Теория вероятностей и математическая статистика. Коллоквиум №3

Lev Khoroshansky

Содержание

Билет №1. **3**
- Вступление . . . 3
- Закон больших чисел . . . 3
  - Доказательство с помощью неравенства Чебышёва . . . 3
  - Вариация для схемы Бернулли . . . 4
- Неравенство Чернова . . . 4
- Комментарий . . . 6

Билет №2. **7**
- Вступление и сходимость почти наверное . . . 7
- Сохранение сходимости по вероятности при подстановке случайных величин в непрерывную функцию . . . 8
- Сходимость по вероятности выборочного среднего . . . 9
- Сходимость по вероятности выборочной дисперсии . . . 9
- Сходимость в среднем . . . 9
- Теорема Лебега . . . 10
- Усиленный закон больших чисел . . . 10

Билет №3. **11**
- Сходимость по распределению . . . 11
- Эквивалентные определения . . . 11
- Сохранение сходимости по распределению при подстановке случайных величин в непрерывную функцию . . . 12
- Сходимость двух последовательностей случайных величин, одна из которых сходится по распределению, а другая сходится по вероятности к константе . . . 12
  - Их сумма . . . 12
  - Их произведение . . . 14

Билет №4. **15**
- Характеристическая функция и её свойства . . . 15
- Характеристическая функция нормального распределения . . . 16
- Центральная предельная теорема . . . 17

Билет №5. **18**
- Вступление . . . 18
- Асимптотическая нормальность выборочного среднего . . . 18
- Теорема о непрерывности (не требуется) . . . 18
- Асимптотический доверительный интервал . . . 19

Билет №6. 20
Вступление 20
Характеристическая функция случайного вектора 20
Проверка независимости компонент случайного вектора с помощью характеристической функции 20
Изменение характеристической функции при афинном преобразовании случайного вектора 21

Билет №7. 22
Многомерное нормальное распределение 22
Равенство нулю ковариации и независимость 23
Распределение вектора, полученного линейным преобразованием из гауссовского . 24

Билет №8. 25
Многомерное нормальное распределение 25
Ортогонализация и получение гауссовского вектора линейным преобразованием из вектора, компоненты которого являются независимыми нормально распределёнными случайными величинами с параметрами 0 и 1 25
Вид плотности распределения в невырожденном случае 26

Билет №9. 28
Независимость выборочного среднего и выборочной дисперсии независимых нормально распределённых величин 28
Распределение хи-квадрат 28

Билет №10. 29
Условные математические ожидания относительно дискретных величин 29
Свойства 30
Решение задачи о наилучшем приближении 32
Геометрический смысл 33

Билет №1.

Вступление

Для начала вспомним, что такое неравенство Чебышёва **(лекция)**.

Th. *Пусть дана неотрицательная случайная величина* ξ *с конечным ожиданием* $\mathbb{E}\xi < \infty$. *Тогда для любого* $t > 0$ *справедливо неравенство*

$$\mathbb{P}(\xi \geqslant t) \leqslant \frac{\mathbb{E}\xi}{t}.$$

Доказательство. Рассмотрим интересующее нас множество исходов $A = \{\xi \geqslant t\}$.
Тогда заметим, что $\xi \geqslant t \cdot \mathrm{Ind}_A$. Возьмём с обеих сторон ожидание: $\mathbb{E}\xi \geqslant t \cdot \mathbb{P}(A)$. **[:|||:]**

Из этого сразу следует усиленное неравенство Чебышёва.

Cor. *Пусть дана случайная величина* ξ *с конечным ожиданием* $\mathbb{E}\xi < \infty$ *и конечной дисперсией* $\mathbb{D}\xi < \infty$. *Тогда для любого* $t > 0$ *справедливо неравенство*

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geqslant t) \leqslant \frac{\mathbb{D}\xi}{t^2}.$$

Доказательство. Достаточно возвести неравенство под вероятностной мерой в квадрат и воспользоваться обычным неравенством Чебышёва:

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geqslant t) = \mathbb{P}(|\xi - \mathbb{E}\xi|^2 \geqslant t^2) \leqslant \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^2}{t^2} = \frac{\mathbb{D}\xi}{t^2}.$$

[:|||:]

Также нам понадобится сходимость по вероятности.

Def. *Последовательность случайных величин* $\{\xi_n\}$ *сходится к случайной величине* ξ *по вероятности, если для всех* $\delta > 0$ *справедливо равенство* $\lim\limits_{n\to\infty} \mathbb{P}(|\xi_n - \xi| \geqslant \delta) = 0$.

Обозначение: $\xi_n \xrightarrow{\mathbb{P}} \xi$.

Закон больших чисел

Th. *Пусть* $\{\xi_n\}$ *— последовательность независимых в совокупности и одинаково распределённых случайных величин с конечным ожиданием* $\mu = \mathbb{E}\xi_n < \infty$ *и конечной дисперсией* $\sigma^2 = \mathbb{D}\xi_n < \infty$. *Тогда для любого* $\delta > 0$ *справедливо равенство*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \mu\right| \geqslant \delta\right) = 0$$

Доказательство с помощью неравенства Чебышёва

Доказательство. **(лекция)**
Сразу воспользуемся усиленным неравенством Чебышёва и независимостью:

$$\mathbb{P}\left(\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \mu\right| \geqslant \delta\right) \leqslant \frac{\mathbb{E}\left(\frac{\xi_1+\cdots+\xi_n}{n} - \mu\right)^2}{\delta^2} = \frac{\mathbb{D}\left(\frac{\xi_1+\cdots+\xi_n}{n}\right)}{\delta^2} = \frac{n \cdot \mathbb{D}\xi_k}{n^2 \cdot \delta^2} = \frac{\sigma^2}{n\delta^2} \to 0.$$

[:|||:]

Вариация для схемы Бернулли

(лекция) Рассмотрим последовательность независимых случайных величин $\{\xi_n\}$ такую, что

$$\xi_n = \begin{cases} 1, & \text{с вероятностью } p, \\ 0, & \text{с вероятностью } q = 1 - p. \end{cases}$$

Иными словами, это самая обычная схема Бернулли, в которой чаще всего интересным объектом выступает количество успехов $S_n = \xi_1 + \dots + \xi_n$. Мы знаем, что $\mathbb{E}\xi = p$, а $\mathbb{D}\xi = pq$. Тогда, используя только что доказанный закон больших чисел, получаем, что для любого $\delta > 0$ справедливо неравенство

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \delta\right) \leqslant \frac{pq}{n\delta^2} \to 0.$$

Видно, что скорость стремления к нулю слишком маленькая — всего порядка $\dfrac{1}{n}$.
С более точной оценкой нам поможет неравенство Чернова.

Неравенство Чернова

Th. (лекция) *Для схемы Бернулли существует следующая оценка сходимости по вероятности:*

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \delta\right) \leqslant 2e^{-2n\delta^2}, \text{ где } \delta > 0.$$

Мы докажем только один из случаев раскрытия модуля — тот, при котором $\dfrac{S_n}{n} - p \geqslant \delta$.
Перед тем, как переходить к непосредственному доказательству, приведём план действий:

1. Преобразуем неравенство к виду $\mathbb{P}\left(e^{\lambda S_n} \geqslant e^{\lambda n(p+\delta)}\right) \leqslant (g(\lambda))^n$.
2. Найдём минимум функции $g(\lambda)$.
3. Оценим функцию с помощью разложения по Тейлору.

Всё готово, можем начинать.

Доказательство. Сначала домножим всё неравенство на n:

$$\mathbb{P}\left(\frac{S_n}{n} - p \geqslant \delta\right) = \mathbb{P}(S_n \geqslant n(p + \delta)).$$

Далее, возьмём $\lambda > 0$ и применим "хорошую" функцию (а именно — экспоненту) к каждой из частей неравенства:

$$\mathbb{P}(S_n \geqslant n(p + \delta)) = \mathbb{P}\left(e^{\lambda S_n} \geqslant e^{\lambda n(p+\delta)}\right).$$

Так можно сделать ввиду того, что экспонента при $\lambda > 0$ является строго возрастающей монотонной функцией. Теперь применим неравенство Чебышёва:

$$\mathbb{P}\left(e^{\lambda S_n} \geqslant e^{\lambda n(p+\delta)}\right) \leqslant e^{-\lambda n(p+\delta)} \cdot \mathbb{E}e^{\lambda S_n}.$$

Теперь вспомним две полезных вещи: $\{\xi_k\}$ независимы и имеют одинаковое распределение. Воспользуемся этим:

$$e^{-\lambda n(p+\delta)} \cdot \mathbb{E}e^{\lambda S_n} = e^{-\lambda n(p+\delta)} \cdot \prod_{k=1}^{n} \mathbb{E}e^{\lambda \xi_k} = e^{-\lambda n(p+\delta)} \cdot \left(\mathbb{E}e^{\lambda \xi_k}\right)^n.$$

Посчитать подобное ожидание не составляет труда:

$$\mathbb{E}e^{\lambda \xi_k} = e^{\lambda \cdot 1} \cdot \mathbb{P}(\xi_k = 1) + e^{\lambda \cdot 0} \cdot \mathbb{P}(\xi_k = 0) = e^{\lambda} \cdot p + q.$$

Посмотрим на промежуточный результат:

$$\mathbb{P}\left(\frac{S_n}{n} - p \geqslant \delta\right) \leqslant \left(e^{-\lambda(p+\delta)} \cdot (e^{\lambda} \cdot p + q)\right)^n.$$

Теперь мы ищем такую λ, что это будет точкой минимума. Введём следующее обозначение:

$$g(\lambda) = e^{-\lambda(p+\delta)} \cdot (e^{\lambda} \cdot p + q).$$

Найдём производную и приравняем её к нулю:

$$\begin{aligned} g'(\lambda_{\min}) = -(p+\delta)e^{-\lambda_{\min}(p+\delta)} \cdot (e^{\lambda_{\min}} \cdot p + q) + e^{-\lambda_{\min}(p+\delta)} \cdot (e^{\lambda_{\min}} \cdot p) = 0, \\ -(p+\delta)(e^{\lambda_{\min}} \cdot p + q) + e^{\lambda_{\min}} \cdot p = 0. \end{aligned}$$

Выразим $e^{\lambda_{\min}}$:

$$e^{\lambda_{\min}} = \frac{q(p+\delta)}{p(q-\delta)} \implies \lambda_{\min} = \ln\left(\frac{q(p+\delta)}{p(q-\delta)}\right).$$

Нашли, можно подставлять. Только для начала разберёмся с правой скобкой:

$$(e^{\lambda_{\min}} \cdot p + q) = \frac{q(p+\delta)}{(q-\delta)} + q = q \cdot \frac{p+\delta+q-\delta}{q-\delta} = \frac{q}{q-\delta}.$$

Подставляем:

$$g(\lambda_{\min}) = e^{-\lambda_{\min}(p+\delta)} \cdot (e^{\lambda_{\min}} \cdot p + q) = e^{-(p+\delta)\cdot\ln\left(\frac{q(p+\delta)}{p(q-\delta)}\right)} \cdot \frac{q}{q-\delta} = e^{-(p+\delta)\cdot\ln\left(\frac{q(p+\delta)}{p(q-\delta)}\right) + \ln\left(\frac{q}{q-\delta}\right)}$$

Рассмотрим показатель как отдельную функцию и приведём подобные слагаемые:

$$H(x) = -(p+x)\ln\left(\frac{q(p+x)}{p(q-x)}\right) + \ln\left(\frac{q}{q-x}\right) = (p+x)\ln\frac{p}{p+x} + (q-x)\ln\frac{q}{q-x}.$$

Заметим, что $H(0) = p \cdot \ln 1 + q \cdot \ln 1 = 0$. Посмотрим на первую производную:

$$\begin{aligned} &H'(x) = \ln\frac{p}{p+x} + (p+x)\cdot\left(-\frac{1}{p+x}\right) - \ln\frac{q}{q-x} + (q-x)\cdot\left(\frac{1}{q-x}\right) = \ln\frac{p}{p+x} - \ln\frac{q}{q-x}, \\ &H'(0) = \ln 1 + \ln 1 = 0. \end{aligned}$$

И на вторую производную:

$$H''(x) = -\frac{1}{p+x} - \frac{1}{q-x} = \frac{-1}{(p+x)(q-x)}.$$

Далее, необходимо решить такую задачу: $a+b=1$, $a \geqslant 0, b \geqslant 0$, какой константой можно ограничить произведение ab? Решение:

$$a+b=1 \implies b = 1-a \implies ab = -a^2 + a \implies a_{\max} = \frac{1}{2} \implies b_{\max} = \frac{1}{2} \implies a_{\max} b_{\max} = \frac{1}{4}.$$

Следовательно, $H''(x) \leqslant -4$. Теперь разложим функцию в ряд Тейлора с остаточным членом в форме Лагранжа в окрестности точки $x_0 = 0$:

$$H(x-0) = H(0) + H'(0) \cdot (x-0) + \frac{H''(c) \cdot (x-0)^2}{2} = 0 + 0 + \frac{H''(c) \cdot x^2}{2}.$$

Используя только что полученную оценку, получим, что

$$H(x-0) = H(x) = \frac{H''(c) \cdot x^2}{2} \leqslant \frac{1}{2} \cdot x^2 \cdot (-4) = -2x^2 \implies H(x) \leqslant -2x^2.$$

Таким образом,

$$\mathbb{P}\left(\frac{S_n}{n} - p \geqslant \delta\right) \leqslant \left(e^{H(\delta)}\right)^n \leqslant e^{-2n\delta^2}.$$

[:|||:]

Комментарий

Двойка перед экспонентой в исходном неравенстве возникает из-за второго случая раскрытия модуля — того, при котором $\frac{S_n}{n} - p \leqslant -\delta$. Там всё более-менее аналогично.

