}^{\infty}\xi_i \xrightarrow[n\to\infty]{P} \mu$.

2. Пусть ξ_i — число успехов при одном испытании в схеме Бернулли. Тогда $\mathsf{M}\xi_i = p$, $\mathsf{D}\xi_i = pq$ и

$$\frac{1}{n}\sum_{i=1}^{\infty} \xi_i \xrightarrow[n\to\infty]{P} p.$$

3. Теорема Маркова. Пусть $\mathsf{M}\xi_n = \mu$. Снимем условие независимости, но потребуем, чтобы $\mathsf{D}\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i \to 0$. Тогда из неравенства Чебышёва получаем непосредственно

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i \xrightarrow[n\to\infty]{P} \frac{1}{n}\sum_{i=1}^{n} \mathsf{M}\xi_i = \mu.$$

Характеристические функции.

Определение. Характеристической функцией называется

$$f_\xi(t) = \mathsf{M}e^{i\xi t},$$

причем этот момент существует всегда, так как $|e^{i\xi t}| = 1$.

Примеры.

Распределение Пуассона:

$$f_\xi(t) = \mathsf{M}e^{i\xi t} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{e^{ikt}\lambda^k}{n!} = e^{\lambda(e^{it}-1)}.$$

Нормальное распределение:

$$f_\xi(t) = \mathsf{M}e^{i\xi t} = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{ixt}e^{-\frac{x^2}{2}}dx = e^{-\frac{t^2}{2}}.$$

Из очевидных свойств отметим следующие: $f(0) = 1$, $|f(t)| \leqslant 1$, четность и вещественность $f(t)$ эквивалентны и в этом случае $f(t) = \mathsf{M}\cos(\xi t)$. Следующие свойства сформулированы в виде теорем.

Теорема 1. $f_{\sigma\xi+\mu}(t) = e^{i\mu t}\mathsf{M}e^{i\xi\sigma t}$.

Теорема 2. Пусть ξ_i, $i = 1, 2, \ldots$ независимы в совокупности. Тогда $f_{\xi_1+\xi_2+\cdots+\xi_n}(t) = f_{\xi_1}(t)f_{\xi_2}(t)\ldots f_{\xi_n}(t)$.

Теорема 3. Пусть существует момент k-го порядка. Тогда существует k-я производная характеристической функции $f_\xi^{(k)}(t)$, она равномерно непрерывна на $-\infty, \infty$ и имеет место равенство

$$f_\xi^{(k)}(0) = i^k\mathsf{M}\xi^k. \tag{5}$$

Первое и третье свойства очевидны, второе доказывается цепочкой равенств (неравенств):

$$|f_\xi^{(k)}(t+h) - f_\xi^{(k)}(t)| \leqslant \int\limits_{-A}^{A} |x|^k|e^{ixh}-1|p(x)dx + 2\int\limits_{|x|>A} |x|^k p(x)dx. \tag{6}$$

Выберем $A > 0$ так, чтобы второе слагаемое было менее $\varepsilon/2$ (следствие сходимости $\mathsf{M}|\xi|^k$). Заметим, что

$$|e^{ixh} - 1| = |\int\limits_0^{xh} e^{i\alpha} d\alpha| \leqslant |xh| \leqslant A|h|.$$

Выбрав теперь $|h| \leqslant \frac{\varepsilon}{2A\mathsf{M}|\xi|^k}$, мы получаем $|f_\xi^{(k)}(t+h) - f_\xi^{(k)}(t)| \leqslant \varepsilon$ независимо от t.

Теорема 4. *Без доказательства.*

Если х.ф. $f_\xi(t)$ абсолютно интегрируема на R_1, то сл. величина ξ непрерывна, а ее плотность вероятности равна

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-ixt} f_\xi(t) dt \tag{7}$$

и равномерно непрерывна на R_1.

(По существу — это теорема об обратном Фурье-преобразовании.)

Теорема 5. Теорема непрерывности для характеристических функций. *Без доказательства*

Пусть последовательность $\{f_{\xi_n}(t)\}$ х.ф. случайных величин сходится при $n \to \infty$ к х.ф. $f_\xi(t)$ случайной величины ξ равномерно по t в каждом конечном интервале $|t| \leqslant T$. Тогда при $n \to \infty$

$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x) \tag{8}$$

в точках непрерывности $F_\xi(x)$, а если последняя непрерывна, то равномерно по $x \in R_1$.

(**Определение**). Если $\lim\limits_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$, то говорят, что последовательность сл. величин ξ_n сходится к случайной величине ξ по распределению, $\xi_n \xrightarrow[n\to\infty]{d} \xi$.

Следствие. Если у двух сл. величин совпадают х.ф., то совпадают и законы распределения (фукции распределения).

Центральная предельная теорема (ЦПТ).

Пусть $\{\xi_n\}$ — последовательность одинаково распределенных, независимых в совокупности сл. величин, $\mathsf{M}\xi_n = \mu$, $\mathsf{D}\xi_n = \sigma^2$. Тогда

$$\eta_n = \frac{\sum\limits_{i=1}^{n} (\xi_i - \mu)}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} \eta \sim \mathcal{N}(0,1), \tag{9}$$

т.е., $P\{\eta_n < x\} \approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp(-\frac{s^2}{2}) ds.$

Доказательство.
Пусть $\varphi(t)$ — х.ф. сл. величины $\xi_i - \mu$. Тогда

$$f_{\eta_n}(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[\varphi(0) + \frac{t}{\sigma\sqrt{n}}\varphi'(0) + \frac{t^2}{2\sigma^2 n}\varphi''(0) + o\left(\frac{1}{n}\right)\right]^n =$$

$$= \left[1 + 0 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \underset{n\to\infty}{\longrightarrow} e^{-\frac{t^2}{2}}$$

— х.ф. нормального распределения. Далее воспользуемся теоремой непрерывности для характеристических функций. □

Следствие. Теорема Муавра-Лапласа. Рассмотрим схему Бернулли, где m — число успехов и $n \to \infty$ при фиксирванных $0 < p < 1$ и $0 < p < 1$. Тогда справедливо

$$P\left\{a \leqslant \frac{m - np}{\sqrt{npq}} < b\right\} \approx \Phi(b) - \Phi(a).$$

Центральные предельные теоремы для неодинаково распределенных случайных величин. *Обзор, без доказательств.*

Пусть сл. величины $\{\xi_k\}$ независимы, существуют моменты 1 и 2 порядка $\mathsf{M}\xi_k = \mu_k$ и $\mathsf{D}\xi_k = \sigma_k^2$

Теорема А.М.Ляпунова. Пусть для некоторого $\delta > 0$ существуют $c_k^{2+\delta} = \mathsf{M}|\xi_k - \mu_k|^{2+\delta}$. Обозначим $C_n^{2+\delta} = \sum\limits_{k=1}^n c_k^{2+\delta}$, $B_n^2 = \sum\limits_{k=1}^n \sigma_k^2$. Если $\frac{C_n^{2+\delta}}{B_n^{2+\delta}} \underset{n\to\infty}{\longrightarrow} 0$ (условие Ляпунова), то $P\{\eta_n < x\} \underset{n\to\infty}{\longrightarrow} \Phi(x)$.

Замечание 1. Условие Ляпунова не является необходимым, однако его практически легче проверять и доказывать теорему, особенно при[13] $\delta = 1$.

Замечание 2. Еще более слабым является условие Линдеберга, которым можно заменить условие Ляпунова: для всякого $\tau > 0$

$$\frac{1}{B_n^2}\sum_{k=1}^n \int\limits_{|x-\mu_k|\geqslant\tau B_n} (x-\mu_k)^2 dF_k(x) \underset{n\to\infty}{\longrightarrow} 0,$$

но и оно не является необходимым, т.е. существуют распределения, которые ему не удовлетворяют, но сходимость к нормальному закону имеет место. Однако, если добавить требование пренебрежимой (асимптотической) малости величин $\zeta_k = \frac{\xi_k - \mathsf{M}\xi_k}{B_n}$,

$$\max_{1\leqslant k\leqslant n} P\{|\zeta_k| > \varepsilon\} \underset{n\to\infty}{\longrightarrow} 0,$$

то условие Линдеберга становится необходимым.

[13] См. доказательство в учебнике Пытьева, Шишмарева, стр. 128.

Л. 8 **Перечислим виды сходимости случайных величин**.

1\. Среднеквадратичная $\lim\limits_{n\to\infty} \mathsf{M}|\xi_n-\xi|^2=0$ или $\xi_n \xrightarrow[n\to\infty]{\text{с.к.}} \xi$ или $\underset{n\to\infty}{\text{l.i.m.}}\,\xi_n=\xi$.

2\. С вероятностью 1, почти наверное, по модулю P $(\mathrm{mod}P)$ [14]: $P\{\omega:\ \lim\limits_{n\to\infty}\{\xi_n(\omega)=\xi(\omega)\}=1$ или $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$.

3\. По вероятности $\forall\varepsilon>0$: $\lim\limits_{n\to\infty} P\{|\xi-\xi_n|>\varepsilon\}=0$ или $\xi_n \xrightarrow[n\to\infty]{P} \xi$.

4\. По распределению («слабая») $F_{\xi_n}(x)\to F_\xi(x)$ или $\xi_n \xrightarrow[n\to\infty]{d} \xi$.

Отношения между видами сходимости.

$$\begin{array}{lcl}
\xi_n \xrightarrow[n\to\infty]{\text{с.к.}} \xi & & \\
 & \searrow & \\
 & & \xi_n \xrightarrow[n\to\infty]{P} \xi \Longrightarrow \xi_n \xrightarrow[n\to\infty]{d} \xi. \\
 & \nearrow & \\
\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi & &
\end{array}$$

Других отношений без дополнительных условий не существует.

Напомним, что речь всюду идет о случайных величинах $\xi_n(\omega)$, зависящих от элементарного исхода $\omega\in\Omega$. При каждом элементарном исходе мы получаем числовую последовательность, названмую выборочной.

Если последовательность $\{\xi_n\}$ сходится в с.к., по вероятности или по распределению, то может так случиться, что соответствующие выборочные последовательности не будут сходиться ни при каком $\omega\in\Omega$.

Однако, если $\xi_n \xrightarrow{P} \xi$, то существует подпоследовательность $\{\xi_{n_k}\}$ последовательности $\{\xi_n\}$, сходящаяся к ξ с вероятностью единица.

Связь между сходимостью по вероятности и с вероятностью единица. Покажем, что из сходимости с вероятностью единица следует сходимость по вероятности. Действительно, пусть $A_k=\{|\xi_k-\xi|>\varepsilon\}$, $\varepsilon>0$. Сходимость с вероятностью единица эквивалентна равенству $P(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k\geqslant n}A_k)=0$, т.е. $\lim\limits_{n\to\infty}P(\bigcup\limits_{k\geqslant n}A_k)=0$. Но $P(\bigcup\limits_{k\geqslant n}A_k)\geqslant P(A_n)$, следовательно, $\lim\limits_{n\to\infty}P(A_n)=0$. $\square$

Однако, если $\xi=\mathrm{const}$ с вероятностью единица, то $\xi_n \xrightarrow{P} \xi \Leftrightarrow \xi_n \xrightarrow{d} \xi$ и, в частности, $\xi_n-\xi \xrightarrow{P} 0 \Leftrightarrow \xi_n-\xi \xrightarrow{d} 0$

Поэтому $\xi_n \xrightarrow{d} \xi$, $\xi_n-\xi \xrightarrow{d} 0$ — разные сходимости.

Если последовательность $\{\xi_n\}$ с.к. сходится к ξ "достаточно быстро", так что $\sum\limits_{k=1}^{\infty}\mathsf{M}|\xi_k-\xi|^2<\infty$, то она сходится к ξ и почти наверное: $\xi_n \xrightarrow{\text{с.к.}} \xi$, $\sum\limits_{k=1}^{\infty}\mathsf{M}|\xi_k-\xi|^2<\infty \Rightarrow \xi_n \xrightarrow{\text{п.н.}} \xi$.

Кроме того, $\xi_n \xrightarrow{\text{с.к.}} \xi \Leftrightarrow \xi_n \xrightarrow{P} \xi$ и $\mathsf{M}\xi_n^2\to\mathsf{M}\xi^2$ (Лоэв).

---

[14] Заметим, что множество A тех $\omega\in\Omega$, для которых существует $\lim\limits_{n\to\infty}\xi_n$ действительно является событием, так как его можно представить в виде $A=\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=1}^{\infty}\bigcap\limits_{m_1,m_2>n}\{\omega:|\xi_{m_1}-\xi_{m_2}|<\frac{1}{k}\}$.

A — множество ω, удовлетворяющее условиям: для $\forall k=1,2,\ldots$ $\exists n\geqslant 1$, такое, что $\forall m_1,m_2>n:\ |\xi_{m_1}(\omega)-\xi_{m_2}(\omega)|<\frac{1}{k}$.

В каждой точке $\omega\in A$ последовательность $\{\xi_n\}$ фундаментальна и, следовательно, сходится. Измеримость A следует из его представления, а в случае сходимости $\{\xi_n\}$ почти наверное $P(A)=1$.

Для доказательства того, что из сходимости по вероятности следует сходимость по распределению, уточним еще понятие "слабой"сходимости: последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ, если для любой ограниченной непрерывной функции $f(x)$ $\lim\limits_{n\to\infty} \mathsf{M}f(\xi_n) = \mathsf{M}f(\xi)$.

Выберем в качестве $f(x)$ непрерывную функцию $g_\varepsilon(t)$, которая равна 1 при $t \leqslant x$, 0 при $t \geqslant x+\varepsilon$ и линейна при $x \in (x, x+\varepsilon)$.

Из двух неравенств

$F_n(x) = \int\limits_{-\infty}^{x} g_\varepsilon dF_n(x) \leqslant \int\limits_{-\infty}^{\infty} g_\varepsilon dF_n(x)$ и $\int\limits_{-\infty}^{\infty} g_\varepsilon dF(x) \leqslant F(x+\varepsilon)$ следует, что $\lim\limits_{n\to\infty} \sup F_n(x) \leqslant F(x+\varepsilon)$.

Аналогично можно получить $\lim\limits_{n\to\infty} \inf F_n(x) \geqslant F(x-\varepsilon)$.

Объединяя последние два неравенства, получим

$$F(x-\varepsilon) \leqslant \lim_{n\to\infty} \inf F_n(x) \leqslant \lim_{n\to\infty} \sup F_n(x) \leqslant F(x+\varepsilon),$$

откуда в силу произвольности ε следует $F_n(x) \to F(x)$, $n \to \infty$ в точках непрерывности x.

Наконец, из $\xi_n \xrightarrow[n\to\infty]{P} \xi$ следует, что

$$\lim_{n\to\infty} \mathsf{M}f(\xi_n) = \mathsf{M}f(\xi). \tag{10}$$

Действительно, пусть $|f(x)| \leqslant c$, $\varepsilon > 0$ и N таково, что $P(|\xi| > N) \leqslant \varepsilon/4c$. Выберем δ таким, чтобы для всех $|x| \leqslant N$ и $|x-y| \leqslant \delta$ было выполнено неравенство $|f(x) - f(y)| \leqslant \varepsilon/4c$. Тогда

$$\begin{gathered}
\mathsf{M}|f(\xi_n) - f(\xi)| = \mathsf{M}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta, |\xi| \leqslant N) + \\
+\mathsf{M}|f(\xi_n) - f(\xi)| = \mathsf{M}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta, |\xi| > N) \\
+\mathsf{M}|f(\xi_n) - f(\xi)| = \mathsf{M}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| > \delta) \leqslant \\
\leqslant \varepsilon/2 + \varepsilon/2 + 2cP(|\xi_n - \xi| > \delta).
\end{gathered}$$

Для достаточно больших n последнее слагаемое может быть сделано менее ε, что в силу произвольности ε доказывает (10). □

Закон больших чисел может пониматься и в смысле сходимости почти наверное ("Усиленный ЗБЧ"). Различные его варианты изложены в изящных теоремах, однако он мало интересен с прикладной точки зрения.

В ЦПТ мы показали сходимость η_n лишь по распределению. К сожалению, никакая другая сходимость не имеет места: последовательность η_n не сходится ни к какой сл. величине по вероятности, а следовательно и с вероятностью единица и в среднем квадратичном. Можно показать, что сходимость $\eta_n \xrightarrow{P} \eta$, а следовательно, и сходимости $\eta_n \overset{\text{с.к.}}{\to} \eta$ и $\eta_n \overset{\text{п.н.}}{\to} \eta$ не имеют места.

Цепи Маркова.

Рассмотрим последовательность испытаний, где элементарным событием является цепочка $\omega_{i_1}\omega_{i_2}...\omega_{i_n}$. Согласно теореме умножения вероятностей, $P(\omega_{i_1}\omega_{i_2}...\omega_{i_n}) = P(\omega_{i_n} \mid \omega_{i_1}...\omega_{i_{n-1}})...p(\omega_{i_1})$. Пусть последовательные испытания "минимально"зависимы: $P(\omega_{i_s} \mid \omega_{i_1}...\omega_{i_{s-1}}) = P(\omega_{i_s} \mid \omega_{i_{s-1}})$ и не зависит от s, то есть $P(\omega_j \mid \omega_i) = p_{ij}$. Пусть известно начальное распределение вероятностей $P(\omega_i) = a_i$. Тогда

$$P(\omega_{i_1}\omega_{i_2}...\omega_{i_n}) = a_{i_1}p_{i_1i_2}...p_{i_{n-1}i_n}. \tag{11}$$

Определение. Пусть задано вероятностное пространство с конечным набором элементарных событий $\{\omega_i\}$, $i = 1, ..., N$. Конечной однородной цепью Маркова называется вероятностное пространство, в котором $\Omega_n = \overset{n}{\times}\ \Omega$, $\mathcal{F}_n = \overset{n}{\times}\ \mathcal{F}$, а вероятность $P_n(\omega_{i_1}\omega_{i_2}...\omega_{i_n})$ определяется формулой (11).

Для вычисления (11), в частности, должны быть заданы переходные вероятности p_{ij} или переходная матрица

$$\pi = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$$

Это так наз. стохастическая матрица ($\sum\limits_j p_ij = 1$). Отсюда следует корректность задания (11) — сумма по всем индексам равна 1. (Проверить!) Марковскую цепь можно интерпретировать как модель системы с N состояниями и вероятностями перехода p_{ij} из i-го в j-е состояние.

Примеры: случайное дискретное блуждание по прямой.

Частный случай — независимые испытания ($p_{ij} = p_j$).

Замечание. Для практики удобен следующий критерий «Марковости».

Обозначим буквами П, Н и Б — события: прошлое, настоящее и будущее. Тогда $P(\text{ПБ}|\text{Н}) = P(\text{П}|\text{Н})P(\text{Б}|\text{Н})$ — при фиксированном настоящем прошлое и будущее независимы. Доказательство следует из равенств

$$P(\text{ПНБ}) = P(\text{ПБ}|\text{Н})P(\text{Н}) = P(\text{Б}|\text{ПН})P(\text{П}|\text{Н})P(\text{Н}) = \\ = P(\text{Б}|\text{Н})P(\text{П}|\text{Н})P(\text{Н}).$$

(Докажите самостоятельно необходимость и достаточность!)

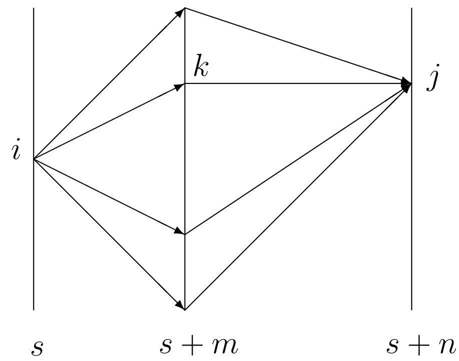
Переход за n шагов. Обозначим через $p_{ij}^{(n)}$ вероятность перехода из i-го в j-е состояние за n шагов. Легко предположить (однородность!), что она не зависит от абсолютного номера шага s.

При переходе из состояния i на s-м шаге в состояние j на $s+n$-м шаге система на $s+m$-м шаге может побывать в любом состоянии k, $k = 1, \dots, N$.

Отсюда следует, что $\pi^{(n)} = \pi^n$ и $\pi^n = \pi^m\pi^{n-m}$ (в данном случае это просто степень матрицы).

По формуле полной вероятности имеем

$$p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik}^{(m)} p_{kj}^{(n-m)} \text{ или } \pi^{(n)} = \pi^{(m)}\pi^{(n-m)}. \tag{12}$$



Отметим, что π^n — тоже стохастическая матрица (Доказать!).

Эргодичность. Рассмотрим *абсолютную* вероятность $p_j^n = \sum\limits_{k=1}^{N} a_i p_{ij}^{(n)}$ (нахождения системы в состоянии j на n-м шаге от начала) и предположим, что[15] $\lim\limits_{n\to\infty} p_{ij}^n = p_j$.

Тогда $\lim\limits_{n\to\infty} p_j^n = \lim\limits_{n\to\infty} \sum\limits_{k=1}^{N} a_k p_{kj}^n = \sum\limits_{k=1}^{N} a_k p_j = p_j$ — *финальное* распределение вероятностей. Из равенства $\sum\limits_{k=1}^{N} p_i^{n-1} p_{ij} = p_j^n$ ((12), $m = n-1$) при $n \to \infty$ следует

$$\sum_{k=1}^{N} p_i p_{ij} = p_j. \tag{13}$$

Это значит, что финальное распределение стационарно (инвариантно) т.е. является единственным решением системы алгебраических уравнений[16] с дополнительными условиями $p_j \geqslant 0$ и $\sum\limits_{i=1}^{N} p_i = 1$.

Если при этом $p_j > 0$, $j = 1, \ldots, N$, то говорят, что Марковская цепь (система) эргодична. В этом случае безусловные (абсолютные) вероятности $p_j^{(n)}$ стремятся к p_j независимо от начального распределения, а система (13) при указанных условиях имеет единственное решение.

Полезна следующая теорема для конечных марковских цепей:

Теорема Маркова. Пусть существуют целые ν и j_0, такие, что $\min\limits_i p_{ij_0}^{\nu} = \delta > 0$. Тогда существуют числа p_j, $j = 1, \ldots, N$, такие, что $\lim\limits_{n\to\infty} p_{ij}^n = p_j$, $j = 1, \ldots, N$ независимо от i.

Доказательство. Обозначим $M_j^r = \max\limits_i p_{ij^r}$, $m_j^r = \min\limits_i p_{ij^r}$, $m_j^r \leqslant M_j^r$.

[15] Далее опустим скобки у верхнего индекса.

[16] Левый собственный вектор матрицы π, отвечающий собственному значению, равному 1.

Далее, $m_j^{r+1} = \min\limits_i \sum\limits_k p_{ik}p_{ki}^r \geqslant \min\limits_i \sum\limits_k p_{ik} \min\limits_k p_{ki}^r = m_j^r$. Аналогично, $M_j^{r+1} \leqslant M_j^r$.

Таким образом, $m_j^1 \leqslant m_j^2 \leqslant \ldots \leqslant m_j^r \leqslant \ldots \leqslant M_j^r \leqslant \ldots \leqslant M_j^2 \leqslant M_j^1$.

Из монотонности и ограниченности следует, что существуют пределы $\lim\limits_{n\to\infty} M_j^n = M_j$ и $\lim\limits_{n\to\infty} m_j^n = m_j$, причем $m_j \leqslant M_j$.

Теперь фиксируем две строки матрицы π^ν с номерами α и β (т.е. рассмотрим $p_{\alpha k}^\nu$ и $p_{\beta k}^\nu$), $k = 1, \ldots, N$.

Обозначим $\sum^+ = \sum\limits_{k:\, p_{\alpha k}^\nu \geqslant p_{\beta k}^\nu}$ и $\sum^- = \sum\limits_{k:\, p_{\alpha k}^\nu < p_{\beta k}^\nu}$.