(лекция) В момент нахождения $\lambda_{\min}$ мы делаем предположение, что $0 < \delta < q$. Следовательно, нам необходимо рассмотреть два случая:

$\delta > q$: $\mathbb{P}\left(\frac{S_n}{n} - p \geqslant \delta\right) = \mathbb{P}(S_n \geqslant n(p+\delta)) \leqslant \mathbb{P}(S_n > n) = 0$, так как $n(p+\delta) > n(p+q) = n$.

$\delta = q$: $\mathbb{P}(S_n \geqslant n(p+\delta)) = \mathbb{P}(S_n = n) = p^n \leqslant e^{-2nq^2}$ — необходимо это проверить:

$$p = (1-q) \leqslant e^{-2q^2} \quad \Leftrightarrow \quad \mu(q) = \ln(1-q) + 2q^2 \leqslant 0$$

Рассмотрим производную:

$$\mu(q)' = (\ln(1-q) + 2q^2)' = -\frac{1}{1-q} + 4q = \frac{-1+4q-4q^2}{1-q} = \frac{-(2q-1)^2}{1-q} \leqslant 0 \ \ \forall q \in [0;1)$$

$\mu(0) = 0, \ \ \mu'(q) \leqslant 0 \implies \mu(q)$ будет убывать при всех $0 \leqslant q \leqslant 1$.

Следовательно, неравенство верно.

(лекция) Также, надо пояснить, почему $\lambda_{\min}$ действительно является точкой минимума:

1. Мы рассматриваем функцию $g(\lambda) = e^{-\lambda(p+\delta)} \cdot \mathbb{E}e^{\lambda\xi_k} = \mathbb{E}e^{\lambda(\xi_k - p - \delta)}$.
2. Экспонента является выпуклой функцией.
3. Ожидание от выпуклой функции является выпуклой функцией.
4. У выпуклой функции может быть либо отрезок, состоящий из точек минимума, либо только одна точка минимума — это именно наш случай.

Билет №2.

Вступление и сходимость почти наверное

Для начала вспомним определение сходимости почти наверное.

Def. *Последовательность случайных величин* $\{\xi_n\}$ *сходится к случайной величине* ξ *почти наверное, если справедливо равенство*

$$\mathbb{P}\left(\lim_{n\to\infty} \xi_n = \xi\right) = 1.$$

Обозначение: $\xi_n \xrightarrow{a.s.} \xi$ *(almost surely).*

Теперь докажем вспомогательный факт.

Th. (лекция) *Если последовательность случайных величин* $\{\xi_n\}$ *сходится по вероятности к случайным величинам* ξ *и* η, *то* $\xi = \eta$ *почти наверное.*

Доказательство. Для всех $\delta > 0$ рассмотрим следующее событие: $\{|\xi - \eta| \geqslant \delta\}$.
Заметим, что оно лежит в таком объединении:

$$\{|\xi - \eta| \geqslant \delta\} \subset \left\{|\xi_n - \xi| \geqslant \frac{\delta}{2}\right\} \cup \left\{|\xi_n - \eta| \geqslant \frac{\delta}{2}\right\}.$$

Действительно, если какой-либо исход не попадает ни в одно из этих множеств, то для него справедливы неравенства

$$|\xi_n - \xi| < \frac{\delta}{2} \quad \text{и} \quad |\xi_n - \eta| < \frac{\delta}{2} \implies |\xi - \eta| < \delta, \text{ в силу неравенства треугольника.}$$

Теперь сделаем следующую оценку:

$$0 \leqslant \mathbb{P}(|\xi - \eta| \geqslant \delta) \leqslant \mathbb{P}\left(|\xi_n - \xi| \geqslant \frac{\delta}{2}\right) + \mathbb{P}\left(|\xi_n - \eta| \geqslant \frac{\delta}{2}\right).$$

Важно понимать, что центральная часть неравенства является числом, в то время как элементы последовательности есть только в правой части. Применяя лемму о двух хранителях правопорядка, получим, что

$$\mathbb{P}(|\xi - \eta| \geqslant \delta) = 0, \text{ так как, по условию, } \mathbb{P}\left(|\xi_n - \xi| \geqslant \frac{\delta}{2}\right) \to 0 \quad \text{и} \quad \mathbb{P}\left(|\xi_n - \eta| \geqslant \frac{\delta}{2}\right) \to 0.$$

Заметим, что

$$\mathbb{P}(\xi \neq \eta) = \mathbb{P}\left(\bigcup_n \left\{|\xi - \eta| \geqslant \frac{1}{n}\right\}\right) = 0 \implies \mathbb{P}(\xi = \eta) = 1.$$

[:|||:]

Сохранение сходимости по вероятности при подстановке случайных величин в непрерывную функцию

Th. (лекция) *Пусть дана непрерывная функция $g \in C(\mathbb{R})$, а последовательность случайных величин $\{\xi_n\}$ сходится к ξ по вероятности. Тогда $\{g(\xi_n)\}$ сходится к $g(\xi)$ по вероятности. Более того, это также справедливо для многомерного случая.*

Идея доказательства заключается в том, что мы используем непрерывность функции по определению, откуда можно будет оценить сверху искомую вероятность (различие значений функции при подстановке) уже известной (различие случайных величин), которая стремится к нулю.

Доказательство. Распишем то, что от нас требуется доказать, по определению:

$$\{g(\xi_n)\} \xrightarrow{\mathbb{P}} g(\xi) \quad \Leftrightarrow \quad \mathbb{P}\left(|g(\xi_n) - g(\xi)| \geqslant \delta\right) \to 0 \ \ \forall \delta > 0.$$

Пусть $t > 0$, тогда (в силу того, что $g \in C[-t;t]$) g будет равномерно непрерывной на $[-t;t]$. Иными словами,

$$\forall \delta > 0 \ \ \exists \hat{\delta} > 0\colon |x-y| < \hat{\delta} \implies |g(x) - g(y)| < \delta \text{ на } [-t;t].$$

Теперь рассмотрим интересующее нас событие: $\{|g(\xi_n) - g(\xi)| \geqslant \delta\}$.
Заметим, что оно лежит в следующем объединении:

$$\{|g(\xi_n) - g(\xi)| \geqslant \delta\} \subset \{|\xi_n| > t\} \cup \{|\xi| > t\} \cup \{|\xi_n - \xi| \geqslant \hat{\delta}\}.$$

Действительно, если исход не принадлежит первым двум множества, то ξ_n и ξ попадают на отрезок $[-t;t]$, на котором g равномерно непрерывна. Если для этого исхода справедливо, что существует такое $\hat{\delta} > 0$, что $|\xi_n - \xi| < \hat{\delta}$, то $|g(\xi_n) - g(\xi)| < \delta$. Следовательно, подобный исход не попадёт в интересующее нас событие. Тогда для него должно выполняться неравенство $|\xi_n - \xi| \geqslant \hat{\delta}$.

Теперь нужно немного модифицировать это объединение:

$$\{|g(\xi_n) - g(\xi)| \geqslant \delta\} \subset \left\{|\xi_n - \xi| \geqslant \frac{t}{2}\right\} \cup \left\{|\xi| \geqslant \frac{t}{2}\right\} \cup \{|\xi_n - \xi| \geqslant \hat{\delta}\}.$$

Это верно потому, что

1. $\{|\xi| > t\} \subset \left\{|\xi| \geqslant \frac{t}{2}\right\}$ (совсем очевидно),
2. $\{|\xi_n| > t\} \subset \left\{|\xi_n - \xi| \geqslant \frac{t}{2}\right\} \cup \left\{|\xi| \geqslant \frac{t}{2}\right\}$ в силу неравенства треугольника.

Теперь можно навесить знак вероятности слева и справа:

$$\mathbb{P}\left(|g(\xi_n) - g(\xi)| \geqslant \delta\right) \leqslant \mathbb{P}\left(|\xi_n - \xi| \geqslant \frac{t}{2}\right) + \mathbb{P}\left(|\xi| \geqslant \frac{t}{2}\right) + \mathbb{P}\left(|\xi_n - \xi| \geqslant \hat{\delta}\right).$$

Видно, что при увеличении t события $\left\{|\xi| \geqslant \frac{t}{2}\right\}$ будут вкладываться друг в друга, стремясь к пустому. Воспользовавшись непрерывностью меры, мы можем выбрать t так, что $\mathbb{P}\left(|\xi| \geqslant \frac{t}{2}\right) < \varepsilon$. Далее, зафиксируем это t и выберем $\hat{\delta}$, исходя из определения равномерной непрерывности. Теперь, используя тот факт, что $\xi_n \xrightarrow{\mathbb{P}} \xi$, получим, что

$$\forall \varepsilon > 0 \ \ \exists N\colon \ \ \forall n > N \quad \mathbb{P}\left(|\xi_n - \xi| \geqslant \frac{t}{2}\right) < \varepsilon \ \text{ и } \ \mathbb{P}\left(|\xi_n - \xi| \geqslant \hat{\delta}\right) < \varepsilon$$

По итогу:

$$\mathbb{P}\left(|g(\xi_n) - g(\xi)| \geqslant \delta\right) < \varepsilon + \varepsilon + \varepsilon = 3\varepsilon \implies \mathbb{P}\left(|g(\xi_n) - g(\xi)| \geqslant \delta\right) \to 0 \ \ \forall \delta > 0.$$

[:|||:]

Из этой теоремы вытекает пара арифметических операций.

Cor. (лекция) *Если* $\xi_n \xrightarrow{\mathbb{P}} \xi$ *и* $\eta_n \xrightarrow{\mathbb{P}} \eta$, *то* $\xi_n + \eta_n \xrightarrow{\mathbb{P}} \xi + \eta$ *и* $\xi_n \cdot \eta_n \xrightarrow{\mathbb{P}} \xi \cdot \eta$.

Теперь несколько определений.

Def. (лекция) *Пусть* $\xi_1, \ldots, \xi_n$ *— независимые одинаково распределённые случайные величины с конечным ожиданием* $\mu = \mathbb{E}\xi_k < \infty$ *и конечной дисперсией* $\sigma^2 = \mathbb{D}\xi_k < \infty$. *Тогда* $\xi_1, \ldots, \xi_n$ *называются простой выборкой.*

Def. *Пусть* $\xi_1, \ldots, \xi_n$ *— простая выборка, тогда* $\overline{\xi} = \dfrac{\xi_1 + \cdots + \xi_n}{n}$ *— выборочное среднее.*

Def. *Пусть* $\xi_1, \ldots, \xi_n$ *— простая выборка, тогда* $S^2 = \dfrac{1}{n-1} \sum\limits_{k=1}^{n} \left(\xi_k - \overline{\xi}\right)^2$ *— выборочная дисперсия.*

Сходимость по вероятности выборочного среднего

Утверждается, что выборочнее среднее сходится к ожиданию по вероятности. Ровно этот же факт был доказан в законе больших чисел.