Тогда

$\sum^+ p_{\alpha k}^\nu + \sum^+ p_{\alpha k}^\nu = 1, \quad \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) + \sum^+ (p_{\alpha k} - p_{\beta k}^\nu) = 0,$

$\sum^+ p_{\alpha k} - \sum^+ p_{\beta k} = 1 - \sum^- p_{\alpha k} - \sum^+ p_{\beta k} \leqslant 1 - \delta.$

Далее,

$$
\begin{aligned}
M_j^{\nu+n} - m_j^{\nu+n} &= \max_\alpha \sum_k p_{\alpha k}^\nu p_{kj}^n - \min_\beta \sum_k p_{\beta k}^\nu p_{kj}^n = \\
&= \max_{\alpha,\beta} \sum_k (p_{\alpha k}^\nu p_{kj}^n - p_{\beta k}^\nu p_{kj}^n) = \\
&= \max_{\alpha,\beta} \left\{ \sum\nolimits^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) p_{kj}^n + \sum\nolimits^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) p_{kj}^n \right\} \leqslant \\
&\leqslant \max_{\alpha,\beta} \left\{ \sum\nolimits^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) M_j^n + \sum\nolimits^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) m_j^n \right\} = \\
&= \max_{\alpha,\beta} \left\{ \sum\nolimits^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) M_j^n - \sum\nolimits^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) m_j^n \right\} = \\
&= \max_{\alpha,\beta} \sum\nolimits^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu)(M_j^n - n_j^n) \leqslant \\
&\leqslant (1-\delta)(M_j^n - m_j^n).
\end{aligned}
$$

Переходя к пределу при $n \to \infty$, получим

$$M_j - m_j \leqslant (1-\delta)(M_j - m_j),$$

что возможно лишь при $M_j = m_j$. □

Следствие. Если условия теоремы Маркова усилить, потребовав выполнения неравенства $\min\limits_{ij} p_{ij}^\nu = \delta > 0$ для всех j, а не только для j_0, то из $p_j \geqslant m_j^\nu \geqslant \delta > 0$, $j = 1, \ldots, N$ следует эргодичность. □

Теорема (без доказательства). Пусть A — подмножество состояний эргодической системы, T_A — время нахождения системы в A, T — общее время функционирования системы. Тогда

$$\lim_{T\to\infty} \frac{T_A}{T} = \sum_{\omega_i \in A} p_i.$$

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Распределение ортогональных проекций нормального вектора. Л. 9

Рассмотрим n-мерное евклидово пространство $\mathcal{R}_n$, $x = (x_1, \ldots, x_n) \in \mathcal{R}_n$, $(x, y) = \sum\limits_{k=1}^{n} x_k y_k$. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор в этом пространстве, причем $\xi \sim \mathcal{N}(0, \sigma^2 I)$.

Напомним, что в общем случае, когда $\xi \sim \mathcal{N}(0, \Sigma)$,

$$p_\xi(x) = \frac{(\det \Sigma^{-1})^{1/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(x, \Sigma^{-1}x)\right\}.$$

Если $\Sigma = \sigma^2 I$, где I — единичный оператор, то координаты ξ_i независимы и

$$p_{\xi_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}x_i^2\right\}.$$

Ортогональный проектор.

Пусть L — линейное подпространство $\mathcal{R}_n$ и $L^\perp$ — ортогональное дополнение L в $\mathcal{R}_n$, т. е. множество векторов $x \in \mathcal{R}_n$, ортогональных всем векторам из L:

$$L^\perp = \{x \in \mathcal{R}_n,\ (x, y) = 0,\ y \in L\}.$$

Очевидно, $\mathcal{L}^\perp$ — также линейное пространство $\mathcal{R}_n$. Как известно, всякий вектор $x \in \mathcal{R}_n$ может быть представлен в виде суммы

$$x = x_1 + x_2, \quad x_1 \in \mathcal{L}, x_2 \in \mathcal{L}^\perp. \tag{14}$$

Разложение (14) единственно. Действительно, равенство $x = x_1 + x_2 = x_1' + x_2'$, $x_1' \in L$, $x_2' \in L^\perp$, совместно с (14) влечет $(x_1 - x_1')^2 + (x_2 - x_2')^2 = 0$. Слагаемые в последнем равенстве ортогональны, так как $x_1 - x_1' \in L$, $x_2 - x_2' \in L^\perp$, поэтому

$(x_1 - x_1')^2 + (x_2 - x_2')^2 = 0$, т. е. $x_1' = x_1$, $x_2' = x_2$. Следовательно каждому $x \in \mathcal{R}_n$ разложением (14) ставится в соответствие единственный вектор $x_1 \in L$:

$$x_1 = \Pi x. \tag{15}$$

Π называется оператором ортогонального проецирования на L, или ортогональным проектором на L. Отметим следующие свойства оператора Π.

1) Π — линейный оператор. Действительно, пусть

$$x = x_1 + x_2,\ y = y_1 + y_2,\ x_1, y_1 \in L,\ x_2, y_2 \in L^\perp. \tag{16}$$

Тогда

$$\alpha x + \beta y = (\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2),\ \alpha x_1 + \beta y_1 \in L,\ \alpha x_2 + \beta y_2 \in L^\perp.$$

Следовательно, согласно определению (15) $\Pi \alpha x + \beta y = \alpha x_1 + \beta y_1 = \alpha \Pi x + \beta \Pi y$.

2) Π — самосопряженный оператор, т.е. для любых $x, y \in \mathcal{R}_n$ $(\Pi x, y) = (x, \Pi y)$. Действительно, воспользовавшись разложением (16), найдем

$$(\Pi x, y) = (x_1, y) = (x_1, y_1) = (x, y_1) = (x, \Pi y).$$

3) Оператор Π удовлетворяет уравнению $\Pi^2 = \Pi$ («идемпотентность»). Действительно, для всякого $x \in \mathcal{R}_n$: $\Pi x = x_1 = \Pi x_1 = \Pi(\Pi x)$, поскольку для $\in L$ разложение (14) имеет вид $x_1 = x_1 = \Pi x_1$.

На самом деле свойства 1) — 3) не только необходимы, но и достаточны для того, чтобы оператор Π был ортогональным проектором. Для доказательства предположим, согласно свойству 1, что Π — линейный оператор. Обозначим через L множество решений уравнения $\Pi x = x$. через N — множество решений уравнения $\Pi x = 0$. Легко убедиться, что L и N — линейные подпространства $\mathcal{R}_n$, причем ортогональные, если Π удовлетворяет условию 2). В самом деле, если $x \in L$, $y \in N$. то $(x, y) = (\Pi,) = (, \Pi y) = (x, 0) = 0$. Для всякого вектора $x \in \mathcal{R}_n$ можно записать тождество

$$x = \Pi x + (I - \Pi)x. \tag{17}$$

Если Π удовлетворяет условию 3), то $\Pi(\Pi x) = \Pi x$, т. е. $\Pi x \in L$ и $\Pi(I - \Pi)x = (\Pi - \Pi^2)x = 0$, т. е. $(I - \Pi)x \in N$.

Следовательно, Π — оператор ортогонального проектирования на $L = \{x \in \mathcal{R}_n,\ \Pi x = x\}$. Из разложения (17) следует также, что оператор $I - \Pi$ ортогонально проецирует на $N = \{x \in \mathcal{R}_n\ (I - \Pi)x = x\} = L^\perp$.

Отметим следующее важное свойство ортогонального проектора. Пусть Π — оператор ортогонального проецирования на линейное подпространство L и

$$\rho(x, L) = \inf\{|x - y|| \mid y \in L\} \tag{18}$$

— расстояние от x до L. Тогда

$$\rho(x, L) = ||x - \Pi x||. \tag{19}$$

Действительно, пусть $y \in L$. Тогда

$$\Pi x - y \in L,\ x - \Pi x = (I - \Pi)x \in L^\perp$$

и, следовательно,

$$||x - y||^2 = ||x - \Pi x + \Pi x - y||^2 = ||x - \Pi x||^2 + ||\Pi x - y||^2 \geqslant ||x - \Pi x||^2,$$

причем равенство здесь выполняется лишь в случае $\Pi x = y$.

<u>Ортогональное преобразование.</u>

Обозначим через U оператор ортогонального преобразования (оператор перехода от одного ортонормированного базиса к другому), $\tilde{e}_i = Ue_i$. Он обладает свойствами: $U^* = U^{-1}$ и $||Ux|| = ||x||$ $(||U^*x|| = ||x||)$, $\det \hat{U} = \pm 1$. (Здесь $\hat{U}$ — матрица оператора U).

Теорема. Распределение вектора $\eta = U\xi$ совпадает с распределением вектора ξ (т.е. $\eta_i \sim \mathcal{N}(0, \sigma^2)$ и независимы).

Доказательство.

$$p_\eta(y) = p_\xi(U^{-1}y)|\det \hat{U}^{-1}| = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}||U^{-1}y||^2\right\} =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{||y||^2}{2\sigma^2}\right\}. \quad \square$$

Определение 1. Распределением χ^2_m или Пирсона (E.S.Person) с m степенями свободы называется распределение сл. величины, равной сумме квадратов $\sum\limits_{k=1}^{m} \xi_k^2$ независимых сл. величин $\xi_k \sim \mathcal{N}(0,1)$.

Свойства[17]. Плотность:

$$p_{\chi^2_m}(x) = \frac{1}{2^{m/2}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}.$$

Моменты: $\mathsf{M}\chi^2_m = m$, $\mathsf{D}\chi^2_m = 2m$.

Следствие 1. $||\Pi_m\xi||^2 = \sigma^2\chi^2_m$.

Доказательство. Пусть $\Pi_m x \in L_m \subset \mathcal{R}_n$. Рассмотрим новый базис $\{\tilde{e}_i\}$, $i = 1, \ldots, n$, удовлетворяющий условиям $\tilde{e}_i \in L_m$, $i = 1, \ldots, m$, $\tilde{e}_i \in L_m^\perp$, $i = m+1, \ldots, n$ и пусть U — ортогональный оператор перехода от старого базиса $\{e_i\}$ к новому $\{\tilde{e}_i\}$, $\tilde{e}_i = Ue_i$, $i = 1, \ldots, n$. Тогда $||\Pi_m\xi||^2 = \sum\limits_{i=1}^{m} (\xi, \tilde{e}_i)^2 = \sigma^2\chi^2_m$, т.к. по теореме $(\xi, \tilde{e}_i) \sim \mathcal{N}(0, \sigma^2)$. $\square$

Определение 2. Пусть χ^2_m и χ^2_k и независимы. Тогда сл. величина $F_{m,k} = \frac{\frac{1}{m}\chi^2_m}{\frac{1}{k}\chi^2_k}$ контролируется распределением Фишера (R.A.Fischer).

Свойства. Плотность:

$$p_{F_{m,k}}(x) = \frac{k^{\frac{k}{2}-1} m^{\frac{k}{2}-1} x^{\frac{m}{2}-1} (kx+m)^{-\frac{k+m}{2}}}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})} \Gamma(\frac{k+m}{2}), \quad x > 0.$$

Моменты: $\mathsf{M}F_{k,m} = \frac{m}{m-2}$, $m > 2$ $\mathsf{D}F_{k,m} = \frac{2m^2}{(m-2)^2(m-4)}(1 + \frac{m-2}{k})$, $m > 4$.

Следствие 2. Пусть $\mathcal{R}_n = L_k \oplus L_m \oplus L_s$, $k+m+s = n$, Π_k — ортогональный проектор на L_k, Π_m — ортогональный проектор на L_m. Тогда $F_{k,m} = \frac{\frac{1}{k}||\Pi_k\xi||^2}{\frac{1}{m}||\Pi_m\xi||^2}$.

Доказательство. Пусть новый базис $\{\tilde{e}_i\}$, $i = 1, \ldots, n$ таков, что $\tilde{e}_i \in L_k$, если $i = 1, \ldots, k$, $\tilde{e}_i \in L_m$, если $i = k+1, \ldots, k+m$, $\tilde{e}_i \in L_s$, если $i = k+m+1, \ldots, n$. Тогда $||\Pi_k\xi||^2 = \sigma^2\chi^2_k$, $||\Pi_m\xi||^2 = \sigma^2\chi^2_m$ и независимы. $\square$

Определение 3. Пусть $\xi \sim \mathcal{N}(0,1)$ и χ^2_m независимы. Тогда сл. величина $t_m = \frac{\xi}{\sqrt{\frac{\chi^2_m}{m}}}$ контролируется распределением Стьюдента (В.С.Госсета) с m степенями свободы.

Свойства. Плотность:

$$p_{t_m}(x) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \frac{1}{\sqrt{\pi m}} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \ x > 0.$$

Моменты: $\mathsf{M}t_m = 0$, $\mathsf{D}t_m = \frac{m}{m-2}$.

[17] Вывод соответствующих формул см. в учебнике Пытьева, Шишмарева, стр. 97 − 102.

Следствие 3. Обозначим $e = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$, L_1 — одномерное подпространство, определенное вектором e, Π_1 — ортогональный проектор на L_1, $\Pi_1\xi = (e,\xi)e = (\frac{1}{\sqrt{n}}\sum\limits_{i=1}^n \xi_i)e = (\frac{1}{\sqrt{n}}\sum\limits_{i=1}^n \xi_i, \dots, \frac{1}{\sqrt{n}}\sum\limits_{i=1}^n \xi_i)$.

Пусть $\xi \sim \mathcal{N}(0, \sigma^2 I)$, тогда

$$\frac{(\xi, e)}{[\frac{1}{n-1}\|(I-\Pi_1)\xi\|^2]^{1/2}} = t_{n-1} \tag{20}$$

— распределение Стьюдента с $n-1$ степенью свободы.

Доказательство. Перейдем от базиса $\{e_i\}$ к базису $\{e, \tilde{e}_1, \dots, \tilde{e}_{n-1}\}$, так что $L = \mathcal{L}(e)$, $L^\perp = \mathcal{L}(\tilde{e}_1, \dots, \tilde{e}_{n-1})$. Тогда координаты в этом новом базисе независимы и распределены как $\mathcal{N}(0,\sigma^2)$, а $\|(I-\Pi_1)\xi\|^2 = \sigma^2\chi^2_{n-1}$ и по определению 3 получаем (20).

Перепишем (20) еще раз

$$\frac{(\xi, e)}{[\frac{1}{n-1}\|(I-\Pi_1)\xi\|^2]^{1/2}} = \frac{\frac{1}{\sqrt{n}}\sum\limits_{i=1}^n \xi_i}{[\frac{1}{n-1}\sum\limits_{k=1}^n(\xi_k - \frac{1}{n}\sum\limits_{i=1}^n \xi_i)^2]^{1/2}} =$$

$$= \frac{\frac{1}{n}\sum\limits_{i=1}^n \xi_i}{[\frac{1}{n(n-1)}\sum\limits_{k=1}^n(\xi_k - \frac{1}{n}\sum\limits_{i=1}^n \xi_i)^2]^{1/2}}. \quad \square$$

Если $\xi \sim \mathcal{N}(\bar{\mu}, \sigma^2 I)$, где $\bar{\mu} = (\mu, \dots, \mu)$, то последнюю формулу можно заменить на

$$t_{n-1} = \frac{\frac{1}{n}\sum\limits_{i=1}^n (\xi_i - \mu)}{[\frac{1}{n(n-1)}\sum\limits_{k=1}^n(\xi_k - \frac{1}{n}\sum\limits_{i=1}^n \xi_i)^2]^{1/2}} = \frac{\hat{\mu} - \mu}{\sqrt{\frac{1}{n}\hat{\sigma}^2}}, \tag{21}$$

где $\hat{\mu} = \frac{1}{n}\sum\limits_{i=1}^n \xi_i$, $\quad \hat{\sigma}^2 = \frac{1}{n-1}\sum\limits_{k=1}^n(\xi_k - \frac{1}{n}\sum\limits_{i=1}^n \xi_i)^2 = \frac{1}{n-1}\|(I-\Pi_1)\xi\|^2$.

Интервальные оценки нормального распределения.

Пусть $\{\xi_i\}$, $i = 1.2.\dots$ — последовательность независимых нормально распределенных $\mathcal{N}(\mu, \sigma^2)$ сл. величин (независимых измерений). Требуется оценить значения неизвестных параметров μ и σ^2. Рассмотрим четыре случая.

1. Оценивание μ при известном σ^2.

Очевидно, $\frac{\sum_{i=1}^n(\xi_i - \mu)}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0,1)$.

Тогда

$$P\left(\left|\frac{\sum\limits_{i=1}^n(\xi_i - \mu)}{\sqrt{n\sigma^2}}\right| < \varepsilon\right) = \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon),$$

или, преобразуя неравенство, получаем

$$P\left(\hat{\mu}-\varepsilon\sqrt{\frac{\sigma^2}{n}}<\mu<\hat{\mu}+\varepsilon\sqrt{\frac{\sigma^2}{n}}\right)=\alpha(\varepsilon)=1-2\Phi(-\varepsilon),$$

где $\hat{\mu}=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ — середина интервала, шириной $2\varepsilon\sqrt{\frac{\sigma^2}{n}}$, которому с вероятностью $1-\alpha(\varepsilon)=1-2\Phi(-\varepsilon)$ принадлежит неизвестный параметр μ.

2. Оценивание σ^2 при известном μ. Из определения 1 следует, что $\frac{\sum\limits_{i=1}^{n}(\xi_i-\mu)^2}{\sigma^2}\sim\chi_n^2$, поэтому

$$P\left(\varepsilon_1<\frac{\sum\limits_{i=1}^{n}(\xi_i-\mu)^2}{\sigma^2}<\varepsilon_2\right)=1-\alpha(\varepsilon_1,\varepsilon_2),$$

или

$$P\left(\frac{\sum\limits_{i=1}^{n}(\xi_i-\mu)^2}{\varepsilon_2}<\sigma^2<\frac{\sum\limits_{i=1}^{n}(\xi_i-\mu)^2}{\varepsilon_1}\right)=1-\alpha(\varepsilon_1,\varepsilon_2),$$

причем обычно ε_1 и ε_2 выбирают так, чтобы $P(\chi_n^2<\varepsilon_1)=P(\chi_n^2>\varepsilon_2)$. Интервал, которому удовлетворяет σ^2 с вероятностью $1-\alpha$, называется интервальной оценкой σ^2.

3. Оценивание μ при неизвестном σ^2.

Воспользуемся формулой (21)

$$P(|t_{n-1}|<\varepsilon)=P\left(\left|\frac{\hat{\mu}-\mu}{\sqrt{\frac{1}{n}\hat{\sigma}^2}}\right|<\varepsilon\right)=1-\alpha(\varepsilon)$$

и получаем выражение, аналогичное пункту 1:

$$P\left(\hat{\mu}-\varepsilon\sqrt{\frac{\hat{\sigma}^2}{n}}<\mu<\hat{\mu}+\varepsilon\sqrt{\frac{\hat{\sigma}^2}{n}}\right)=1-\alpha_{n-1}(\varepsilon),$$

но с тем отличием, что вместо σ^2 стоит $\hat{\sigma}^2$ и $1-\alpha_{n-1}(\varepsilon)$ соответствует распределению Стьюдента с $n-1$ степенями свободы.

4. Оценивание σ^2 при неизвестном μ.

Здесь по аналогии с пунктом 2 получаем

$$P\left(\frac{\sum\limits_{i=1}^{n}(\xi_i-\hat{\mu})^2}{\varepsilon_2}<\sigma^2<\frac{\sum\limits_{i=1}^{n}(\xi_i-\hat{\mu})^2}{\varepsilon_1}\right)=1-\alpha_{n-1}(\varepsilon_1,\varepsilon_2),$$

где $\alpha_{n-1}(\varepsilon_1,\varepsilon_2)$ вычисляется по распределению χ_{n-1}^2 с $n-1$ степенями свободы.

Точечные оценки. Лекц. 10

Пусть $\{\xi_i\}$, $i = 1, 2, \ldots, n$ — независимая выборка из распределения $P(x, \theta)$, где θ — неизвестный параметр. Нас интересует оценка $t(\xi)$ величины $\tau(\theta)$ (здесь $\tau(\cdot)$ — известная функция), роль которой играет некоторая *статистика* $t(\xi)$.

Терминология: X — выборочное пространство, n — объем выборки, всякая измеримая функция t от выборки ξ называется статистикой, следовательно по определению любая точечная оценка — статистика.

Желательные свойства оценок:

1. Несмещенность $\mathsf{M}t(\xi) = \tau(\theta)$. (Гарантирует от накопления систематических ошибок).

2. Состоятельность $t_n(\xi) \xrightarrow[n\to\infty]{P} \tau(\theta)$.

3. Минимальность дисперсии (если оценка несмещенная) — качество оценки при фиксированном объеме выборки

Примеры.

Несмещенность $\hat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i$ очевидна. Состоятельность $\hat{\mu}$ — утверждение З.Б.Ч.

Если $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\hat{\sigma}^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(\xi_k - \hat{\mu})^2$ — несмещенная и состоятельная оценка. Действительно, при этом $\hat{\sigma}^2 = \frac{1}{n-1}\sigma^2\chi^2_{n-1}$, $\mathsf{M}\hat{\sigma}^2 = \sigma^2$, а $\mathsf{D}\hat{\sigma}^2 = \frac{\sigma^4}{(n-1)^2}\mathsf{D}\chi^2_{n-1} = \frac{\sigma^4}{(n-1)^2}2(n-1)$ и по неравенству Чебышёва $P\{|\hat{\sigma}^2 - \sigma^2| > \varepsilon\} < \frac{2\sigma^4}{\varepsilon^2(n-1)} \xrightarrow[n\to\infty]{} 0$. Если ξ не является нормальной, то несмещенность оценки сохраняется:

$$\mathsf{M}\sum(\xi_i - \hat{\mu})^2 = \mathsf{M}\sum[(\xi_i - \mu)^2 - 2(\xi_i - \mu)(\hat{\mu} - \mu) + (\hat{\mu} - \mu)^2] =$$

$$= \mathsf{M}[\sum(\xi_i - \mu)^2 - 2(\hat{\mu} - \mu)\sum(\xi_i - \mu) + \sum(\hat{\mu} - \mu)^2] =$$

$$= \mathsf{M}[\sum(\xi_i - \mu)^2 - \frac{2}{n}\sum(\xi_i - \mu)\sum(\xi_i - \mu) + \frac{n}{n^2}\sum(\xi_i - \mu)\sum(\xi_i - \mu)] =$$
$$= n\sigma^2 - 2\sigma^2 + \sigma^2 = (n-1)\sigma^2,$$

а состоятельность — нет (вообще говоря).

Минимальность дисперсии — желательное свойство, однако заметим, что смещение может уменьшить ср. кв. уклонение. Например, задача

$$\mathsf{M}(k\sum(\xi_i - \hat{\mu})^2 - \sigma^2)^2 \sim \min_k$$

имеет решение для $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$ при $k = \frac{1}{n+1}$. (Доказать).

Рассмотрим специально несмещенные оценки минимальной дисперсии (НОМД).

Лемма. Если существует НОМД, то она единственна (с вер. 1).

Доказательство. Пусть $t_1(\xi)$ и $t_2(\xi)$ — НОМД, т.е. $\mathsf{M}t_i(\xi) = \tau(\theta)$, $\mathsf{D}t_i(\xi) = \delta$, $i = 1, 2$.

Рассмотрим $t_3 = \frac{1}{2}(t_1 + t_2)$. Тогда $\mathsf{D}t_3 =$

$$= \frac{1}{4}(\mathsf{D}t_1 + 2\mathrm{cov}t_1t_2 + \mathsf{D}t_2) \leqslant \frac{1}{4}(\mathsf{D}t_1 + 2\sqrt{\mathsf{D}t_1\mathsf{D}t_2} + \mathsf{D}t_2) = \frac{1}{4}(\sqrt{\delta} + \sqrt{\delta})^2 = \delta,$$

но $\mathsf{D}t_3 \geqslant \delta$, т.е. возможно лишь равенство, откуда следует, что $t_1(\xi) - \tau(\theta) = k(\theta)(t_2(\xi) - \tau(\theta))$ с вер. 1 ($k^2 = 1$), и далее, из $\mathrm{cov} t_1 t_2 = \delta$ получаем $k = 1$. □

Определение. Назовем статистику $t(\xi)$, удовлетворяющую условию $\mathsf{M}(t(\xi))^2 < +\infty$, гильбертовой.

Теорема. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — выборка из распределения $P(x, \theta)$. Для того, чтобы гильбертова статистика $t(\xi)$ была НОМД, необходимо и достаточно, чтобы для всякой центрированной гильбертовой статистики $\eta = s(\xi)$ (такой, что $\mathsf{M}\eta = 0$), выполнялось $\mathsf{M}t(\xi)\eta = 0$.

Доказательство. Пусть $t(\xi)$ — гильбертова несмещенная оценка $\tau(\theta)$. Тогда $t(\xi) + \lambda\eta$ — тоже гильбертова несмещенная оценка $\tau(\theta)$ для всех λ. Обозначим

$$\varphi_\lambda = \mathsf{M}(t(\xi) - \tau(\theta) + \lambda\eta)^2.$$

$$\min_\lambda \varphi_\lambda = \mathsf{M}(t - \tau)^2 - \frac{[\mathsf{M}t\eta]^2}{\mathsf{M}\eta^2} \quad \left(\text{при} \lambda = \lambda_* = -\frac{\mathsf{M}t\eta}{\mathsf{M}\eta^2}\right).$$

(Необходимость). Если последнее слагаемое не равно нулю, то существует несмещенная статистика с дисперсией, меньшей, чем у статистики $t(\xi)$.