Сходимость по вероятности выборочной дисперсии

Th. (лекция) *Выборочная дисперсия* S^2 *сходится к* $\sigma^2 = \mathbb{D}\xi_k$ *по вероятности (если* $\mathbb{E}\xi_k^4 < \infty$*).*

Доказательство. Распишем выборочную дисперсию чуть подробнее:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(\xi_k - \overline{\xi}\right)^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(\xi_k^2 - 2\xi_k \cdot \overline{\xi} + \overline{\xi}^2\right)$$

$$\sum_{k=1}^{n} 2\xi_k \cdot \overline{\xi} = 2\overline{\xi} \cdot (\xi_1 + \cdots + \xi_n) = 2\overline{\xi} \cdot n \cdot \frac{\xi_1 + \cdots + \xi_n}{n} = 2n\overline{\xi}^2$$

$$S^2 = \frac{1}{n-1} \left(\left(\sum_{k=1}^{n} \xi_k^2 \right) - 2n\overline{\xi}^2 + n\overline{\xi}^2 \right)$$

$$= \frac{n}{n-1} \cdot \frac{\xi_1^2 + \cdots + \xi_n^2}{n} - \frac{n}{n-1} \cdot \overline{\xi}^2$$

Заметим, что $\xi_1^2 + \cdots + \xi_n^2$ — простая выборка, откуда следует, что $\dfrac{\xi_1^2 + \cdots + \xi_n^2}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_k^2$ в силу закона больших чисел. Теперь, используя теорему о сохранении сходимости по вероятности при подстановке случайных величин в непрерывную функцию, можем заключить, что $\overline{\xi}^2 \xrightarrow{\mathbb{P}} (\mathbb{E}\xi_k)^2$. Опять же используя эту теорему, получим, что $S^2 \xrightarrow{\mathbb{P}} \mathbb{E}\xi_k^2 - (\mathbb{E}\xi_k)^2 = \sigma^2$, так как $\frac{n}{n-1}$ сходится к 1.

[:|||:]

Сходимость в среднем

Def. (лекция) *Последовательность случайных величин* $\{\xi_n\}$ *сходится к случайной величине* ξ *в среднем, если справедливо равенство*

$$\lim_{n \to \infty} \mathbb{E}|\xi_n - \xi| = 0.$$

Обозначение: $\xi_n \xrightarrow{L_1} \xi$.

Теорема Лебега

Теорема идёт в курсе без доказательства.

Th. (лекция) *Пусть* $\xi_n \xrightarrow{\mathbb{P}} \xi$ *и* $|\xi_n| \leqslant \eta$, *причём* $\mathbb{E}\eta < \infty$, *тогда* $\mathbb{E}|\xi_n - \xi| \to 0$, *или же* $\xi_n \xrightarrow{L_1} \xi$.

Усиленный закон больших чисел

Альтернативное название — закон больших чисел в форме Колмогорова.
Тоже без доказательства.

Th. (лекция) *Пусть* $\xi_1, \ldots, \xi_n$ *— независимые одинаково распределённые случайные величины с конечным ожиданием* $\mu = \mathbb{E}\xi_k < \infty$, *тогда* $\dfrac{\xi_1 + \cdots + \xi_n}{n} \xrightarrow{a.s} \mu$.

Билет №3.

Сходимость по распределению

Def. *Последовательность случайных величин $\{\xi_n\}$ сходится к ξ по распределению, если для всех точек непрерывности t функции распределения $F_\xi(t)$ справедливо равенство*

$$\lim_{n\to\infty} F_{\xi_n}(t) = F_\xi(t).$$

Обозначение: $\xi_n \xrightarrow{d} \xi$ (distribution).

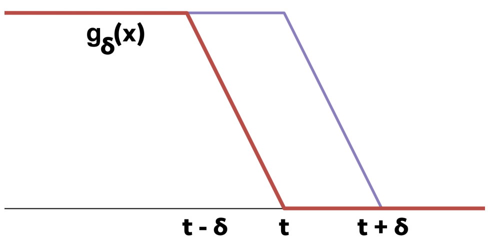
Эквивалентные определения

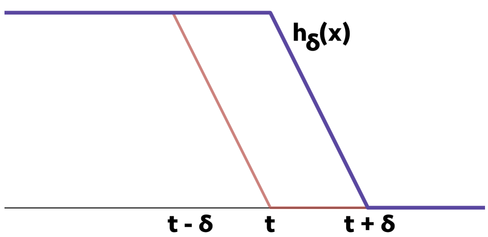
Th. (лекция) $\xi_n \xrightarrow{d} \xi$ *тогда и только тогда, когда для любой непрерывной и ограниченной функции $f \in C_B(\mathbb{R})$ справедливо равенство*

$$\lim_{n\to\infty} \mathbb{E} f(\xi_n) = \mathbb{E} f(\xi).$$

Доказательство. Сначала докажем в обратную сторону. Пусть t — точка непрерывности $F_\xi(t)$. Заметим, что $F_\xi(t) = \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi)$. Тогда для любого $\delta > 0$ определим следующие функции:

$$g_\delta = \begin{cases} 1, & x < t-\delta, \\ \dfrac{t-x}{\delta}, & t-\delta \leqslant x \leqslant t, \\ 0, & x > t. \end{cases} \qquad h_\delta = \begin{cases} 1, & x < t, \\ \dfrac{t+\delta-x}{\delta}, & t \leqslant x \leqslant t+\delta, \\ 0, & x > t+\delta. \end{cases}$$





Далее нужно воспроизвести серию неравенств:

$$\mathbb{E} g_\delta(\xi_n) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \leqslant \mathbb{E} h_\delta(\xi_n)$$
$$F_\xi(t-\delta) \leqslant \mathbb{E} g_\delta(\xi) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi) \leqslant \mathbb{E} h_\delta(\xi) \leqslant F_\xi(t+\delta)$$

По условию, $\mathbb{E} g_\delta(\xi_n) \to \mathbb{E} g_\delta(\xi)$ и $\mathbb{E} h_\delta(\xi_n) \to \mathbb{E} h_\delta(\xi)$, откуда следует, что

$$\mathbb{E} g_\delta(\xi_n) \to \mathbb{E} g_\delta(\xi) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \leqslant \mathbb{E} h_\delta(\xi) \leftarrow \mathbb{E} h_\delta(\xi_n)$$
$$\mathbb{E} g_\delta(\xi) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \leqslant \mathbb{E} h_\delta(\xi)$$
$$F_\xi(t-\delta) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \leqslant F_\xi(t+\delta)$$

Так как t — точка непрерывности, то мы можем устремить δ к 0:

$$F_\xi(t-0) \leqslant \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \leqslant F_\xi(t+0) \implies \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \to \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi)$$

Осталось заметить, что последнее выражение эквивалентно $F_{\xi_n}(t) \to F_\xi(t)$.

Теперь пусть нам известно, что $\xi_n \xrightarrow{d} \xi$. Иными словами, для точки непрерывности t

$$F_{\xi_n}(t) \to F_\xi(t) \;\;\Leftrightarrow\;\; \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi_n) \to \mathbb{E}\,\mathrm{Ind}_{(-\infty;t]}(\xi)$$

Раз мы знаем ожидание лучей, то можем найти ожидание отрезков:

$$F_{\xi_n}(t) - F_{\xi_n}(s) \to F_\xi(t) - F_\xi(s) \;\;\Leftrightarrow\;\; \mathbb{E}\,\mathrm{Ind}_{(s;t]}(\xi_n) \to \mathbb{E}\,\mathrm{Ind}_{(s;t]}(\xi)$$

Теперь мы можем найти их линейные комбинации:

$$\sum_j c_j \cdot \mathbb{E}\,\mathrm{Ind}_{(s_j;t_j]}(\xi_n) \to \sum_j c_j \cdot \mathbb{E}\,\mathrm{Ind}_{(s_j;t_j]}(\xi)$$

Воспользуемся линейностью ожидания:

$$\mathbb{E}\sum_j c_j\,\mathrm{Ind}_{(s_j;t_j]}(\xi_n) \to \mathbb{E}\sum_j c_j\,\mathrm{Ind}_{(s_j;t_j]}(\xi)$$

Под суммами находятся приближения функций с помощью константных отрезков. Иными словами, на каждом отрезке $(s_j;t_j]$ выбирается значение c_j. Любую непрерывную и ограниченную функцию можно приблизить подобным образом сколь угодно точно. **[:|||:]**

Сохранение сходимости по распределению при подстановке случайных величин в непрерывную функцию

Th. (лекция) *Если* $\xi_n \xrightarrow{d} \xi$, *то для любой непрерывной функции* $g \in C(\mathbb{R})$ *последовательность* $\{g(\xi_n)\}$ *сходится к* $g(\xi)$ *по распределению, или же* $g(\xi_n) \xrightarrow{d} g(\xi)$.

Доказательство. Важно понимать, что равно как ξ_n — случайные величины, так и $g(\xi_n)$ — случайные величины. Отметим, для любой непрерывной и ограниченной функции f композиция $h = f \circ g$ будет обладать такими же свойствами, тогда

$$\lim_{n\to\infty} \mathbb{E}f(g(\xi_n)) = \mathbb{E}f(g(\xi)) \Leftrightarrow \lim_{n\to\infty} \mathbb{E}h(\xi_n) = \mathbb{E}h(\xi)$$

[:|||:]

Сходимость двух последовательностей случайных величин, одна из которых сходится по распределению, а другая сходится по вероятности к константе

Переформулировка условия в принятых обозначениях: $\xi_n \xrightarrow{\mathbb{P}} c = \mathrm{const}$, а $\eta_n \xrightarrow{d} \eta$.

Их сумма

Th. $\xi_n + \eta_n \xrightarrow{d} c + \eta$.

Доказательство №1. Пусть $t \in \mathbb{R}$ и $\delta > 0$, тогда

$$\begin{aligned} F_{\xi_n+\eta_n}(t) &= \mathbb{P}(\xi_n + \eta_n \leqslant t) \\ &\leqslant \mathbb{P}\left(\{\xi_n + \eta_n \leqslant t\} \cap \{|\xi_n - c| < \delta\}\right) + \mathbb{P}(|\xi_n - c| \geqslant \delta) \end{aligned}$$

Последнее неравенство верно в силу того, что всё множество исходов разбивается на два: $|\xi_n - c| < \delta$ и $|\xi_n - c| \geqslant \delta$. В самом неравенстве мы не пересекаем последнее событие с тем, которое нас интересует, поэтому вероятность будет не меньше.

Неравенство $|\xi_n - c| < \delta$ равносильно $\xi_n \in (c-\delta; c+\delta)$. Заметим, что $t-c+\delta > t-c-\delta$, тогда

$$\mathbb{P}\left(\{\xi_n + \eta_n \leqslant t\} \cap \{|\xi_n - c| < \delta\}\right) + \mathbb{P}(|\xi_n - c| \geqslant \delta) \leqslant \mathbb{P}(\eta_n \leqslant t - c + \delta) + \mathbb{P}(|\xi_n - c| \geqslant \delta)$$

Аналогичным способом можно ограничить функцию снизу:

$$\begin{aligned} 1 - F_{\xi_n+\eta_n}(t) &= \mathbb{P}(\xi_n + \eta_n > t) \\ &\leqslant \mathbb{P}\left(\{\xi_n + \eta_n > t\} \cap \{|\xi_n - c| < \delta\}\right) + \mathbb{P}(|\xi_n - c| \geqslant \delta) \\ &\leqslant \mathbb{P}(\eta_n > t - c - \delta) + \mathbb{P}(|\xi_n - c| \geqslant \delta) \\ &= 1 - \mathbb{P}(\eta_n \leqslant t - c - \delta) + \mathbb{P}(|\xi_n - c| \geqslant \delta) \\ F_{\xi_n+\eta_n}(t) &\geqslant \mathbb{P}(\eta_n \leqslant t - c - \delta) - \mathbb{P}(|\xi_n - c| \geqslant \delta) \end{aligned}$$

Вспомнив условие, можем понять, что $\mathbb{P}(|\xi_n - c| \geqslant \delta) \to 0$.

Если точки $t-c, t-c+\delta$ и $t-c-\delta$ являются точками непрерывности F_η, то

$$\mathbb{P}(\eta_n \leqslant t - c \pm \delta) = F_{\eta_n}(t - c \pm \delta) \to F_\eta(t - c \pm \delta) \text{ по условию, откуда следует, что}$$

$$F_\eta(t - c - \delta) \leqslant \liminf_{n\to\infty} F_{\xi_n+\eta_n}(t) \leqslant \limsup_{n\to\infty} F_{\xi_n+\eta_n}(t) \leqslant F_\eta(t - c + \delta).$$

Исходя из предположения о непрерывности функции распределения в этих точках, можем выбрать δ сколь угодно малым, тогда

$$\lim_{n\to\infty} F_{\xi_n+\eta_n}(t) = F_\eta(t-c) = \mathbb{P}(\eta \leqslant t - c) = \mathbb{P}(\eta + c \leqslant t) = F_{\eta+c}(t)$$

[:|||:]

Доказательство №2 (с лекции). Потребуем от f непрерывную и ограниченную производную: $f, f' \in C_B(\mathbb{R})$. Полагая $\xi'_n := \xi_n - c$, можем считать, что $\xi'_n \xrightarrow{\mathbb{P}} 0$. Далее мы проверим следующее утверждение:

$$\mathbb{E}f(\xi'_n + \eta_n) \to \mathbb{E}f(\eta)$$

Рассмотрим их разность:

$$\begin{aligned} \mathbb{E}f(\xi'_n + \eta_n) - \mathbb{E}f(\eta) &= \mathbb{E}f(\xi'_n + \eta_n)\left[-\mathbb{E}f(\eta_n) + \mathbb{E}f(\eta_n)\right] - \mathbb{E}f(\eta) \\ &= \left[\mathbb{E}f(\xi'_n + \eta_n) - \mathbb{E}f(\eta_n)\right] + \left[\mathbb{E}f(\eta_n) - \mathbb{E}f(\eta)\right] \end{aligned}$$

Последняя скобка стремится к нулю по условию, поэтому рассмотрим первую. Ввиду ограниченности f и её производной, а также в силу теоремы Лагранжа, справедливо неравенство

$$|f(u) - f(v)| \leqslant C \cdot \min\{|u - v|, 1\}$$

Модифицируем его под нашу ситуацию:

$$|\mathbb{E}(f(\xi'_n + \eta_n) - f(\eta_n))| \leqslant \mathbb{E}|f(\xi'_n + \eta_n) - f(\eta_n)| \leqslant \mathbb{E}\, C \cdot \min\{|\xi'_n + \eta_n - \eta_n|, 1\} \leqslant C \cdot \mathbb{E}\min\{|\xi'_n|, 1\}$$

Заметим, что $\min\{|\xi'_n|, 1\} \geqslant \delta$ тогда и только тогда, когда и $|\xi'_n| \geqslant \delta$ и $1 \geqslant \delta$, следовательно

$$\mathbb{P}(\min\{|\xi'_n|, 1\} \geqslant \delta) = \mathbb{P}(\{|\xi'_n| \geqslant \delta\} \cap \{1 \geqslant \delta\}) \leqslant \mathbb{P}(|\xi'_n| \geqslant \delta) \to 0 \text{ (по условию.)}$$

Важно понимать, что $\min\{|\xi'_n|, 1\} = \min\{|\xi'_n|, 1\} - 0$, тогда $\min\{|\xi'_n|, 1\} \xrightarrow{\mathbb{P}} 0$.
Тогда введём следующие обозначения: $X_n = \min\{|\xi'_n|, 1\}, X = 0, Y = 1,$ — следовательно,

$$X_n \xrightarrow{\mathbb{P}} X, \quad |X_n| \leqslant Y, \quad \mathbb{E}Y < \infty$$

Тогда по теореме Лебега $\mathbb{E}|X_n - X| \to 0 \iff \mathbb{E}\left|\min\{|\xi'_n|, 1\} - 0\right| \to 0 \iff \mathbb{E}\min\{|\xi'_n|, 1\} \to 0$. Последний предел означает, что $|\mathbb{E}(f(\xi'_n + \eta_n) - f(\eta_n))| \to 0$, тогда $\mathbb{E}f(\xi'_n + \eta_n) \to \mathbb{E}f(\eta)$. **[:|||:]**

Их произведение

Th. $\xi_n \cdot \eta_n \xrightarrow{d} c \cdot \eta$.

Доказательство. Добавим и вычтем $c \cdot \eta_n$:

$$\xi_n \cdot \eta_n = \xi_n \cdot \eta_n + [c \cdot \eta_n - c \cdot \eta_n] = \eta_n \cdot (\xi_n - c) + c \cdot \eta_n$$

Обозначим $\zeta_n = c \cdot \eta_n$. Заметим, что $\zeta_n \xrightarrow{d} c \cdot \eta$. Действительно, для $c > 0$

$$\begin{aligned} F_{\zeta_n}(t) = \mathbb{P}(\zeta_n \leqslant t) = \mathbb{P}(c \cdot \eta_n \leqslant t) = \mathbb{P}\left(\eta_n \leqslant \frac{t}{c}\right) = F_{\eta_n}\left(\frac{t}{c}\right) \\ \to F_\eta\left(\frac{t}{c}\right) = \mathbb{P}\left(\eta \leqslant \frac{t}{c}\right) = \mathbb{P}(c \cdot \eta \leqslant t) = F_{c \cdot \eta}(t) \end{aligned}$$

И для $c < 0$

$$\begin{aligned} F_{\zeta_n}(t) = \mathbb{P}(\zeta_n \leqslant t) = \mathbb{P}\left(c \cdot \eta_n \leqslant t\right) = \mathbb{P}\left(\eta_n \geqslant \frac{t}{c}\right) = 1 - F_{\eta_n}\left(\frac{t}{c}\right) \\ \to 1 - F_\eta\left(\frac{t}{c}\right) = \mathbb{P}\left(\eta \geqslant \frac{t}{c}\right) = \mathbb{P}(c \cdot \eta \leqslant t) = F_{c \cdot \eta}(t) \end{aligned}$$

Рассуждения выше справедливы для таких t, что $\dfrac{t}{c}$ — точка непрерывности F_η.
Обозначим $\zeta'_n = \eta_n \cdot (\xi_n - c)$. Если воспользоваться результатами для суммы, то нам достаточно показать, что $\zeta'_n \xrightarrow{\mathbb{P}} 0$, откуда сразу будет следовать, что $\zeta_n + \zeta'_n \xrightarrow{d} \zeta \Leftrightarrow \zeta_n + (\zeta'_n + c') \xrightarrow{d} \zeta + c'$.
Пусть $t \in \mathbb{R}_+$ и $\delta > 0$, тогда

$$\mathbb{P}\left(|\eta_n \cdot (\xi_n - c)| > \delta\right) \leqslant \mathbb{P}(|\eta_n| > t) + \mathbb{P}(t \cdot |\xi_n - c| > \delta)$$

Это верно, потому что иначе исход не попадает ни в первое событие, ни во второе, тогда

$$|\eta_n| \leqslant t \text{ и } t \cdot |\xi_n - c| \leqslant \delta \implies t \cdot |\eta_n \cdot (\xi_n - c)| \leqslant t \cdot \delta \implies |\eta_n \cdot (\xi_n - c)| \leqslant \delta \text{ — противоречие.}$$

Вспомним, что $\xi_n - c \xrightarrow{\mathbb{P}} 0$, тогда $\mathbb{P}(t \cdot |\xi_n - c| > \delta)$ стремится к 0 по условию. Оценим $\mathbb{P}(|\eta_n| > t)$:

$$\begin{aligned} \mathbb{P}(|\eta_n| > t) = 1 - \mathbb{P}(|\eta_n| \leqslant t) &= 1 - (\mathbb{P}(\eta_n \leqslant t) - \mathbb{P}(\eta_n \leqslant -t)) \\ &= 1 - (F_{\eta_n}(t) - F_{\eta_n}(-t)) \\ &\to 1 - (F_\eta(t) - F_\eta(-t)) \end{aligned}$$

Если t стремится к бесконечности, то $1 - (F_\eta(t) - F_\eta(-t))$ стремится к $1 - (1 - 0) = 0$.
Таким образом, выбирая для функции F_η достаточно большие точки непрерывности t, можем добиться того, чтобы $\mathbb{P}(|\eta_n| > t)$ было сколь угодно малым, тогда и $\mathbb{P}\left(|\eta_n \cdot (\xi_n - c)| > \delta\right)$ будет сколь угодно малым, что означает сходимость $\eta_n \cdot (\xi_n - c) = \zeta'_n$ к 0 по вероятности. **[:|||:]**

Билет №4.

Характеристическая функция и её свойства

Def. *Пусть* ξ — *случайная величина, тогда функция* $\varphi_\xi(y) = \mathbb{E}e^{i\xi y}$ *называется характеристической функцией.*

Рассмотрим свойства характеристической функции:

1. В нуле принимает значение, равное 1: $\varphi_\xi(0) = 1$.

Доказательство. $\varphi_\xi(0) = \mathbb{E}e^{i\xi\cdot 0} = \mathbb{E}1 = 1$. **[:|||:]**

2. По модулю не превосходит единицы: $|\varphi_\xi(y)| \leqslant 1$

Доказательство. $|\varphi_\xi(y)| = \left|\mathbb{E}e^{i\xi y}\right| = |\mathbb{E}(\cos(\xi y) + i\cdot\sin(\xi y)| \leqslant \mathbb{E}\left|\cos(\xi y) + i\cdot\sin(\xi y)\right|$.
Модуль комплексного числа $a+i\cdot b$ считается следующим образом: $|a+i\cdot b| = \sqrt{a^2+b^2}$, — тогда $\mathbb{E}\left|\cos(\xi y) + i\cdot\sin(\xi y)\right| = \mathbb{E}\sqrt{(\cos(\xi y))^2 + (\sin(\xi y))^2} = \mathbb{E}1 = 1$. **[:|||:]**

3. Если существует k-тый момент (то есть $\mathbb{E}|\xi|^k < \infty$), то $\varphi_\xi(y)$ дифференцируема k раз, при этом справедливо равенство $\varphi_\xi^{(k)}(0) = i^k\,\mathbb{E}\xi^k$.

Доказательство. Продифференцируем в общем виде:

$$\lim_{\Delta y\to 0}\frac{\varphi_\xi(y+\Delta y)-\varphi_\xi(y)}{\Delta y} = \lim_{\Delta y\to 0}\mathbb{E}\left[\frac{e^{i\xi(y+\Delta y)} - e^{i\xi y}}{\Delta y}\right] = \lim_{\Delta y\to 0}\mathbb{E}\left[e^{i\xi y}\cdot\left(\frac{e^{i\xi\Delta y}-1}{\Delta y}\right)\right]$$

Заметим, что

$$\left|e^{i\xi y}\cdot\left(\frac{e^{i\xi\Delta y}-1}{\Delta y}\right)\right| = \left|e^{i\xi y}\right|\cdot\left|\frac{\exp(i\xi\Delta y) - \exp(i\xi\cdot 0)}{\Delta y}\right| \leqslant 1\cdot C\cdot\frac{\min\{|i\xi\Delta y|, 1\}}{|\Delta y|}$$
$$= C\cdot\frac{\min\{|\xi|\cdot|\Delta y|, 1\}}{|\Delta y|} \leqslant C\cdot|\xi|$$

Последнее неравенство верно, потому что при $|\xi|\cdot|\Delta y| \geqslant 1$ имеем $\dfrac{1}{|\Delta y|} \leqslant |\xi|$, тогда

$$C\cdot\frac{\min\{|\xi|\cdot|\Delta y|, 1\}}{|\Delta y|} = C\cdot\frac{1}{|\Delta y|} \leqslant C\cdot|\xi|$$

В то время как при $|\xi|\cdot|\Delta y| < 1$ неравенство очевидно. Вспомним, что, по условию, $\mathbb{E}|\xi| < \infty$. Далее, по свойству второго замечательного предела,

$$\lim_{\Delta y\to 0}\frac{e^{i\xi\Delta y}-1}{\Delta y} = i\xi.$$

Следовательно, по теореме Лебега,

$$\lim_{\Delta y\to 0}\mathbb{E}\left[e^{i\xi y}\cdot\left(\frac{e^{i\xi\Delta y}-1}{\Delta y}\right)\right] = \mathbb{E}\left(i\xi\cdot e^{i\xi y}\right)$$

Обобщение на случаи $k > 1$ проводится аналогичным способом. Подставим 0:

$$\varphi'_\xi(0) = \mathbb{E}\left(i\xi\cdot e^{i\xi\cdot 0}\right) = i\cdot\mathbb{E}\xi$$

[:|||:]

4. Можно считать линейное преобразование: $\varphi_{A\xi+B}(y) = e^{iBy} \cdot \varphi_\xi(Ay)$.

Доказательство. Честно распишем:

$$\varphi_{A\xi+B}(y) = \mathbb{E}e^{i(A\xi+B)y} = \mathbb{E}\left[e^{i\xi(Ay)} \cdot e^{iBy}\right] = e^{iBy}\varphi_\xi(Ay)$$

[:|||:]

5. Пусть ξ и η — независимые случайные величины, тогда $\varphi_{\xi+\eta}(y) = \varphi_\xi(y) \cdot \varphi_\eta(y)$.

Доказательство. Точно так же,

$$\varphi_{\xi+\eta}(y) = \mathbb{E}e^{i(\xi+\eta)y} = \mathbb{E}\left[e^{i\xi y} \cdot e^{i\eta y}\right] = \left[\mathbb{E}e^{i\xi y}\right] \cdot \left[\mathbb{E}e^{i\eta y}\right] \text{ (в силу независимости) } = \varphi_\xi(y) \cdot \varphi_\eta(y)$$

[:|||:]

6. Распределения двух случайных величин ξ и η совпадают тогда и только тогда, когда $\varphi_\xi = \varphi_\eta$.