(Достаточность). Если последнее слагаемое при любых η, $\mathsf{M}\eta = 0$ равно нулю, то статистика $t(\xi)$ имеет наименьшую дисперсию среди всех гильбертовых несмещенных оценок типа $t(\xi) + \lambda\eta$, а следовательно, всех гильбертовых несмещенных оценок, поскольку любую гильбертову несмещенную оценку можно представить в таком виде. □

Иногда качество оценки можно оценить, зная минимально возможное значение ее дисперсии (неравенство Рао-Крамера).

Определение. Функцией правдоподобия для некоторого распределения $P(x, \theta)$ называется $L(x, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$, где $f(x_i, \theta)$ — либо плотность распределения $p_\xi(x, \theta)$ сл. величины ξ, либо $P_\theta\{\xi = x\}$[18].

Теорема Рао-Крамера. Пусть $L(x, \theta)$ — функция правдоподобия, $\theta \in R_1$ и выполнены условия:

1. $t(\xi)$ — несмещенная оценка $\tau(\theta)$.
2. Функции $L(x, \theta)$ и $\tau(\theta)$ дифференцируемы по θ.
3. Можнество тех x, для которых $L(x, \theta) > 0$ не зависит от θ и

$$\frac{d}{d\theta} \int L(x, \theta) dx = \int \frac{d}{d\theta} L(x, \theta) dx$$

и

$$\frac{d}{d\theta} \int t(x) L(x, \theta) dx = \int t(x) \frac{d}{d\theta} L(x, \theta) dx.$$

Тогда

$$\mathsf{D}t(\xi) \geqslant \frac{|\tau'(\theta)|^2}{\mathsf{M}\left[\frac{\partial \ln(\xi, \theta)}{\partial \theta}\right]^2}, \tag{22}$$

[18] Обычно фукнция правдоподобия рассматривается как функция от θ, а значения $x_1, x_2, \ldots, x_n$ (выборка) — параметры.

причем знак равенства имеет место тогда и только тогда, когда

$$\frac{\partial \ln(\xi,\theta)}{\partial\theta} = a(\theta)[t(\xi) - \tau(\theta)] \tag{23}$$

с вероятностью единица для некоторого $a(\theta)$.

Доказательство. Дифференцируя тождества $\int L(x,\theta)dx = 1$ и $\int t(x)L(x,\theta)dx = \tau(\theta)$, получим

$$\int \frac{d\ln L(x,\theta)}{d\theta}L(x,\theta)dx = 0, \int t(x)\frac{d\ln L(x,\theta)}{d\theta}L(x,\theta)dx = \tau'(\theta)$$

или

$$\int [t(x) - \tau(\theta)]\frac{d\ln L(x,\theta)}{d\theta}L(x,\theta)dx = \tau'(\theta)$$

и отсюда по неравенству Коши-Буняковского получаем (22). □

Если в условиях теоремы Рао-Крамера имеет место равенство, то справедливо (23) с вероятностью единица. В этом случае оценка называется эффективной и ее дисперсия равна $\mathsf{D}t(\xi) = \frac{|\tau'(\theta)|}{|a(\theta)|}$.

К таким оценкам, например, приводят распределения, плотности которых можно представить в виде

$$f(x,\theta) = \exp\{a(\theta)b(x) + c(\theta) + d(x)\},\ x \in R_1,\ \theta \in R_1.$$

(Они называются экспоненциальными семействами.) Тогда

$$L(x,\theta) = \exp\{a(\theta)\sum b(x_i) + nc(\theta) + \sum d(x_i)\}$$

и

$$\frac{\partial \ln(\xi,\theta)}{\partial\theta} = a'(\theta)n\left\{\frac{1}{n}\sum b(x_i) + \frac{c'(\theta)}{a'(\theta)}\right\}.$$

Пусть условия теоремы Рао-Крамера выполнены, тогда $t(\xi) = \frac{1}{n}\sum b(x_i)$ есть эффективная оценка $\tau(\theta) = -\frac{c'(\theta)}{a'(\theta)}$ с дисперсией

$$\left|\frac{\tau'(\theta)}{na'(\theta)}\right|.$$

Экспоненциальному семейству принадлежат многие важные для практики распределения: нормальное, Пуассона, Бернулли (биномиальное), гамма-распределение и другие.

К сожалению, класс эффективных оценок весьма узок: если такая оценка существует для функции $\tau(\theta)$, то она не существует ни для какой функции, отличной от $c_1\tau(\theta) + c_2$.

Оценки максимального правдоподобия. Значение параметра $\theta = \hat{\theta}$, при котором функция правдоподобия имеет максимум, называется оценкой максимального правдоподобия. Она, к сожалению, не связана с каким-либо принципом оптимальности и не обладает, например, свойством несмещенности, однако легко находится и имеет хорошие асимптотические свойства (состоятельность).

Лекц. 11

Теорема Гаусса-Маркова.

Предположим, что наблюдению доступны лишь линейные комбинации неизвестных величин (наблюдения *косвенные*)

$$\xi_i = \sum_{j=1}^{k} a_{ij}\alpha_j + \nu_i,\ i = 1, 2, \ldots, n. \tag{24}$$

Пусть ν_i — независимые сл. величины, $\mathsf{M}\mu = 0$, $\mathsf{D}\nu = \sigma^2$, $i = 1, 2, \ldots, n$.

Требуется оценить α_j, $j = 1, 2, \ldots, k$, точнее, найти линейные несмещенные оценки $\hat\alpha_j$ с минимальной дисперсией.

Запишем (24) в виде

$$\xi = \sum_{j=1}^{k} a_j\alpha_j + \nu, \tag{25}$$

где $a_j = (a_{1j}, a_{2j}, \ldots, a_{nj})^*$, $\xi, \nu, a_j \in \mathcal{R}_n$, $n \geqslant k$, векторы-столбцы a_j линейно независимы, или в виде

$$\xi = A\alpha + \nu, \quad \alpha \in \mathcal{R}_k, \tag{26}$$

причем $\mathsf{M}\nu = 0$, $\mathsf{M}\nu\nu^* = \sigma^2 I$.

1. (Линейность) Будем искать оценку α_j в виде $\hat\alpha_j = \sum\limits_{i=1}^{n} b_{ji}\xi_i = (b_j, \xi)$, $b_j = (b_{j1}, b_{j1}, \ldots, b_{jn})^*$.

2. Требование несмещенности дает:

$$\mathsf{M}\hat\alpha_j = \sum_{i=1}^{n} b_{ji}\sum_{s=1}^{k} a_{is}\alpha_s = \sum_{s=1}^{k}\left(\sum_{i=1}^{n} b_{ji}a_{is}\right) = \alpha_j,\ j = 1, \ldots, k. \tag{27}$$

Отсюда $(b_j, a_s) = \delta_{js}$, $j, s = 1, 2, \ldots, k$.

3. Вычислим дисперсию $\mathsf{D}\hat\alpha_j = \mathsf{D}\sum\limits_{i=1}^{n} b_{ji}\xi_i = \sigma^2\sum\limits_{i=1}^{n} b_{ji}^2 = \sigma^2||b_j||^2$.

Требование минимальности дисперсии приводит к следующей задаче на условный экстремум:

Для каждого $j = 1, 2, \ldots, k$ найти $\min||b_j||^2$ при условиях $(b_j, a_s) = \delta_{js}$, $s = 1, 2, \ldots, k$. Воспользуемся методом множителей Лагранжа[19]: введем функцию Лагранжа

$$\mathcal{L} = ||b_j||^2 - 2\sum_{s=1}^{k}\lambda_{js}(b_j, a_s) \tag{28}$$

[19]Для нахождения минимума $\varphi(x)$ при условиях $g_i(x) = 0$, $i = 1, 2, \ldots, m$, нужно, чтобы градиент $\operatorname{grad}\varphi(x)$ был ортогонален всем поверхностям $g_i(x) = 0$, $i = 1, 2, \ldots, m$, т.е. градиент $\operatorname{grad}\varphi(x)$ может быть разложен по векторам $\operatorname{grad} g_i(x))$, $i = 1, 2, \ldots, m$: $\operatorname{grad}[\varphi(x) - \sum\limits_{i=1}^{m}\lambda_i g_i(x)] = 0$ при некоторых λ_i. Выражение в квадратных скобках — так называемая функция Лагранжа.

и, дифференцируя по b_{ji}, получаем $b_j = \sum\limits_{s=1}^{k} \lambda_{js} a_s$. Используем условие несмещенности: $(b_j, a_p) = \sum\limits_{s=1}^{k} \lambda_{js}(a_s, a_p) = \delta_{jp}$. откуда $\lambda_{js} = (a_j, a_s)^{-}$ и окончательно[20]

$$\hat{\alpha}_j = \sum_{s=1}^{k} (a_j, a_s)^{-}(a_s, \xi).$$

Поскольку в векторно-матричной форме $(a_j, a_s) = (A^*A)_{js}$, то

$$\hat{\alpha} = (A^*A)^{-1}A^*\xi. \tag{29}$$

Найдем матрицу ковариаций $\hat{\alpha}$. Поскольку

$$\hat{\alpha} - \alpha = (A^*A)^{-1}A^*\xi - \alpha = (A^*A)^{-1}A^*(A\alpha + \nu) - \alpha =$$
$$= (A^*A)^{-1}A^*A\alpha + (A^*A)^{-1}A^*\nu - \alpha = (A^*A)^{-1}A^*\nu,$$

то

$$\mathsf{M}(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^* = \mathsf{M}(A^*A)^{-1}A^*\nu\nu^*A(A^*A)^{-1} = \sigma^2(A^*A)^{-1}. \tag{30}$$

Рассмотрим *метод наименьших квадратов*. Пусть $\tilde{\alpha}$ выбираются из условия[21]

$$\sum_{i=1}^{n} (\xi_i - \sum_{j=1}^{k} a_{ij}\alpha_j)^2 \sim \min_{\alpha_j}.$$

Дифференцируя по α_s, получим

$$2\sum_{i=1}^{n}(\xi_i - \sum_{j=1}^{k} a_{ij}\tilde{\alpha}_j)a_{is} = 0 \Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{k} a_{ij}a_{is}\tilde{\alpha}_j = \sum_{i=1}^{n} a_{is}\xi_i, \quad s = 1, 2, \ldots, k. \tag{31}$$

Отсюда получаем

$$\tilde{\alpha} = (A^*A)^{-1}A^*\xi,$$

т.е. ту же оценку,что и $\hat{\alpha}$.

Таким образом, справедлива **Теорема Гаусса-Маркова**:

Пусть ξ измеряется по схеме (24). Тогда ЛНОМД дается формулой (29), а матрица ковариаций — формулой (30). □

Как оценить σ^2?

Заметим, что из (31) следует $(\xi - A\hat{\alpha})a_j = 0$, $j = 1, 2, \ldots, k$, т.е.

$$(I - A(A^*A)^{-1}A^*)\xi \perp \mathcal{L}(a_1, \ldots, a_k) = (I - \Pi_a)\xi.$$

Таким образом, $\Pi_a = A(A^*A)^{-1}A^*$ — отогональный проектор на $\mathcal{L}(a_1, \ldots, a_k)$ (это можно проверить непосредственно).

[20] Знак $^-$ говорит о том, что берется соответствующий элемент матрицы, обратной к матрице $||(a_j, a_s)||$.

[21] Здесь не делается никаких предположений о ξ_i, $i = 1, 2, \ldots, n$.

Пусть $k < n$. Обозначим

$$\begin{aligned} s^2 &= ||\xi - A\alpha||^2 = ||\nu||^2, \\ s_1^2 &= ||\xi - \Pi_a\xi||^2 = ||\xi - \Pi_a(A\alpha+\nu)||^2 = ||(I-\Pi_a)\nu||^2, \\ s_2^2 &= ||\Pi_a\xi - A\alpha||^2 = ||\Pi_a(\xi - A\alpha)||^2 = ||\Pi_a\nu||^2. \end{aligned}$$

Далее, $\mathsf{M}s^2 = \operatorname{tr}\sigma^2 I = n\sigma^2$, $\mathsf{M}s_1^2 = \sigma^2\operatorname{tr}(I-\Pi_a) = \sigma^2(n-k)$.
Отсюда $\hat\sigma^2 = \frac{1}{n-k}s_1^2 = \frac{1}{n-k}||\xi - \Pi_a\xi||^2$ — несмещеная оценка σ^2.