Доказательство. Заметим, что если нам известна $\varphi_\xi(y)$, то в силу того, что

$$\varphi_\xi(y) = \mathbb{E}e^{i\xi y} = \mathbb{E}\left[\cos(\xi y) + i \cdot \sin(\xi y)\right] = \mathbb{E}\cos(\xi y) + i \cdot \mathbb{E}\sin(\xi y),$$

нам известны $\mathbb{E}\cos(\xi y)$ и $\mathbb{E}\sin(\xi y)$ как вещественная и мнимая части комплексного числа $\varphi_\xi(y)$. Функции $\xi \mapsto \cos(\xi y)$ и $\xi \mapsto \sin(\xi y)$ непрерывны и ограничены, тогда мы можем применить теорему об эквивалентных определениях сходимости по распределению, чтобы получить требуемое. **[:|||:]**

7. $\xi_n \xrightarrow{d} \xi \Leftrightarrow \varphi_{\xi_n}(y) \to \varphi_\xi(y)\ \forall y$.

Доказательство. Следует напрямую из теоремы об эквивалентных определениях сходимости по распределению, где $\mathbb{E}f(\xi) = \mathbb{E}e^{i\xi y}$. **[:|||:]**

Характеристическая функция нормального распределения

Пусть дана случайная величина $\eta \sim \mathrm{N}(\mu; \sigma^2)$. Обозначим $\xi = \dfrac{\eta - \mu}{\sigma}$, тогда $\xi \sim \mathrm{N}(0; 1)$.

Th. *Для случайной величины* $\xi \sim \mathrm{N}(0;1)$ *справедливо равенство* $\varphi_\xi(y) = e^{-y^2/2}$.

Доказательство. Немного матанализа:

$$\begin{aligned}
\varphi_\xi(y) = \mathbb{E}e^{i\xi y} &= \int\limits_{-\infty}^{+\infty} e^{ixy} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \\
&= \int\limits_{-\infty}^{+\infty} (\cos(xy) + i \cdot \sin(xy)) \cdot \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \\
&= \left[\text{в силу нечётности } \sin(xy)\text{, интеграл с ним обнулится, тогда}\right] \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(xy) \cdot e^{-x^2/2}\, dx = I(y) \\
I'(y) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \sin(xy) \cdot \left[-xe^{-x^2/2}\right]\, dx
\end{aligned}$$

Часть в квадратных скобках равна $\left(e^{-x^2/2}\right)'$, тогда, используя интегрирование по частям, получим

$$I'(y) = \frac{\sin(xy)}{\sqrt{2\pi}} \cdot e^{-x^2/2}\Bigg|_{-\infty}^{+\infty} - \frac{y}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(xy) \cdot e^{-x^2/2}\, dx = -\frac{y}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(xy) \cdot e^{-x^2/2}\, dx = -y \cdot I(y)$$

Нам необходимо решить задачу Коши (дифференциальное уравнение с начальным условием):

$$\begin{cases} I'(y) = -y \cdot I(y), \\ I(0) = \varphi_\xi(0) = 1 \end{cases} \quad \Leftrightarrow \quad I(y) = e^{-y^2/2}$$

в силу теоремы о существовании и единственности решений дифференциального уравнения. **[:|||:]**

Центральная предельная теорема

Th. *Пусть $\xi_1, \ldots, \xi_n$ — простая выборка с $\mathbb{E}\xi_k = \mu$ и $\mathbb{D}\xi_k = \sigma^2$, тогда*

$$\frac{\xi_1 + \cdots + \xi_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \xi \sim \mathrm{N}(0;1)$$

Доказательство. Отнормируем наши величины: $\eta_k = \dfrac{\xi_k - \mu}{\sigma}$. Заметим, что $\mathbb{E}\eta_k = 0, \mathbb{D}\eta_k = 1$. Тогда нам нужно удостовериться в том, что

$$\frac{\eta_1 + \cdots + \eta_n}{\sqrt{n}} \xrightarrow{d} \xi \sim \mathrm{N}(0;1)$$

Сделаем это с помощью характеристических функций:

$$\begin{aligned} \varphi_{\frac{\eta_1 + \cdots + \eta_n}{\sqrt{n}}}(y) &= \varphi_{\eta_1}\left(\frac{y}{\sqrt{n}}\right) \ldots \varphi_{\eta_n}\left(\frac{y}{\sqrt{n}}\right) \text{ (в силу независимости)} \\ &= \left[\varphi_{\eta_1}\left(\frac{y}{\sqrt{n}}\right)\right]^n \text{ (в силу одинакового распределения)} \\ &= \left[1 - \frac{y^2}{2n} + o\left(\frac{1}{n}\right)\right]^n, \end{aligned}$$

в силу формулы для производных характеристических функций и разложения по Тейлору. Заметим, что

$$\lim_{n\to\infty} \left[1 - \frac{y^2}{2n} + o\left(\frac{1}{n}\right)\right]^n = e^{-y^2/2} \text{ как второй замечательный предел.}$$

Мы знаем, что $\varphi_{\mathrm{N}(0;1)}(y) = e^{-y^2/2}$, тогда

$$\varphi_{\frac{\eta_1 + \cdots + \eta_n}{\sqrt{n}}}(y) \to \varphi_{\mathrm{N}(0;1)}(y) \quad \Leftrightarrow \quad \frac{\eta_1 + \cdots + \eta_n}{\sqrt{n}} \xrightarrow{d} \xi \sim \mathrm{N}(0;1)$$

[:|||:]

Билет №5.

Вступление

Этот билет является сборкой информации, которую удалось найти. На самих лекциях билет явно не выделялся, а был разбит на несколько отдельных частей, которые пришлось собрать.

Асимптотическая нормальность выборочного среднего

Пусть нам дана простая выборка $\xi_1, \ldots, \xi_n$ с ожиданием $\mathbb{E}\xi_k = \mu$ и дисперсией $\mathbb{D}\xi_k = \sigma^2$. **(лекция)** Выборочое среднее $\overline{\xi} = \dfrac{\xi_1 + \cdots + \xi_n}{n} \xrightarrow[\text{ЗБЧ}]{\mathbb{P}} \mu$. Немного преобразований:

$$\mathrm{N}(0;1) \sim \xi \xleftarrow[\text{ЦПТ}]{d} \frac{\xi_1 + \cdots + \xi_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{\xi_1+\cdots+\xi_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma} \cdot (\overline{\xi} - \mu)$$

Нам интересно, с какой скоростью $\overline{\xi}$ стремится к μ. ЦПТ даёт наивную оценку порядка $\dfrac{1}{\sqrt{n}}$ (если домножить $\overline{\xi}$ на $\sqrt{n}$, то она будет сходиться к чему-то ненулевому, но и не будет уходить на бесконечность). Иными словами, $\overline{\xi}$ разбросана около μ в пределе $\pm\dfrac{1}{\sqrt{n}}$.

(лекция)
Для простоты будем считать, что $\mu = 0, \sigma = 1$.
У выборочного среднего следующие характеристики: $\mathbb{E}\overline{\xi} = 0, \mathbb{D}\overline{\xi} = \dfrac{1}{n} \to 0$.
Теперь рассмотрим другую случайную величину $\zeta = \dfrac{\xi_1 + \cdots + \xi_n}{\sqrt{n}} \implies \mathbb{E}\zeta = 0, \mathbb{D}\zeta = 1$.
Процесс нормировки и приведения случайных величин к виду ζ называется стандартизацией данных. Получается, что при росте количества данных ($n \to \infty$), график распределения ζ приближается к графику стандартного нормального распределения.

Теорема о непрерывности (не требуется)

Th. (лекция) *Пусть $\xi_n \xrightarrow{d} \xi$, a — фиксированное число, $\{b_n\}$ — числовая последовательность, которая стремится к 0, а f — непрерывная и дифференцируемая функция. Тогда*

$$\frac{f(a + b_n\xi_n) - f(a)}{b_n} \xrightarrow{d} f'(a) \cdot \xi$$

Доказательство. Из курса матанализа знаем, что

$$\frac{f(a + b_n\xi_n) - f(a)}{b_n} = \frac{(a + b_n\xi_n - a) \cdot \int\limits_0^1 f'(a + b_n\xi_n t)\,dt}{b_n} = \xi_n \cdot \int\limits_0^1 f'(a + b_n\xi_n t)\,dt$$

По теореме о сходимости двух последовательностей случайных величин, одна из которых сходится по распределению, а другая сходится по вероятности к константе, можем заключить, что $b_n\xi_n \xrightarrow{\mathbb{P}} 0$.

Используя теорему о сохранении сходимости по вероятности при подстановке случайной величины в функцию, получаем, что $f'(a + b_n\xi_n t) \xrightarrow{\mathbb{P}} f'(a)$.

Далее, можем заключить, что $\int\limits_0^1 f'(a+b_n\xi_n t)\,dt \xrightarrow{\mathbb{P}} f'(a)$. Для строгого доказательства этого факта потребуем, чтобы f была дважды дифференцируема, тогда

$$\int\limits_0^1 |f'(a+b_n\xi_n t)-f'(a)|\,dt \leqslant C\cdot|b_n\xi_n|$$

Так как $b_n\xi_n \xrightarrow{\mathbb{P}} 0$, то и разность интеграла и значения производной в точке стремится к нулю. Тогда, опять же используя теорему о сходимости двух последовательностей случайных величин, одна из которых сходится по распределению, а другая сходится по вероятности к константе, можем заключить, что $\xi_n\cdot\int\limits_0^1 f'(a+b_n\xi_n t)\,dt \to \xi\cdot f'(a)$. **[:|||:]**

Асимптотический доверительный интервал

(лекция) Пусть $\xi_1,\ldots,\xi_n$ — простая выборка, тогда знаем, что

$$\frac{\sqrt{n}}{\sigma}\cdot(\overline{\xi}-\mu)\xrightarrow{d}\xi\sim\mathrm{N}(0;1)$$

Def. *Пусть задано такое α, что $1-\alpha$ лежит около единицы, тогда интервал $(\mu-\delta;\mu+\delta)$ называется доверительным интервалом, если*

$$\mathbb{P}(\overline{\xi}-\delta<\mu<\overline{\xi}+\delta)=1-\alpha$$

Обычно хочется, чтобы $\alpha\simeq 0$ и $\delta\simeq 0$, но так бывает крайне редко. Но на практике мы мало что знаем про само распределение $\overline{\xi}$. Как в таком случае подбирать δ? Здесь нам поможет ЦПТ:

$$\mathbb{P}\left(-z<\frac{\sqrt{n}}{\sigma}\cdot(\overline{\xi}-\mu)<z\right)\to\int\limits_{-z}^{z}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx$$

Тогда при больших n мы можем выбирать такие z_α, что

$$\mathbb{P}\left(-z_\alpha<\frac{\sqrt{n}}{\sigma}\cdot(\overline{\xi}-\mu)<z_\alpha\right)\to 1-\alpha \quad\Leftrightarrow\quad \mathbb{P}\left(\overline{\xi}-\frac{z_\alpha\sigma}{\sqrt{n}}<\mu<\overline{\xi}+\frac{z_\alpha\sigma}{\sqrt{n}}\right)\simeq 1-\alpha$$

Чтобы уменьшать промежуток, надо увеличивать n, чтобы он превосходил z_α. На данный момент, наша вероятность стремится к $1-\alpha$ со скоростью порядка $\dfrac{1}{\sqrt{n}}$.

Но у нас есть ещё одна проблема: не всегда известна σ. Но на самом деле, не всё так плохо, потому что мы знаем о выборочной дисперсии: $S^2=\dfrac{1}{n-1}\cdot\sum\limits_{k=1}^{n}(\overline{\xi}-\xi_k)^2$. Используя все те теоремы, что мы доказали ранее, можем заключить, что

$$\frac{\sqrt{n}}{S}\cdot(\overline{\xi}-\mu)\xrightarrow{d}\xi\sim\mathrm{N}(0;1).$$

Это верно потому, что

$$\frac{\sqrt{n}}{S}\cdot(\overline{\xi}-\mu)=\frac{\sigma}{S}\cdot\frac{\sqrt{n}}{\sigma}\cdot(\overline{\xi}-\mu),\ \text{где}\ \frac{\sigma}{S}\xrightarrow{\mathbb{P}}1,\ \text{а}\ \frac{\sqrt{n}}{\sigma}\cdot(\overline{\xi}-\mu)\xrightarrow{d}\xi\sim\mathrm{N}(0;1).$$

Результат таков:

$$\mathbb{P}\left(\overline{\xi}-\frac{z_\alpha S}{\sqrt{n}}<\mu<\overline{\xi}+\frac{z_\alpha S}{\sqrt{n}}\right)\simeq 1-\alpha$$

Билет №6.

Вступление

(лекция) Рассматриваются только двумерные векторы, так как для более общего случая всё получается аналогично.

Def. *Случайным вектором называется вектор, составленный из случайных величин. Иными словами, если ξ, η — случайные величины, то (ξ, η) — случайный вектор.*

У случайного вектора есть распределение: $\mu_{(\xi,\eta)}(B) = \mathbb{P}((\xi, \eta) \in B)$.
В терминах функции распределения: $F_{(\xi,\eta)}(t, s) = \mathbb{P}(\xi \leqslant t; \eta \leqslant s)$.

Характеристическая функция случайного вектора

Def. (лекция) *Для случайного вектора $x = (\xi, \eta)$ характеристической функцией называется функция вида $\varphi_x(y) = \mathbb{E}e^{i\langle x,y\rangle}$, где $y = (y_1, y_2)$ — вектор чисел, а $\langle \cdot, \cdot \rangle$ — скалярное произведение.*

(лекция) Рассмотрим несколько свойств:

1. Векторы u и v имеют одинаковое распределение тогда и только тогда, когда $\varphi_u = \varphi_v$.
[Без доказательства.]

Проверка независимости компонент случайного вектора с помощью характеристической функции

2. Компоненты ξ и η случайного вектора $x = (\xi, \eta)$ независимы тогда и только тогда, когда $\varphi_x(y) = \varphi_\xi(y_1) \cdot \varphi_\eta(y_2)$.

Доказательство. **(лекция)** Пусть компоненты независимы, тогда

$$\varphi_x(y) = \mathbb{E}e^{i\langle x,y\rangle} = \mathbb{E}\left[e^{i\xi y_1} \cdot e^{i\eta y_2}\right] = \left[\mathbb{E}e^{i\xi y_1}\right] \cdot \left[\mathbb{E}e^{i\eta y_2}\right] \text{ (в силу независимости)} = \varphi_\xi(y_1) \cdot \varphi_\eta(y_2)$$

Теперь докажем в обратную сторону: пусть $\varphi_x(y) = \varphi_\xi(y_1) \cdot \varphi_\eta(y_2)$, тогда сконструируем такой новый вектор $\zeta = (\zeta_1,\zeta_2)$, что распределение ζ_1 совпадает с распределением ξ, а распределение ζ_2 совпадает с распределением η, причём ζ_1, ζ_2 независимы, тогда пусть функция распределения этого вектора будет задаваться так:

$$F_\zeta(t, s) = F_{\zeta_1}(t) \cdot F_{\zeta_2}(s)$$

Посчитаем его характеристическую функцию:

$$\begin{aligned}\varphi_\zeta(y) &= \varphi_{\zeta_1}(y_1) \cdot \varphi_{\zeta_2}(y_2) \text{ (в силу независимости компонент)}\\ &= \varphi_\xi(y_1) \cdot \varphi_\eta(y_2) \text{ (в силу одинаковых распределений)}\\ &= \varphi_x(y) \text{ (по условию)}\end{aligned}$$

По предыдущему свойству можем утверждать, что распределения этих векторов совпадают, откуда следует, что и ξ с η являются независимыми. **[:|||:]**

Изменение характеристической функции при афинном преобразовании случайного вектора

3. Можем делать афинные преобразования: $\varphi_{Ax+b}(y) = e^{i\langle b,y\rangle}\varphi_x\left(A^T y\right)$.

Доказательство. **(лекция)** Подробно распишем:

$$\varphi_{Ax+b}(y) = \mathbb{E}e^{i\langle Ax+b,y\rangle} = e^{i\langle b,y\rangle}\mathbb{E}e^{i\langle Ax,y\rangle} = e^{i\langle b,y\rangle}\mathbb{E}e^{i\langle x,A^T y\rangle} = e^{i\langle b,y\rangle}\varphi_x\left(A^T y\right)$$

[:|||:]

Билет №7.

Многомерное нормальное распределение

Def. (лекция) *Вектор x имеет нормальное распределение с параметрами μ и R, где μ — вектор, а R — симметричная неотрицательно определённая матрица ($\langle Ry, y\rangle \geqslant 0$), если справедливо равенство*

$$\varphi_x(y) = \exp\left(i\langle \mu, y\rangle - \frac{\langle Ry, y\rangle}{2}\right)$$

Поймём, что же из себя представляют вектор μ и матрица R. С одной стороны,

$$\varphi_x(y) = \mathbb{E}e^{i(\xi y_1 + \eta y_2)} \implies \frac{\partial \varphi_x}{\partial y_1}(0,0) = \left(\mathbb{E}\left[e^{i\eta y_2} \cdot i\xi \cdot e^{i\xi y_1}\right]\right)(0,0) = i \cdot \mathbb{E}\xi$$
$$\frac{\partial \varphi_x}{\partial y_2}(0,0) = \left(\mathbb{E}\left[e^{i\xi y_1} \cdot i\eta \cdot e^{i\eta y_2}\right]\right)(0,0) = i \cdot \mathbb{E}\eta$$
$$\frac{\partial^2 \varphi_x}{\partial y_1^2}(0,0) = \left(\mathbb{E}\left[-e^{i\eta y_2} \cdot \xi^2 \cdot e^{i\xi y_1}\right]\right)(0,0) = -\mathbb{E}\xi^2$$
$$\frac{\partial^2 \varphi_x}{\partial y_2^2}(0,0) = \left(\mathbb{E}\left[-e^{i\xi y_1} \cdot \eta^2 \cdot e^{i\eta y_2}\right]\right)(0,0) = -\mathbb{E}\eta^2$$
$$\frac{\partial^2 \varphi_x}{\partial y_1 \partial y_2}(0,0) = \left(\mathbb{E}\left[i\eta \cdot e^{i\eta y_2} \cdot i\xi \cdot e^{i\xi y_1}\right]\right)(0,0) = -\mathbb{E}(\xi \cdot \eta)$$

С другой стороны,

$$\varphi_x(y) = \exp\left(i \cdot (\mu_1 y_1 + \mu_2 y_2) - \frac{1}{2}(r_{11}y_1^2 + 2r_{12}y_1y_2 + r_{22}y_2^2)\right) = \exp(\dots)$$
$$\frac{\partial \varphi_x}{\partial y_1}(0,0) = \left[\exp(...) \cdot (i\mu_1 - r_{11}y_1 - r_{12}y_2)\right](0,0) = i\mu_1 \implies \mu_1 = \mathbb{E}\xi, \mu_2 = \mathbb{E}\eta \text{ (аналогично)}$$

Следовательно, μ — вектор ожиданий (вектор средних). Теперь разберёмся с R:

$$\frac{\partial^2 \varphi_x}{\partial y_1^2}(0,0) = \left[\exp(\dots) \cdot (i\mu_1 - r_{11}y_1 - r_{12}y_2)^2 - \exp(\dots) \cdot r_{11}\right](0,0) = -\mu_1^2 - r_{11}$$
$$\frac{\partial^2 \varphi_x}{\partial y_1 \partial y_2}(0,0) = \left[\exp(\dots) \cdot (i\mu_1 - r_{11}y_1 - r_{12}y_2) \cdot (i\mu_2 - r_{12}y_1 - r_{22}y_2) - \exp(\dots) \cdot r_{12}\right](0,0) = -\mu_1\mu_2 - r_{12}$$

В таком случае, получается, что

$$-\mathbb{E}\xi^2 = -\mu_1^2 - r_{11} \implies r_{11} = \mathbb{E}\xi^2 - \mu_1^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \mathbb{D}\xi \implies r_{22} = \mathbb{D}\eta$$
$$-\mathbb{E}(\xi \cdot \eta) = -\mu_1\mu_2 - r_{12} \implies r_{12} = \mathbb{E}(\xi \cdot \eta) - \mathbb{E}\xi \cdot \mathbb{E}\eta = \operatorname{cov}(\xi, \eta) \implies r_{21} = \operatorname{cov}(\eta, \xi)$$

Выходит, что $R = (r_{ij}) = (\operatorname{cov}(\xi_i, \xi_j))$.
В этом случае матрица R называется *ковариационной матрицей*. По итогу:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}\xi \\ \mathbb{E}\eta \end{pmatrix} \quad R = \begin{pmatrix} \operatorname{cov}(\xi,\xi) & \operatorname{cov}(\xi,\eta) \\ \operatorname{cov}(\eta,\xi) & \operatorname{cov}(\eta,\eta) \end{pmatrix} = \begin{pmatrix} \mathbb{D}\xi & \operatorname{cov}(\xi,\eta) \\ \operatorname{cov}(\eta,\xi) & \mathbb{D}\eta \end{pmatrix}$$

Th. (лекция) *Для любых случайных величин ковариационная матрица является неотрицательно определённой, то есть для любого вектора y (составленного из чисел, необязательно из случайных величин) справедливо неравенство*

$$\langle Ry, y\rangle \geqslant 0$$

Доказательство. Рассмотрим следующее выражение:

$$0 \leqslant \mathbb{D}(y_1\xi + y_2\eta) = \mathbb{D}(y_1\xi) + \mathbb{D}(y_2\eta) + 2\operatorname{cov}(y_1\xi, y_2\eta) = y_1^2 \cdot \mathbb{D}\xi + y_2^2 \cdot \mathbb{D}\eta + 2y_1y_2\operatorname{cov}(\xi,\eta)$$

Теперь внимательно посмотрим на скалярное произведение:

$$\langle Ry, y\rangle = \left\langle \begin{pmatrix} y_1 \cdot \mathbb{D}\xi + y_2\operatorname{cov}(\xi,\eta) \\ y_1\operatorname{cov}(\eta,\xi) + y_2 \cdot \mathbb{D}\eta \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = y_1^2 \cdot \mathbb{D}\xi + y_2^2 \cdot \mathbb{D}\eta + 2y_1y_2\operatorname{cov}(\xi,\eta) = \mathbb{D}(y_1\xi + y_2\eta) \geqslant 0$$

[:|||:]

Равенство нулю ковариации и независимость

(лекция) Важно заметить, что если случайный вектор x нормально распределён, то есть $x \sim N(\mu; R)$, то его компоненты нормально распределены. Это верно, потому что

$$\begin{aligned}
\varphi_x(t,0) &= \exp\left(i \cdot (\mu_1 t + \mu_2 \cdot 0) - \frac{1}{2}(r_{11}t^2 + 2r_{12}t \cdot 0 + r_{22} \cdot 0^2)\right) \\
&= \exp\left(it\mu_1 - \frac{t^2 r_{11}}{2}\right) \\
&= \exp\left(it\mu_1 - \frac{t^2\sigma_1^2}{2}\right) = \varphi_\xi(t), \text{ где } \xi \sim \mathrm{N}(\mu_1, \sigma_1^2). \\
\varphi_x(0,s) &= \exp\left(is\mu_2 - \frac{s^2\sigma_2^2}{2}\right) = \varphi_\eta(s), \text{ где } \eta \sim \mathrm{N}(\mu_2, \sigma_2^2).
\end{aligned}$$

(лекция) Пусть дан нормально распределённый случайный вектор $x \sim \mathrm{N}(\mu; R)$. Что будет с его характеристической функцией, если r_{12} станет равным 0? Рассмотрим это подробнее:

$$\begin{aligned}
\varphi_x(y) &= \exp\left(i(\mu_1 y_1 + \mu_2 y_2) - \frac{1}{2}\left(r_{11}y_1^2 + r_{22}y_2^2\right)\right) \\
&= \exp\left(i\mu_1 y_1 - \frac{r_{11}y_1^2}{2}\right) \cdot \exp\left(i\mu_2 y_2 - \frac{r_{22}y_2^2}{2}\right) \\
&= \varphi_\xi(y_1) \cdot \varphi_\eta(y_2)
\end{aligned}$$

Но по свойству характеристической функции вектора выходит, что ξ и η независимы.
Таким образом, мы задаром получаем следующую теорему.

Th. *Пусть дан нормально распределённый случайный вектор $x \sim \mathrm{N}(\mu; R)$. Тогда его компоненты независимы в том и только в том случае, когда их ковариация равна нулю.*

Доказательство. Пусть компоненты независимы, тогда очевидно.
Пусть ковариация равна нулю — эту ситуацию мы разобрали выше. **[:|||:]**

Распределение вектора, полученного линейным преобразованием из гауссовского

Th. (лекция) *Пусть* $x \sim \mathrm{N}(\mu; R)$, C *— матрица, а* b *— вектор.*
Тогда вектор $b + Cx \sim \mathrm{N}(b + C\mu, C \cdot R \cdot C^T)$.