Доверительные множества в нормальной регрессии.

Доверительные множества — аналог интервалов в интервальных оценках.

Пусть $\nu \sim \mathcal{N}(0,\sigma^2 I)$. Тогда $s_2^2 = ||\Pi_a\xi - A\alpha||^2 = ||\Pi_a\nu||^2 = \sigma^2\chi_k^2$, $s_1^2 = ||\xi - \Pi_a\xi||^2 = ||(I-\Pi_a)\nu||^2 = \sigma^2\chi_{n-k}^2$ и независимы, поэтому

$$F_{k,n-k} = \frac{\frac{1}{k}\chi_k^2}{\frac{1}{n-k}\chi_{n-k}^2}.$$

Пусть $P\{F_{k,n-k} \leqslant \varepsilon\} = \gamma_F(\varepsilon)$, тогда с вероятностью $\gamma_F(\varepsilon)$

$$||A(\alpha-\hat\alpha)||^2 = (A^*A(\alpha-\hat\alpha),(\alpha-\hat\alpha)) \leqslant \varepsilon\frac{k}{n-k}||(I-\Pi_a)\xi||^2. \qquad (32)$$

Левая часть неравенства (32) представляет собой квадратичную форму относительно координат α с матрицей $A^*A > 0$, поэтому (32) определяет в координатах α_j эллипсоид с центром $\hat\alpha$ (доверительный эллипсоид Хотеллинга).

Если нам нужно оценить одну координату α_j, то вспомним, что ее дисперсия равна $\sigma^2(a_j,a_j)^-$, поэтому $\frac{\alpha_j-\hat\alpha_j}{\sqrt{\sigma^2(a_j,a_j)^-}} \sim \mathcal{N}(0,1)$, а

$$\frac{\alpha_j-\hat\alpha_j}{\sqrt{(a_j,a_j)^-\frac{1}{n-k}||(I-\Pi_a)\xi||^2}} = t_{n-k},$$

и если $P\{|t_{n-k}| < \varepsilon\} = \gamma_t(\varepsilon)$, то с вероятностью $\gamma_t(\varepsilon)$ неравенство

$$|\alpha_j-\hat\alpha_j| \leqslant \varepsilon\sqrt{\frac{(a_j,a_j)^-||(I-\Pi_a)\xi||^2}{n-k}}$$

дает интервальную оценку α_j.

Наконец, статистика s_1^2 дает возможность проверить адекватность модели измерения (обозначим ее $[A,\sigma^2]$). Если модель верна, то, как показано ранее, $s_1^2 = \sigma^2\chi_{n-k}^2$ и не зависит от измеряемого параметра (сигнала) α. Поэтому вероятность $P\{\chi_{n-k}^2 < \frac{s_1^2}{\sigma^2}\}$ характеризует адекватность модели данным измерениям.

Задачи редукции измерений. Л. 12

1° Постановка задачи несмещенной редукции измерений.

Для схемы измерений $\xi = Af + \nu$, $\mathsf{M}\nu = 0$, $\mathsf{M}\nu\nu^* = \sigma^2 I$ ставится **задача несмещенной редукции**:

$$\inf\{\mathsf{M}||R\xi - f||^2 \mid R,\ RA = I\} = \inf\{\sigma^2\operatorname{tr}RR^* \mid R,\ RA = I\} = h_0.$$

Пусть $A^*A > 0$ ($\operatorname{rank} A = k \leqslant n$).

Решаем уравнение $RA = I$: $R = R_0 + Y$, где $R_0 = (A^*A)^{-1}A^*$, а Y — решение уравнения $YA = 0 \Leftrightarrow Y\Pi_a = 0 \Leftrightarrow Y = Z(I - \Pi_a)$, $\forall Z$.

Т.о., общее решение $R = (A^*A)^{-1}A^* + Z(I - \Pi_a)$.

В этом случае $\operatorname{tr} RR^* = \operatorname{tr}(A^*A)^{-1} + \operatorname{tr} Z(I - \Pi_a)Z^*$ и $\inf$ достигается на $R = R_0 = (A^*A)^{-1}A^*$ и равен $h_0 = \sigma^2 \operatorname{tr}(A^*A)^{-1}$. Очевидно, этот результат совпадает с результатом, полученным в теореме Гаусса-Маркова.

В этом случае $R\xi = f + R\nu$, где $R\nu$ — шум, суммарная энергия которого равна h_0.

2° Задача редукции с ограничением на уровень шума. Часто шум, полученный при решении задачи несмещенной редукции, неприемлемо велик. Вспомним, что ошибка складывается из двух:

$$R\xi = f + (RA - I)f + R\nu.$$

Введем расстояние между матрицами[22] (операторами) A и B: $\rho^2(A - B) = \operatorname{tr}(A - B)(A - B)^*$.

Поставим задачу **синтеза прибора с ограничением на уровень шума**:

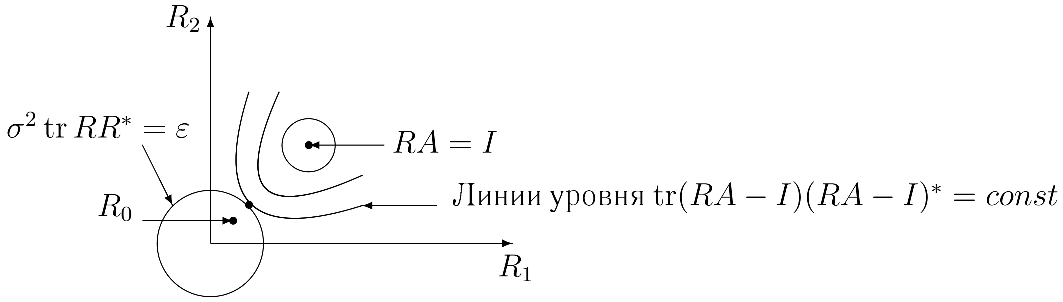
$$\inf\{\operatorname{tr}(RA - I)(RA - I)^* \mid R,\ \mathsf{M}||R\nu||^2 \leqslant \varepsilon\}. \tag{33}$$

Заметим, что $\mathsf{M}||R\nu||^2 = \sigma^2 \operatorname{tr} RR^* = \sigma^2 \sum\limits_{i,j} r_{ij}^2$.

Далее рассмотрим два случая.

1. $h_0 \leqslant \varepsilon$. В этом случае условие в задаче (33) выполняется и $R_0 = (A^*A)^{-1}A^*$ — есть решение, так как любое $R = (A^*A)^{-1}A^* + Z(I - \Pi_a)$ минимизирует $||RA - I||_2^2$.

2. Введем систему координат в пространстве матричных элементов (изобразим лишь два!):



Очевидно, решение есть точка касания, где $\sigma^2 \operatorname{tr} RR^* = \varepsilon$ (равно!). Тогда решаем задачу методом множителей (одного!) Лагранжа. Функция Лагранжа

$$L(R) = \operatorname{tr}(RA - I)(RA - I)^* + \omega\sigma^2 \operatorname{tr} RR^*.$$

$$\nabla_R L = 2(RA - I)A^* + 2\omega\sigma^2 R = 0.$$

$R(AA^* + \omega\sigma^2 I) = A^*$, $R = R(\omega) = A^*(AA^* + \omega\sigma^2 I)^{-1} = (A^*A + \omega\sigma^2 I)^{-1}A^*$. Обозначим $h = \sigma^2 \operatorname{tr} R(\omega)R^*(\omega) = \sigma^2 \operatorname{tr}(A^*A + \omega\sigma^2 I)^{-1}A^*A(A^*A + \omega\sigma^2 I)^{-1}$ и[23] $g(\omega) = \operatorname{tr}(RA - I)(RA - I)^* = \omega^2\sigma^4 \operatorname{tr}(A^*A + \omega\sigma^2 I)^{-2}$.

[22] Можно также ввести скалярное произведение $(AB)_2 = \operatorname{tr} AB^*$ и норму $||A||_2 = \{\operatorname{tr} AA^*\}^{1/2}$, которая называется нормой Гильберта-Шмидта.

[23] Здесь используется равенство $I - RA = I - (A^*A + \omega\sigma^2 I)^{-1}A^*A = \omega\sigma^2(A^*A + \omega\sigma^2 I)^{-1}$.

Пусть далее $\{e_i\}$ — ортонормированный базис из собственных векторов оператора A^*A: $A^*Ae_i = \lambda_i e_i$, $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_k > 0$[24]. Тогда

$$h = \sigma^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + \omega\sigma^2)^2}, \quad g = \sigma^4 \sum_{i=1}^{k} \frac{\omega^2}{(\lambda_i + \omega\sigma^2)^2}.$$

Вычислим призводные: $\frac{dh}{d\omega} = -2\sigma^4 \sum\limits_{i=1}^{k} \frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3} < 0$, $0 < \omega < \infty$, $h \underset{\omega\to 0}{\longrightarrow} h_0 = \sigma^2 \sum\limits_{i=1}^{k} \frac{1}{\lambda_i} = \sigma^2 \operatorname{tr}(A^*A)^{-1}$, $h \underset{\omega\to 0}{\longrightarrow} 0$. Поэтому уравнение $h(\omega) = \varepsilon$ при $\varepsilon < \varepsilon_0 = h_0$ имеет единственное решение.

Кроме того, $\frac{dg}{d\omega} = \sigma^4 \sum\limits_{i=1}^{k} \left[\frac{2\omega}{(\lambda_i+\omega\sigma^2)^2} - \frac{2\omega^2\sigma^2}{(\lambda_i+\omega\sigma^2)^3}\right] = 2\omega\sigma^4 \sum\limits_{i=1}^{k} \frac{\lambda_i}{(\lambda_i+\omega\sigma^2)^3}$ и поэтому имеет место дифференциальный закон сохранения:

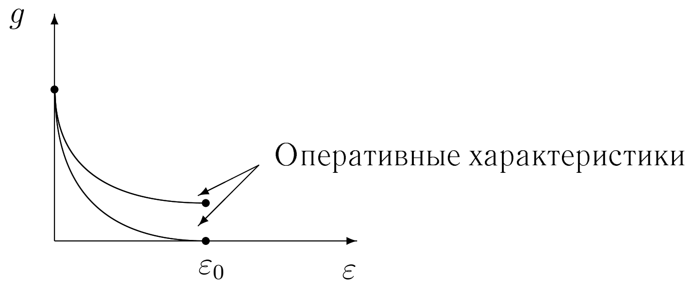
$$\omega\frac{dh}{d\omega} + \frac{dg}{d\omega} = 0. \tag{34}$$

Итак, общее решение задачи (37) имеет вид:

$$R = \begin{cases} R(\omega) = (A^*A + \omega\sigma^2 I)^{-1}, & 0 < \varepsilon < \varepsilon_0 = h_0 = \sigma^2 \operatorname{tr}(A^*A)^{-1}, \\ 0 & \varepsilon - 0, \\ R_0 = (A^*A)^{-1}A^*, & \varepsilon \geqslant \varepsilon_0 = h_0, \end{cases} \tag{35}$$

при этом выполняется (34).

Зависимость g от ε носит название оперативной характеристики. При этом характеристика, график которой лежит ниже, соответствует равномерно лучшему прибору.



[24] $\operatorname{rank} A = k \leqslant n$.

ция 13

СЛУЧАЙНЫЕ ПРОЦЕССЫ.

Пусть задано верятностное пространство $(\Omega, \mathcal{F}, P(\cdot))$ и некоторое множество T значений t (непрер. или дискр.). Случайной функцией называется $\xi(t) = \xi(\omega, t)$, такая, что $\forall t_0 \in T$ $\xi(t_0)$ — сл. величина. Если зафиксировать $t = t_0$, то получим *сечение* сл. функции (или процесса, если t интерпретировать как время). Если зафиксировать $\omega = \omega_0$, то полученная $\xi(\omega_0, t)$ есть *выборочная функция* (реализация) сл. процесса. Распределение вероятностей в каждом сечении t можно задать функцией $F(x, t) = P\{\xi(t) < x\}$, но она не дает никакого представления о связи сл. величин, характеризующих разные сечения. Более полным является задание семейства функций совместного распределения $F(x_1, t_1, \dots, x_n, t_n)$ для n сечений, $n = 1, 2, \dots$. Однако при этом должжно быть выполнено условие *согласованности*, касающееся редукции (уменьшению) числа аргументов и их перестановок.