Доказательство. Обозначим $v = b + Cx$, тогда

$$\begin{aligned}\varphi_v(y) &= e^{i\langle b,y\rangle} \cdot \varphi_x\left(C^T y\right) \text{ (в силу афинного преобразования)}\\ &= \exp\left(i\langle b, y\rangle + i\langle \mu, C^T y\rangle - \frac{1}{2}\langle R \cdot C^T y, C^T y\rangle\right)\end{aligned}$$

Это мы получили с помощью обычной формулы характеристической функции для нормального случайного вектора. Заметим, что $C^T y$ — это вектор, обозначим как u, тогда

$$\langle R \cdot C^T y, C^T y\rangle = \langle Ru, u\rangle \geqslant 0 \text{ (так как } R \text{ — неотрицательно определена)}$$

Далее,

$$\begin{aligned}\ldots &= \exp\left(i\langle b + C\mu, y\rangle - \frac{1}{2}\langle R \cdot C^T y, C^T y\rangle\right) \text{ (объединили два скалярных произведения)}\\ &= \exp\left(i\langle b + C\mu, y\rangle - \frac{1}{2}\langle C \cdot R \cdot C^T y, y\rangle\right) \text{ (перенесли матрицу)}\end{aligned}$$

Заметим, что $(C \cdot R \cdot C^T)^T = (C^T)^T \cdot R^T \cdot C^T = C \cdot R \cdot C^T$. Следовательно, эта матрица симметрична и неотрицательно определена, тогда

$$\ldots = \exp\left(i\langle b + C\mu, y\rangle - \frac{1}{2}\langle C \cdot R \cdot C^T y, y\rangle\right) = \varphi_v(y) = \varphi_{b+Cx}(y)$$

[:|||:]

Cor. *Пусть дан нормальный случайный вектор* $x = (\xi, \eta) \sim \mathrm{N}(\mu; R)$, *тогда линейная комбинация его компонент имеет нормальное распределение. Иными словами, для любых* b, c_1, c_2 *справдливо выражение*

$$b + c_1 \cdot \xi + c_2 \cdot \eta \sim \mathrm{N}(\alpha, \gamma^2)$$

Доказательство. Напрямую следует из предыдущей теоремы и свойства нормального случайного вектора о том, что его компоненты нормально распределены. **[:|||:]**

Билет №8.

Многомерное нормальное распределение

Пусть дан нормальный случайный вектор $x \sim \mathrm{N}(\mu; R)$. Очевидно, что в этом случае вектор $x - \mu \sim \mathrm{N}(0; R)$, так как вычитание вектора не влияет на матрицу ковариаций. Далее будем рассматривать только такие векторы.

Вспомним, что существует линейное пространство над случайными величинами с нулевым ожиданием:

$$L = \left\{\xi \mid \mathbb{E}\xi = 0,\ \ \mathbb{E}\xi^2 < \infty\right\}$$

Скалярное произведение на нём задано так:

$$\langle \xi, \eta \rangle = \mathbb{E}(\xi \cdot \eta)$$

Рассмотрим плоскость

$$\Pi = \{\lambda\xi + \beta\eta \mid (\xi, \eta) \sim \mathrm{N}(0; R),\ \ \alpha, \beta \in \mathbb{R}\}$$

Понятно, что для любых двух случайных величин $\xi, \eta \in \Pi$:

1. (ξ, η) является нормальным вектором;

2. $\xi \perp \eta \Leftrightarrow \langle \xi, \eta \rangle = 0 \Leftrightarrow \xi$ и η независимы.

Заметим, что норма определена так:

$$\|\xi\| = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\mathbb{D}\xi}$$

В этом случае

$$\|\xi\| = 1 \Leftrightarrow \xi \sim \mathrm{N}(0; 1)$$

Ортогонализация и получение гауссовского вектора линейным преобразованием из вектора, компоненты которого являются независимыми нормально распределёнными случайными величинами с параметрами 0 и 1

Th. *В плоскости* Π *существует ортнормированный базис* (e_1, e_2), *то есть для любых случайных величин* ξ, η *существуют такие матрица* C *и независимые случайные величины* $e_1, e_2 \sim \mathrm{N}(0; 1)$, *что*

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = C \cdot \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

Доказательство. Мы будем использовать метод ортогонализации Грама-Шмидта:

$$e_1 = \frac{\xi}{\|\xi\|} = \frac{1}{\sqrt{\mathbb{D}\xi}} \cdot \xi$$
$$e_2 = \frac{\eta - \langle \eta, e_1 \rangle \cdot e_1}{\|\eta - \langle \eta, e_1 \rangle \cdot e_1\|},\ \text{где } \langle \eta, e_1 \rangle = \frac{\mathbb{E}(\eta \cdot \xi)}{\sqrt{\mathbb{D}\xi}}$$

Ожидание у них равно нулю, так как мы рассматриваем только такие случайные величины. Видно, что каждый вектор имеет длину 1. Прямая проверка того, что $\langle e_1, e_2 \rangle = 0$, позволяет убедиться в их независимости. **[:|||:]**

Пусть e_1 и e_2 независимы и имеют распределение $\mathrm{N}(0;1)$, тогда $(e_1, e_2) \sim \mathrm{N}(0; E)$, где E — единичная матрица. Заметим, что

$$\varphi_{(e_1,e_2)}(y) = \exp\left(-\frac{\|y\|^2}{2}\right) = \exp\left(-\frac{y_1^2}{2}\right) \cdot \exp\left(-\frac{y_2^2}{2}\right)$$

Пусть $\begin{pmatrix} \xi \\ \eta \end{pmatrix} = C \cdot \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$, тогда

$$\varphi_{(\xi,\eta)}(y) = \exp\left(-\frac{1}{2}\langle C \cdot C^T y, y\rangle\right) \implies R = C \cdot C^T, \det R = (\det C)^2$$

Если матрица R уже задана, то в качестве C можно взять $\sqrt{R}$.

Cor. (лекция) *Для любой неотрицательно определённой и симметричной матрицы R и любого вектора μ существует случайный вектор $(\xi, \eta) \sim \mathrm{N}(\mu; R)$. Более того, всякий такой вектор имеет вид $\mu + C \cdot \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$, где $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim \mathrm{N}(0; E)$.*

Заметим, что в случае, когда C вырождена (имеет ранг < 2), её определитель равен нулю. Следовательно, матрица R тоже будет иметь нулевой определитель. Это означает, что вектор будет распределён либо в точке, либо на прямой. У обоих этих объектов вероятность равна 0, откуда можно заключить, что при $\det R = 0$ у вектора не существует плотности распределения.

Вид плотности распределения в невырожденном случае

Пусть $\det R \neq 0$, тогда вектор (e_1, e_2), независимые компоненты которого имеют распределение $\mathrm{N}(0;1)$, имеет плотность распределения:

$$\rho_{(e_1,e_2)}(x, y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-y^2/2} = \frac{1}{2\pi} \cdot e^{-(x^2+y^2)/2}$$

Допустим, что мы хотим посчитать следующую вероятность:

$$\mathbb{P}((\xi, \eta) \in A) = \iint\limits_A \rho(x, y)\, dx\, dy$$

Так как любой такой вектор можно представить в виде $\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \mu + C \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$, то

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in A \Leftrightarrow \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \in C^{-1} \cdot (A - \mu) = \{C^{-1} \cdot (a - \mu) \mid a \in A\}$$

Тогда

$$\begin{aligned} \mathbb{P}\left(\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in A\right) &= \mathbb{P}\left(\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \in C^{-1} \cdot (A - \mu)\right) \\ &= \iint\limits_{C^{-1} \cdot (A-\mu)} \frac{1}{2\pi} \cdot e^{-(u^2+v^2)/2}\, du\, dv \end{aligned}$$

Сделаем следующие замены переменных (именно переменных, наши действия сейчас отличаются от воздействия на случайные величины):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mu + C \cdot \begin{pmatrix} u \\ v \end{pmatrix} \implies \begin{pmatrix} u \\ v \end{pmatrix} = C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right)$$
$$\begin{pmatrix} u \\ v \end{pmatrix} \in C^{-1} \cdot (A - \mu) \Leftrightarrow \begin{pmatrix} x \\ y \end{pmatrix} \in A, \quad |J| = |\det C|^{-1}$$

Рассмотрим изменение в функции под интегралом:

$$\exp\left(-\frac{1}{2} \left\langle \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle \right) = \exp\left(-\frac{1}{2} \left\langle C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right), C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\rangle \right)$$

Немного линейной алгебры:

$$\left. \begin{aligned} &(C^T)^{-1} \cdot C^T = E \\ &\left(C^T \cdot (C^{-1})^T\right)^T = C^{-1} \cdot C = E \end{aligned} \right\} \implies (C^T)^{-1} = (C^{-1})^T$$
$$R = C \cdot C^T \implies R^{-1} = (C^T)^{-1} \cdot C^{-1} = (C^{-1})^T \cdot C^{-1}$$

Обратно к функции:

$$\begin{aligned} \ldots &= \exp\left(-\frac{1}{2} \left\langle C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right), C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\rangle \right) \\ &= \exp\left(-\frac{1}{2} \left\langle (C^{-1})^T \cdot C^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right), \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\rangle \right) \\ &= \exp\left(-\frac{1}{2} \left\langle R^{-1} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right), \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\rangle \right) \\ &= \exp\left(-\frac{1}{2} \left\| R^{-\frac{1}{2}} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\|^2 \right) \end{aligned}$$

Итоговый вид интеграла с учётом изменений и якобиана:

$$\mathbb{P}\left(\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in A \right) = \iint\limits_A \frac{1}{2\pi \cdot \sqrt{\det R}} \cdot \exp\left(-\frac{1}{2} \left\| R^{-\frac{1}{2}} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\|^2 \right) dx\, dy$$

Билет №9.

Независимость выборочного среднего и выборочной дисперсии независимых нормально распределённых величин

Пусть $\xi_1, \ldots, \xi_n$ — простая выборка с распределением $\mathrm{N}(0;1)$, тогда

$$\overline{\xi} = \frac{\xi_1 + \cdots + \xi_n}{n} \sim \mathrm{N}\left(0; \frac{1}{n}\right) \qquad S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_i - \overline{\xi})^2$$

Th. *$\overline{\xi}$ и S^2 независимы.*

Доказательство. Возьмём ортогональную матрицу U, у которой первая строка состоит из $\frac{1}{\sqrt{n}}$, и умножим её на столбец, составленный из ξ_k:

$$\begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} = \begin{pmatrix} \sqrt{n}\cdot\overline{\xi} \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

У ортогональных матриц есть несколько замечательных свойств, а именно:

1. $U \cdot U^T = E \Leftrightarrow U^{-1} = U^T$ — обратная матрица равна транспонированной исходной.

2. $\langle Ux, Uy\rangle = \langle x, y\rangle$ — не затрагивает результат скалярного произведения.

3. $\|Ux\| = \sqrt{\langle Ux, Ux\rangle} = \sqrt{\langle x, x\rangle} = \|x\|$ — не влияет на длину (как следствие).

Как дополнить матрицу так, чтобы она стала ортогональной:

1. Дополнить её линейно независимыми векторами.

2. Применить метод ортогонализации Грама-Шмидта.

Теперь поймём, что, в силу ортогональности, матрица ковариации у вектора в правой части равенства равна единичной. Тогда $y_2, \ldots, y_n$ — такие независимые случайные величины, что

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} \sim \mathrm{N}(0; E) \sim \begin{pmatrix} \sqrt{n}\cdot\overline{\xi} \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Так как длина сохранилась, то

$$\left\|\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}\right\| = \left\|\begin{pmatrix} \sqrt{n}\cdot\overline{\xi} \\ y_2 \\ \vdots \\ y_n \end{pmatrix}\right\| \Longrightarrow \sum_{k=1}^{n}\xi_k^2 = n\cdot\overline{\xi}^2 + \sum_{k=2}^{n} y_k^2 \Longrightarrow (n-1)\cdot S^2 = \sum_{k=2}^{n} y_k^2$$

Выходит, что S^2 независима от $\overline{\xi}$. **[:|||:]**

Распределение хи-квадрат

Def. *Если $y_1, \ldots, y_m$ — независимые случайные величины с распределением* $\mathrm{N}(0;1)$, *то величина $y_1^2 + \cdots + y_m^2$ имеет χ^2-распределение с m степенями свободы.*

В теореме выше, S^2 имеет χ^2-распределение с $n-1$ степенью свободы.

Билет №10.

Условные математические ожидания относительно дискретных величин

Пусть на Ω задана дискретная случайная величина ξ, то есть такая случайная величина, что она принимает не более чем счётное количество значений. Тогда её распределение выглядит так:

ξ	ξ_1	$\dots$	ξ_m	$\dots$
$\mathbb{P}$	$\mathbb{P}(A_1)$	$\dots$	$\mathbb{P}(A_m)$	$\dots$

Эту случайную величину можно записать в виде следующей суммы:

$$\xi = \sum_{k=1}^{m} \xi_k \cdot \mathrm{Ind}_{A_k}$$

Пусть нам известно, что произошло такое событие B, что $\mathbb{P}(B) > 0$, тогда условное ожидание ξ при условии события B выглядит так:

$$\begin{aligned}
\mathbb{E}(\xi \mid B) &= \sum_{k=1}^{m} \xi_k \cdot \mathbb{P}(A_k \mid B) \\
&= \sum_{k=1}^{m} \xi_k \cdot \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} \\
&= \sum_{k=1}^{m} \xi_k \cdot \frac{\mathbb{E}\,\mathrm{Ind}_{A_k \cap B}}{\mathbb{P}(B)} \\
&= \sum_{k=1}^{m} \xi_k \cdot \frac{\mathbb{E}(\mathrm{Ind}_{A_k} \cdot \mathrm{Ind}_B)}{\mathbb{P}(B)} \\
&= \frac{1}{\mathbb{P}(B)} \cdot \mathbb{E}\left[\mathrm{Ind}_B \cdot \sum_{k=1}^{m} \xi_k \cdot \mathrm{Ind}_{A_k}\right] \text{ (в силу линейности ожидания)} \\
&= \frac{1}{\mathbb{P}(B)} \cdot \mathbb{E}(\mathrm{Ind}_B \cdot \xi) \\
&= \frac{\mathbb{E}(\xi \cdot \mathrm{Ind}_B)}{\mathbb{P}(B)}
\end{aligned}$$

Обобщим на произвольные случайные величины.

Def. *Пусть произошло такое событие B, что $\mathbb{P}(B) > 0$. Пусть ξ — такая случайная величина, что $\mathbb{E}\xi < \infty$. Тогда условное ожидание ξ при условии B определяется следующим образом:*

$$\mathbb{E}(\xi \mid B) = \frac{\mathbb{E}(\xi \cdot \mathrm{Ind}_B)}{\mathbb{P}(B)}$$

Пусть задана случайная величина $\eta = \sum\limits_{k=1}^{n} \eta_k \cdot \mathrm{Ind}_{B_k}$, $\mathbb{P}(B_k) > 0$. Это задаёт разбиение Ω на B_k.

$$B_k = \{\omega \mid \eta(\omega) = \eta_k\} \implies \mathbb{E}(\xi \mid B_k) = \mathbb{E}(\xi \mid \eta = \eta_k)$$

Вообще говоря, в этом случае возникает случайная величина:

$$\omega \;\longmapsto\; \eta(\omega) = \eta_k \;\longmapsto\; \mathbb{E}(\xi \mid \eta = \eta_k)$$

Но рассматривать по одному значению нам неинтересно, поэтому

$$\omega \;\longmapsto\; \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k}(\omega)$$

Def. *Условным ожиданием случайной величины ξ при условии η называется случайная величина, описанная выше.*
Обозначение: $\mathbb{E}(\xi \mid \eta)$

Для неё можно построить таблицу распределения:

$\mathbb{E}(\xi \mid \eta)$	$\ldots$	$\mathbb{E}(\xi \mid \eta = \eta_k)$	$\ldots$
$\mathbb{P}$	$\ldots$	$\mathbb{P}(\eta = \eta_k)$	$\ldots$

Заметим, что $\mathbb{E}(\xi \mid \eta) = F(\eta)$, то есть условное ожидание случайной величины ξ при условии η является функцией от η. То есть, $F(s) = \mathbb{E}(\xi \mid \eta = s)$.

Свойства

1. (лекция) Линейность: $\mathbb{E}(\alpha\xi + \beta\zeta \mid \eta) = \alpha\mathbb{E}(\xi \mid \eta) + \beta\mathbb{E}(\zeta \mid \eta)$.

Доказательство.

$$\begin{aligned}
\mathbb{E}(\alpha\xi + \beta\zeta \mid \eta) &= \sum_{k=1}^{n} \mathbb{E}(\alpha\xi + \beta\zeta \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \sum_{k=1}^{n} \frac{\mathbb{E}\left[(\alpha\xi + \beta\eta) \cdot \mathrm{Ind}_{\eta=\eta_k}\right]}{\mathbb{P}(\eta = \eta_k)} \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \alpha \cdot \sum_{k=1}^{n} \frac{\mathbb{E}\left[\xi \cdot \mathrm{Ind}_{\eta=\eta_k}\right]}{\mathbb{P}(\eta = \eta_k)} \cdot \mathrm{Ind}_{\eta=\eta_k} + \beta \cdot \sum_{k=1}^{n} \frac{\mathbb{E}\left[\zeta \cdot \mathrm{Ind}_{\eta=\eta_k}\right]}{\mathbb{P}(\eta = \eta_k)} \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \alpha \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} + \beta \sum_{k=1}^{n} \mathbb{E}(\zeta \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \alpha\mathbb{E}(\xi \mid \eta) + \beta\mathbb{E}(\zeta \mid \eta)
\end{aligned}$$

[:|||:]

2. (лекция) Монотонность: $\xi \geqslant \zeta \implies \mathbb{E}(\xi \mid \eta) \geqslant \mathbb{E}(\zeta \mid \eta)$ (в терминах почти наверное).

Доказательство. Рассмотрим условное ожидание разности:

$$\begin{aligned}
\mathbb{E}(\xi - \zeta \mid \eta) &= \sum_{k=1}^{n} \mathbb{E}(\xi - \zeta \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \sum_{k=1}^{n} \frac{\mathbb{E}\left[(\xi - \zeta) \cdot \mathrm{Ind}_{\eta=\eta_k}\right]}{\mathbb{P}(\eta = \eta_k)} \cdot \mathrm{Ind}_{\eta=\eta_k} \geqslant 0
\end{aligned}$$

[:|||:]

3. (лекция) Формула полной вероятности: $\mathbb{E}(\mathbb{E}(\xi \mid \eta)) = \mathbb{E}\xi$.

Доказательство.

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(\xi \mid \eta)) &= \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k}\right] \\
&= \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathbb{P}(\eta = \eta_k) \\
&= \sum_{k=1}^{n} \frac{\mathbb{E}(\xi \cdot \mathrm{Ind}_{\eta=\eta_k})}{\mathbb{P}(\eta = \eta_k)} \cdot \mathbb{P}(\eta = \eta_k) \\
&= \sum_{k=1}^{n} \mathbb{E}(\xi \cdot \mathrm{Ind}_{\eta=\eta_k}) \\
&= \mathbb{E}\left[\xi \cdot \sum_{k=1}^{n} \mathrm{Ind}_{\eta=\eta_k}\right] \\
&= \mathbb{E}\xi \\
\mathbb{E}\xi &= \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathbb{P}(\eta = \eta_k) \text{ — важный факт.}
\end{aligned}$$

[:|||:]

4. (лекция) Если ξ и η независимы, то $\mathbb{E}(\xi \mid \eta) = \mathbb{E}\xi$.

Доказательство.

$$\begin{aligned}
\mathbb{E}(\xi \mid \eta) &= \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
\mathbb{E}(\xi \mid \eta = \eta_k) &= \frac{\mathbb{E}(\xi \cdot \mathrm{Ind}_{\eta=\eta_k})}{\mathbb{P}(\eta = \eta_k)} \\
&= \frac{\mathbb{E}\xi \cdot \mathbb{E}\,\mathrm{Ind}_{\eta=\eta_k}}{\mathbb{P}(\eta = \eta_k)} \text{ (в силу независимости)} \\
&= \frac{\mathbb{E}\xi \cdot \mathbb{P}(\eta = \eta_k)}{\mathbb{P}(\eta = \eta_k)} \\
&= \mathbb{E}\xi \\
\mathbb{E}(\xi \mid \eta) &= \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \mathbb{E}\xi \cdot \sum_{k=1}^{n} \mathrm{Ind}_{\eta=\eta_k} \\
&= \mathbb{E}\xi
\end{aligned}$$

[:|||:]

5. (лекция) Можно выносить функции от η: $\mathbb{E}(\xi \cdot F(\eta) \mid \eta) = F(\eta) \cdot \mathbb{E}(\xi \mid \eta)$.

Доказательство.

$$\mathbb{E}(\xi \cdot F(\eta) \mid \eta) = \sum_{k=1}^{n} \mathbb{E}(\xi \cdot F(\eta) \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k}$$

$$\begin{aligned}
\mathbb{E}(\xi \cdot F(\eta) \mid \eta = \eta_k) &= \frac{\mathbb{E}(\xi \cdot F(\eta) \cdot \mathrm{Ind}_{\eta=\eta_k})}{\mathbb{P}(\eta = \eta_k)} \\
&= \frac{\mathbb{E}(\xi \cdot F(\eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k})}{\mathbb{P}(\eta = \eta_k)} \\
&= \frac{F(\eta_k) \cdot \mathbb{E}(\xi \cdot \mathrm{Ind}_{\eta=\eta_k})}{\mathbb{P}(\eta = \eta_k)} \\
&= F(\eta_k) \cdot \mathbb{E}(\xi \mid \eta = \eta_k)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\xi \cdot F(\eta) \mid \eta) &= \sum_{k=1}^{n} \mathbb{E}(\xi \cdot F(\eta) \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \sum_{k=1}^{n} F(\eta_k) \cdot \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= \sum_{k=1}^{n} F(\eta) \cdot \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= F(\eta) \sum_{k=1}^{n} \mathbb{E}(\xi \mid \eta = \eta_k) \cdot \mathrm{Ind}_{\eta=\eta_k} \\
&= F(\eta) \cdot \mathbb{E}(\xi \mid \eta)
\end{aligned}$$

[:|||:]

Решение задачи о наилучшем приближении

Th. (лекция) $\min\limits_{F(\eta)} [\mathbb{E}(\xi - F(\eta))^2]$ *достигается на единственной случайной величине* $\mathbb{E}(\xi \mid \eta)$, *в предположении, что* $\mathbb{E}\xi^2 < \infty$.

Доказательство. Рассмотрим наше ожидание подробнее (обозначим $G(\eta) = \mathbb{E}(\xi \mid \eta) - F(\eta)$):

$$\begin{aligned}
\mathbb{E}(\xi - F(\eta))^2 &= \mathbb{E}\left[\xi - \mathbb{E}(\xi \mid \eta) + \mathbb{E}(\xi \mid \eta) - F(\eta)\right]^2 \\
&= \mathbb{E}(\xi - \mathbb{E}(\xi \mid \eta))^2 + \mathbb{E}(G(\eta))^2 + 2\mathbb{E}\left[(\xi - \mathbb{E}(\xi \mid \eta)) \cdot G(\eta)\right]
\end{aligned}$$

Покажем, что $\mathbb{E}\left[(\xi - \mathbb{E}(\xi \mid \eta)) \cdot G(\eta)\right] = 0$:

$$\begin{aligned}
\mathbb{E}(\xi \cdot G(\eta)) - \mathbb{E}(G(\eta) \cdot \mathbb{E}(\xi \mid \eta)) &= \mathbb{E}(\xi \cdot G(\eta)) - \mathbb{E}(\mathbb{E}(G(\eta) \cdot \xi \mid \eta)) \text{ (можем вносить функции от } \eta) \\
&= \mathbb{E}(\xi \cdot G(\eta)) - \mathbb{E}(\xi \cdot G(\eta)) \text{ (по полной вероятности)} \\
&= 0
\end{aligned}$$

Теперь имеем:

$$\mathbb{E}(\xi - F(\eta))^2 = \mathbb{E}(\xi - \mathbb{E}(\xi \mid \eta))^2 + \mathbb{E}(G(\eta))^2$$

Ввиду того, что $\mathbb{E}(G(\eta))^2 \geqslant 0$, имеем:

$$\mathbb{E}(\xi - F(\eta))^2 = \mathbb{E}(\xi - \mathbb{E}(\xi \mid \eta))^2 + \mathbb{E}(G(\eta))^2 \geqslant \mathbb{E}(\xi - \mathbb{E}(\xi \mid \eta))^2$$

Видно, что минимум достигается при $G(\eta) = 0 = \mathbb{E}(\xi \mid \eta) - F(\eta) \implies F(\eta) = \mathbb{E}(\xi \mid \eta)$ (почти наверное). **[:|||:]**

Геометрический смысл

(лекция) Случайные величины ζ с $\mathbb{E}\zeta^2 < \infty$ образуют Евклидово пространство со скалярным произведением $\langle \zeta_1, \zeta_2 \rangle = \mathbb{E}(\zeta_1 \cdot \zeta_2)$. Заметим, что случайные величины $F(\eta)$ образуют линейное пространство:

$$L = \{F(\eta) \mid \eta \colon \mathbb{E}\eta^2 < \infty\}$$

Утверждается, что условное ожидание является проекцией ξ на L, причём $F(\eta)$ является данной проекцией тогда и только тогда, когда $\xi - F(\eta) \perp L$, то есть для любых $G(\eta)$ верно равенство

$$\mathbb{E}((\xi - F(\eta)) \cdot G(\eta)) = 0 \iff \mathbb{E}(\xi \cdot G(\eta)) = \mathbb{E}(F(\eta) \cdot G(\eta))$$