При каком n функции совместного распределения достаточно для описания процесса? Если рассматривать дискретные значения параметра t, то для схемы независмых испытаний $n = 1$, для марковских цепей $n = 2$. Обычно этим $(n = 2)$ ограничиваются в случае непрерывного времени для так называемых процессов с независимыми приращениями.

Рассмотрим в качестве примера такие два процсса 2 порядка.

Процесс Пуассона (одномерный случай).

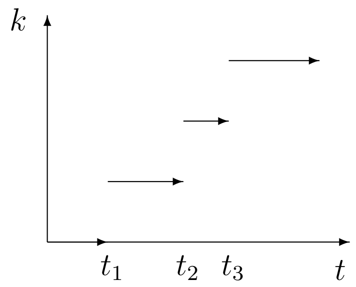
Определение. Сл. функция $\eta(t)$ $0 \leqslant t < \infty$, называется процессом Пуассона или пуассоновским потоком событий, если

1. для любых $0 \leqslant t_1 < \dots < t_n$ сл. величины $\eta(t_i) - \eta(t_{i-1})$, $i = 1, 2, \dots, n$, независимы в совокупности (процесс с независимыми приращениями),

2. сл. величина $\eta(t) - \eta(s)$, $0 \leqslant s < t$, имеет распределение Пуассона с параметром $\lambda(t - s)$:

$$P\{\eta(t) - \eta(s) = k\} = \frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}.$$

3. Если $\eta(0) = 0$, то говорят, что процесс *начинается в нуле*.



Выборочная функция Пуассоновского процесса

Теорема. Пусть $\eta(t)$ — процесс с независимыми приращениями и пусть выполняются условия (при $t \to 0$):

а) $P\{\eta(t) - \eta(0) = 1\} = \lambda t + o(\lambda t)$,

б) $P\{\eta(t) - \eta(0) > 1\} = o(\lambda t)$.

Тогда $\eta(t)$ — процесс Пуассона.

Доказательство. В силу независимости приращений достаточно доказать, что выполняется пункт 2 в определении.

Разобьем интервал $t = n\Delta$ на n (одинаковых) подинтервалов Обозначим через Δ_i, $i = 1, 2, \dots, n$, i-ый подынтервал. При $(n > k)$ событие $\{\eta(t) = k\}$ представим в виде $A + B$, где A — событие, в котором в каждый подынтервал попадает не более 1 точки, B — событие, в котором по крайней мере в один подынтервал попадает более одной точки. Тогда $P(A) = C_n^k p^k (1-p)^{n-k}$, где $p = \lambda\Delta + o(\lambda\Delta)$. Поскольку $pn = \lambda n\Delta + n\ o(\frac{\lambda t}{n}) \underset{n\to\infty}{\longrightarrow} \lambda t$, то применяя теорему Пуассона, получаем $P(A) \underset{n\to\infty}{\longrightarrow} \frac{(\lambda t)^k}{k!} e^{\lambda t}$. Аналогично,

$$P(B) < \sum_{i: P\{\eta(\Delta_i) > 1\}} P\{\eta(\Delta_i) > 1\} < n\ o\left(\frac{\lambda t}{n}\right) \underset{n\to\infty}{\longrightarrow} 0. \quad \square$$

Задача. Найти распределение сл. величины — времени ожидания τ первого события в пуассоновском потоке событий.

Из условия $P\{\tau \geqslant t\} = P\{\eta(t) - \eta(0) = 0\} = e^{-\lambda t}$ получаем $F_\tau(t) = P\{\tau < t\} = 1 - e^{-\lambda t}$, $t \geqslant 0$.

Плотность вероятности $p_\tau(t) = \lambda e^{-\lambda t}$, $t \geqslant 0$.

Среднее время ожидания $\mathsf{M}\tau = \int\limits_0^\infty t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}$.

Заметим, что из свойств сл. процесса Пуассона следует, что время ожидания не зависит от момента начала ожидания, а лишь от величины интервала ожидания.

Радиоактивный распад. Вероятность распада для каждого атома $P\{\eta(t) > 0\} = P\{\tau < t\} = 1 - e^{-\lambda t}$. При $t = t_0$, $1 - e^{-\lambda t_0} = \frac{1}{2}$, половина всего вещества распадается, так что $e^{-\lambda t_0} = \frac{1}{2}$, $\lambda t_0 = \ln 2$, $t_0 = \frac{\ln 2}{\lambda}$ — время полураспада.

Винеровский сл. процесс. Случайный процесс $\xi(t)$ называется Винеровским, если

1. для $0 = t_0 \leqslant t_1 < t_2 < \dots < t_n$ $\eta_i = \xi(t_i) - \xi(t_{i-1})$ — независимы в совокупности,
2. $\xi(t) - \xi(s) \sim \mathcal{N}(0, t-s)$, $0 < s < t$,
3. $\xi(0) = 0$ (начинается в нуле).

Таким образом, вектор $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ распределен с плотностью

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_1 - t_{i-1})}} \exp\left\{-\frac{x_i^2}{2(t_1 - t_{i-1})}\right\}.$$

Если перейти к $\xi_i = \sum\limits_{k=1}^i \xi_k$ с помощью невырожденного преобразования с матрицей A вида

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

то получим для вектора ξ плотность

$$p_\xi(x,t) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(t_1 - t_{i-1})}} \exp\left\{ -\frac{(x_i - x_{i-1})^2}{2(t_1 - t_{i-1})} \right\}.$$

Пример. Броуновское движение (одномерная задача).

Пусть $\xi(t)$ — координата броуновской частицы на прямой, $\xi(0) = 0$.

$$\xi(t+s) = [\xi(t+s) - \xi(s)] + [\xi(s) - \xi(0)], \quad t, s > 0.$$

Из однородности следует, что слагаемые в правой части независимы и поэтому $\mathsf{D}\xi(t+s) = \mathsf{D}\xi(t) + \mathsf{D}\xi(s)$, откуда $\mathsf{D}\xi(t) = \sigma^2 t$, где σ^2 — некоторая константа.

Рассмотрим дискретный аналог — случайное блуждание по одномерной сетке в дискретном времени: $\xi_n = \sum\limits_{i=1}^{n} \xi_i$, $P\{\xi_i = \pm\Delta x\} = \frac{1}{2}$, $t = n\Delta t$.

$\mathsf{M}\xi_n = \sum \mathsf{M}\xi_i = 0$, $\mathsf{D}\xi_n = \sum \mathsf{D}\xi_i = n(\Delta x)^2$, приравнивая $\sigma^2 n\Delta t$, получим, что $\frac{(\Delta x)^2}{\Delta t} = \sigma^2 = const$. Переходя к пределу при $\Delta x \to 0$, $\Delta t \to 0$, $\frac{(\Delta x)^2}{\Delta t} = \sigma^2$, получаем

$$\frac{\xi_n}{\sigma\sqrt{t}} \xrightarrow[n\to\infty]{d} \frac{\xi(t)}{\sigma\sqrt{t}} \sim \mathcal{N}(0,1).$$

Отсюда $\xi(t) \sim \mathcal{N}(0, \sigma^2 t)$ или $\xi(t) - \xi(s) \sim \mathcal{N}(0, \sigma^2(t-s))$.

В общем случае (если нет симметрии)

$\xi(t) - \xi(s) \sim \mathcal{N}(\theta(t-s), \sigma^2(t-s))$, где θ — коэффициент сноса.

Характеризация винеровского процесса.

Теорема. Пусть сл. процесс удовлествояет условиям:

1) для любых непересекающихся промежутков приращения независимы;

2) для любого промежутка вероятность зависит лишь от длины промежутка (однородность);

3) для функции распределения приращения $\xi(t_0+t) - \xi(t_0)$, $F(x,t)$ существует плотность $p(x,t)$, имеющая непрерывные и ограниченные по x, $-\infty < x < \infty$, производные до 3-го порядка $\forall t > 0$;

4) для малых Δt малые приращения более вероятны, чем большие, а именно

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} x p(x, \Delta t) dx = A; \tag{36}$$

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} x^2 p(x, \Delta t) dx = B > 0; \tag{37}$$

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} |x|^3 p(x, \Delta t) dx = 0; \tag{38}$$

Тогда $\xi(t)$ — (обобщенный) винеровский процесс.

Доказательство. Рассмотрим два смежных промежутка $(0, t)$ и $(t, t + \Delta t)$, а также его сумму (объединение). Пользуясь формулой для плотности суммы независимых сл. величин, получаем

$$p(x, t + \Delta t) = \int\limits_{-\infty}^{\infty} p(x - s, t) p(s, \Delta t) ds. \tag{39}$$

Разлагая сомножитель $p(x - s, t)$ в ряд Тейлора по степеням s, получим

$$p(x - s, t) = p(x, t) - \frac{\partial p(x, t)}{\partial x} s + \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2} s^2 - \frac{1}{6} \frac{\partial^3 p(\tilde{x}, t)}{\partial x^3} s^3,$$

подставляя это в (39), получаем

$$p(x, t + \Delta t) - p(x, t) = -\frac{\partial p(x, t)}{\partial x} \int\limits_{-\infty}^{\infty} s p(s, \Delta t) ds +$$

$$+\frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2} \int\limits_{-\infty}^{\infty} s^2 p(s, \Delta t) ds - \frac{1}{6} \int\limits_{-\infty}^{\infty} \left[\frac{\partial^3 p(\tilde{x}, t)}{\partial x^3} \right] s^3 p(s, \Delta t) ds.$$

Разделим на Δt и перейдем к пределу при $\Delta t \to 0$, учитывая свойства (36, 37, 38):

$$\frac{\partial p(x, t)}{\partial t} = -A \frac{\partial p(x, t)}{\partial x} + \frac{B}{2} \frac{\partial^2 p(x, t)}{\partial x^2}, \quad \int\limits_{-\infty}^{\infty} p(x, t) dx = 1, \quad p(x, t) \geqslant 0.$$

Подстановкой $y = x - At$, $\tau = Bt$, $p(x, t) = \bar{p}(y, \tau)$ приводим к

$$\frac{\partial \bar{p}}{\partial \tau} = \frac{1}{2} \frac{\partial^2 \bar{p}}{\partial y^2}, \quad -\infty < y < \infty$$

— уравнению теплопроводности с решением

$$\bar{p}(y, \tau) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{y^2}{2\tau}}, \quad \int\limits_{-\infty}^{\infty} \bar{p} dy = 1.$$

В прежних переменных это выглядит так:

$$p(x, t) = \frac{1}{\sqrt{2\pi Bt}} e^{-\frac{(x - At)^2}{2Bt}}. \quad \square$$

Задача. Найти распределение τ_x — времени первого достижения броуновской частицей точки x $(x > 0)$.

$$P\{\xi(t) > x | \tau_x < t\} = \frac{1}{2} = \frac{P\{\xi(t) > x\}}{P\{\tau_x < t\}}.$$

Отсюда $P\{\tau_x < t\} = 2P\{\frac{\xi(t)}{\sqrt{t}} > \frac{x}{\sqrt{t}}\} = \sqrt{\frac{2}{\pi}} \int\limits_{\frac{x}{\sqrt{t}}}^{\infty}$ и $p_{\tau_x} = \frac{1}{\sqrt{2\pi}} \frac{x}{t^{3/2}} e^{-\frac{x^2}{2t}}$.

Аналогично для $\xi_t = \max\limits_{0 \leqslant s \leqslant t} \xi(s)$ — максимальной координаты броуновской частицы имеем: $P\{\max\limits_{0 \leqslant s \leqslant t} \xi(s) > x\} = P\{\tau_x < t\}$ и $p_{\xi_t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$, $x \geqslant 0$.

Интересно, что $\forall x$: $P\{\tau_x < \infty\} = 1$ и $\forall t$: $P\{\max\limits_{0 \leqslant s \leqslant t} \xi(s) > 0\} = 1$.

ция 14 **Теория второго порядка (корреляционная теория).** Далее будут встречаться комплексные случайные процессы $\xi(t) = \eta(t) + i\zeta(t)$, где $\eta(t)$ и $\zeta(t)$ - действительные случайные процессы.

Корреляционной функцией случайного процесса $\xi(t)$ называется[25]

$$K(t,s) = \mathsf{M}(\xi(t) - \mathsf{M}\xi(t))(\overline{\xi(s)} - \mathsf{M}\overline{\xi(s)}).$$

Примеры. (1) Корреляционная функция винеровского случайного процесса, $\xi(t) - \xi(s) \sim \mathcal{N}(0, t-s)$, $s < t$. При $0 < s < t$

$$\mathsf{M}\xi(t)\xi(s) = \mathsf{M}[(\xi(t) - \xi(s)) + (\xi(s) - \xi(0))](\xi(s) - \xi(0)) = \mathsf{M}(\xi(s) - \xi(0))^2 = s.$$

При $0 < t < s$ аналогично получаем $\mathsf{M}\xi(t)\xi(s) = t$. Следовательно, $\mathsf{M}\xi(t)\xi(s) = \min(t,s)$.

(2). Корреляционная функция пуассоновского процесса,

$$P(\xi(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \mathsf{M}\xi(t) = \lambda t.$$

Для $0 < s < t$:

$$K(t,s) = \mathsf{M}(\xi(t) - \lambda t)(\xi(s) - \lambda s) = \mathsf{M}[(\xi(t) - \xi(s) - \lambda(t-s)) +$$

$$+ (\xi(s) - \lambda s)](\xi(s) - \lambda s) = \mathsf{M}(\xi(s) - \lambda s)^2 = \lambda s.$$

Поэтому и в этом случае $K(t,s) = \lambda \min(t,s)$.

В обоих примерах использована независимость приращений.

Определение. Случайный процесс $\xi(t)$ называется непрерывным в среднем квадратичном в точке t, если

$$\mathsf{M}|\xi(t+h) - \xi(t)|^2 \to 0 \text{ при } h \to 0,$$

или, иначе говоря, если

$$\xi(t) = \underset{h \to 0}{\text{l.i.m.}}\, \xi(t+h).$$

Речь идет о случайных процессах с конечными моментами второго порядка; $|\mathsf{M}\xi| \leqslant \mathsf{M}|\xi| \leqslant [\mathsf{M}|\xi|^2]^{1/2} < \infty$, в этом случае и $\mathsf{M}|\xi| < \infty$.

Напомним, что последовательность $\{\xi_k\}$ сходится в среднем квадратичном к ξ, если $\mathsf{M}|\xi_k - \xi|^2 \to 0,\ k \to \infty;\ \xi = \underset{k \to \infty}{\text{l.i.m.}}\, \xi_k$.

Свойства с.к. сходимости.

Критерий 1. (фундаментальность) Для сходимости в с.к. последовательности $\{\xi_k\}$ необходимо и достаточно, чтобы $\mathsf{M}|\xi_k - \xi_n|^2 \to 0$ при $k, n \to \infty$. [26]

[25]Черта означает комплексное сопряжение.

[26](Если бы M символизировало римановское интегрирование, не было бы полноты).

Критерий 2. Последовательность $\{\xi_k\}$ сходится в среднем квадратичном, если и только если числовая последовательность $\mathsf{M}\xi_k\overline{\xi_n} \to \mathsf{M}\xi\overline{\xi}$ при $k, n \to \infty$ независимо.

Доказательство. $\mathsf{M}(\xi_k - \xi_n)(\overline{\xi_k} - \overline{\xi_n}) = \mathsf{M}\xi_k\overline{\xi_k} + \mathsf{M}\xi_n\overline{\xi_n} - \mathsf{M}\xi_k\overline{\xi_n} - \mathsf{M}\xi_n\overline{\xi_k} \to 0$, если $M\xi_k\overline{\xi}_n$ сходится.

Наоборот, $\mathsf{M}\xi_k\overline{\xi_n} - \mathsf{M}\xi\overline{\xi} = \mathsf{M}(\xi_k - \xi)(\overline{\xi_n} - \overline{\xi}) + \mathsf{M}\xi(\overline{\xi_n} - \overline{\xi}) + \mathsf{M}(\xi_k - \xi)\overline{\xi} \to 0$, если $\xi_k \xrightarrow[n\to\infty]{\text{с.к.}} \xi$. □

Другие свойства: если $\xi = \operatorname*{l.i.m.}_{k\to\infty} \xi_k$, то

(1) $\mathsf{M}\xi = \lim\limits_{k\to\infty} \mathsf{M}\xi_k$. Действительно, $|\mathsf{M}\xi - \mathsf{M}\xi_k| \leqslant \mathsf{M}|\xi - \xi_k| \leqslant (\mathsf{M}|\xi - \xi_k|^2)^{1/2} \to 0,\ k \to \infty$.

(2) $\mathsf{M}|\xi|^2 = \lim\limits_{k\to\infty} \mathsf{M}|\xi_k|^2$. Доказывается на основе критерия 2 или непосредственно: $\mathsf{M}|\xi-\xi_k|^2 = \mathsf{M}|\xi|^2 - \mathsf{M}\xi\overline{\xi_k} - \mathsf{M}\xi_k\overline{\xi} + \mathsf{M}|\xi_k|^2 \geqslant \mathsf{M}|\xi|^2 - 2(\mathsf{M}|\xi|^2\mathsf{M}|\xi_k|^2)^{1/2} + \mathsf{M}|\xi_k|^2 = (\sqrt{\mathsf{M}|\xi|^2} - \sqrt{\mathsf{M}|\xi_k|^2})^{1/2}$. □

(3) Если, кроме того, $\eta = \operatorname*{l.i.m.}_{k\to\infty} \eta_k$, то $\mathsf{M}\xi\overline{\eta} = \lim\limits_{n,k\to\infty} \mathsf{M}\xi_n\overline{\eta_k}$. Действительно, $|\mathsf{M}(\xi_n\bar{\eta}_k - \xi\bar{\eta})| \leqslant \mathsf{M}|\xi_n\bar{\eta}_k - \xi\bar{\eta}| \leqslant \mathsf{M}|\xi_n\bar{\eta}_k - \xi_n\bar{\eta} + \xi_n\bar{\eta} - \xi\bar{\eta}| \leqslant \mathsf{M}|\xi_n\bar{\eta}_k - \xi_n\bar{\eta}| + \mathsf{M}|\xi_n\bar{\eta} - \xi\bar{\eta}| \leqslant \sqrt{\mathsf{M}|\xi_n|^2\mathsf{M}|\eta_k - \eta|^2} + \sqrt{\mathsf{M}|\xi_n - \xi|^2\mathsf{M}|\eta|^2} \xrightarrow[n,k\to\infty]{} 0$. □

Вернемся к сл. процессам. Всюду далее будем считать $\mathsf{M}\xi(t) = 0$.

Теорема.

1. Для того, чтобы случайный процесс $\xi(t)$ был с.к. непрерывен в точке t_0 необходимо и достаточно, чтобы $K(t, s)$ была непрерывна в точке (t_0, t_0).

2. Если $K(t, s)$ непрерывна на диагонали $t = s$, то $K(t, s)$ непрерывна всюду.

Доказательство. 1. Достаточность. Пусть $K(t, s)$ непрерывна в точке (t_0, t_0). Тогда

$$\mathsf{M}|\xi(t_0+h) - \xi(t_0)|^2 = K(t_0+h, t_0+h) - K(t_0+h, t_0) - K(t_0, t_0+h) + K(t_0, t_0) \to 0$$

при $h \to 0$.

Необходимость. Пусть сл. процесс $\xi(t)$ с.к. непрерывен при $t = t_0$, тогда

$$K(t_0 + h, t_0 + k) - K(t_0, t_0) = \mathsf{M}[(\xi(t_0 + h) - \xi(t_0))(\overline{\xi(t_0 + k) - \xi(t_0)}) +$$
$$+ (\xi(t_0 + h) - \xi(t_0))\overline{\xi(t_0)} + \xi(t_0)(\overline{\xi(t_0 + k) - \xi(t_0)})].$$

Далее для каждого из слагаемых правой части воспользуемся свойством (3). Например, для первого:

$$|\mathsf{M}(\xi(t_0 + h) - \xi(t_0))(\overline{\xi(t_0 + k) - \xi(t_0)})| \leqslant$$
$$\leqslant \sqrt{\mathsf{M}|\xi(t_0 + h) - \xi(t_0))|^2\mathsf{M}|\xi(t_0 + k) - \xi(t_0))|^2} \xrightarrow[h\to 0, k\to 0]{} 0.$$

2. Если $K(t, s)$ непрерывна на диагонали, например, в точках $(t, t), (s, s)$, то $\xi(t + h) \xrightarrow[h\to 0]{\text{с.к.}} \xi(t)$ и $\xi(s + k) \xrightarrow[k\to 0]{\text{с.к.}} \xi(s)$. Следовательно, $\mathsf{M}\xi(t + h)\overline{\xi(s + k)} \to \mathsf{M}\xi(t)\overline{\xi(s)}$, т.е. $K(t + h, s + k) \to K(t, s)$ при $h, k \to 0$. □

Следствие. Пуассоновский и винеровский случайные процессы непрерывны в среднем квадратичном. Однако, реализации пуассоновского процесса разрывные (ступенчатые) функции.

Определение. Случайная функция $\xi(t)$ называется с.к. дифференцируемой в точке t, если существует с.к. предел $\frac{\xi(t+h)-\xi(t)}{h}$ при $h \to 0$. Этот предел называется с.к. производной $\xi(t)$:

$$\xi'(t) = \underset{h\to 0}{\mathrm{l.i.m.}} \frac{\xi(t+h)-\xi(t)}{h}, \text{ или } \mathsf{M}|\frac{\xi(t+h)-\xi(t)}{h} - \xi'(t)|^2 \to 0, h \to 0.$$

Согласно критерию 2, $\xi(t)$ с.к. дифференцируема в точке t, если и только если

$$Q = \frac{1}{hk}\mathsf{M}(\xi(t+h)-\xi(t))(\overline{\xi(t+k)-\xi(t)}) \text{ сходится при } h,k \to 0:$$

$$Q = \frac{K(t+h,t+k)-K(t,t+k)-K(t+h,t)+K(t,t)}{hk} \underset{h\to 0, k\to 0}{\longrightarrow} \frac{\partial^2 K(t,s)}{\partial t \partial s}|_{t=s}.$$

Это разностное отношение для второй производной сходится к правой части, если, например, $\frac{\partial^2 K(t,s)}{\partial t \partial s}$ непрерывна.

В этом случае $\frac{\partial^2 K(t,s)}{\partial t \partial s} = \mathsf{M}\xi'(t)\overline{\xi'(s)}$ - корреляционная функция с.к. производной $\xi(t)$. Это утверждение — следствие свойства 2: если $\frac{\xi(t+h)-\xi(t)}{h} \underset{h\to 0}{\longrightarrow} \xi'(t)$, $\frac{\xi(s+k)-\xi(s)}{k} \underset{k\to 0}{\longrightarrow} \xi'(s)$, то

$$\mathsf{M}\frac{\xi(t+h)-\xi(t)}{h}\overline{\frac{\xi(s+k)-\xi(s)}{k}} \underset{h\to 0, k\to 0}{\longrightarrow} \mathsf{M}\xi'(t)\overline{\xi'(s)} = \frac{\partial^2 K(t,s)}{\partial t \partial s},$$

если эта производная непрерывна.

Пример. Корреляционная функция производной (в обобщенном смысле) стандартного Винеровского процесса равна $\frac{\partial^2 \max(t,s)}{\partial t \partial s} = \delta(t-s)$. Такой процесс называется *белым шумом*.

Определение. Случайная функция $\xi(t)$ с.к. интегрируема по Риману на $[a,b]$, если последовательность римановских интегральных сумм $\sum\limits_{[a,b]} \xi(t_k)\delta t_k$ с.к. сходится при $\max\limits_k \delta t_k \to 0$.

Теорема. Случайная функция $\xi(t)$ с.к. интегрируема на $[a,b]$ по Риману, если $K(t,s)$ интегрируема по Риману на $[a,b]\times[a,b]$. Действительно, пусть

$$S = \sum_{[a,b]} \xi(t_k)\delta t_k, \quad S' = \sum_{[a,b]} \xi(t'_k)\delta t'_k$$

Тогда

$$MS\overline{S'} = \sum_{[a,b]\times[a,b]} K(t_i,t'_k)\delta t_i \delta t'_k$$

и факт существования предела справа эквивалентен с.к. интегрируемости $\xi(t)$ (с.к. сходимости S). □