

ИИИ МГУ



AI Masters

Конспект лекций

Антон Рябченко

**Математическая статистика и
приложения**

*Составитель
конспекта:
Вадим Смирнов*

15 февраля 2026 г.

Содержание

1 Введение	1
1.1 Напоминание из теории вероятностей	1
1.1.1 Основные определения	1
1.1.2 Независимость случайных величин	7
1.1.3 Моменты случайных величин и другие функционалы распределений	7
1.1.4 Вероятностные неравенства	12
1.1.5 Линейная и евклидова структура пространств случайных величин	14
1.1.6 Типы сходимости	23
2 Оценивание	27
2.1 Лекция 1	27
2.1.1 Основные определения	27
2.1.2 Задача точечной оценки параметра	29
2.1.3 Эффективные оценки	31
2.1.4 Экспоненциальное семейство распределений	35
2.2 Лекция 2	36
2.2.1 Достаточные статистики	36
3 Тестирование гипотез	36

1 Введение

1.1 Напоминание из теории вероятностей

Этого напоминания не было в лекциях, но мне удобнее вставить его сюда.

1.1.1 Основные определения

Определение 1.1: Сигма-алгебра

σ -алгеброй на множестве Ω называется непустое семейство его подмножеств, замкнутое относительно дополнения до Ω , не более чем счётных объединений и не более чем счётных пересечений.

Определение 1.2: Измеримое множество

Множество с заданной на нём σ -алгеброй называется измеримым.

Определение 1.3: Измеримое подмножество

Подмножество измеримого множества, принадлежащее его σ -алгебре, называется измеримым.

Определение 1.4: Измеримая функция

Пусть (Ω_1, σ_1) , (Ω_2, σ_2) - измеримые множества. Функция $f : \Omega_1 \rightarrow \Omega_2$ называется измеримой, если полный прообраз любого измеримого подмножества в Ω_2 является измеримым в Ω_1 , т.е.

$$\forall s_2 \in \sigma_2 : \{\omega_1 \in \Omega_1 | f(\omega_1) \in s_2\} \in \sigma_1$$

Замечание 1. Измеримые множества образуют категорию **Meas**, морфизмами которой являются измеримые отображения. Это конкретная категория очевидным забывающим функтором $\square : \mathbf{Meas} \rightarrow \mathbf{Sets}$.

Определение 1.5: Вероятностная мера

Вероятностная мера - это функция P из σ -алгебры множества X в $[0, 1] \subset \mathbb{R}$, такая, что:

1. $P(\emptyset) = 0$
2. $P(\bigcup_{k=1}^{\infty} s_k) = \sum_{k=1}^{\infty} P(s_k)$ для всякого не более чем счётного семейства попарно непересекающихся множеств $s_k \in \sigma$
3. $P(\Omega) = 1$

Определение 1.6: Вероятностное пространство

Тройка (Ω, σ, P) , где Ω - множество, σ - σ -алгебра, и P - вероятностная мера, называется вероятностным пространством.

Определение 1.7: Расширение вероятностного пространства

Говорят, что вероятностное пространство (Ω', σ', P') расширяет вероятностное пространство (Ω, σ, P) , если задано сюръективное измеримое отображение $\pi : \Omega' \rightarrow \Omega$, сохраняющее вероятность, т.е. $\forall s \in \sigma : P'(\pi^{-1}(s)) = P(s)$

Пример 1. Вероятностное пространство бросков двух шестигранных кубиков расширяет вероятностное пространство бросков одного шестигранного кубика. Расширение задаётся сюръекцией $\pi : (x_1, x_2) \rightarrow x_1$, где x_1, x_2 - исходы первого и второго броска соответственно. В качестве иллюстрации свойства сохранения вероятности рассмотрим событие в "выпадает единица или четвёрка" в вероятностном пространстве, соответствующем броску одного кубика. Его вероятность - $1/3$. Полный образ этого события - "Выпадает единица или четвёрка на первом кубике, и что угодно на втором". Его вероятность - по-прежнему $1/3$.

Замечание 2. Вероятностные пространства также образуют категорию, морфизмы которой являются расширения вероятностных пространств.

Определение 1.8: Случайная величина

Случайной величиной называется измеримая функция f из множества событий Ω вероятностного пространства (Ω, σ, P) в некоторое измеримое множество (\mathcal{X}, Σ) с σ -алгеброй Σ .

Определение 1.9: Реализация случайной величины

Реализациями случайной величины называются измеримые подмножества $X \subset \mathcal{X}$.

Определение 1.10: Вероятность реализации

Вероятность реализации $X \subset \mathcal{X}$ - это $P(f^{-1}(X))$.

Это определение корректно, так как f измерима.

Теорема 1.1: Расширение сохраняет вероятности

Расширение вероятностных пространств сохраняет вероятность реализаций, т.е.

$$\forall X \subset \mathcal{X} : P'((f \circ \pi)^{-1}(X)) = P(f^{-1}(X))$$

Доказательство. Действительно, пусть $\pi : (\Omega', \sigma', P') \rightarrow (\Omega, \sigma, P)$ - расширение вероятностных пространств.

$$(f \circ \pi)^{-1} = \pi^{-1} \circ f^{-1}$$

$$P'(\pi^{-1}(f^{-1}(X))) = P(f^{-1}(X))$$

по определению расширения. \square

Определение 1.11: Борелевское множество

Борелевским множеством называется подмножество топологического пространства, получаемое из открытых подмножеств этого пространства операциями дополнения, не более чем счётного пересечения и не более чем счётного объединения.

Теорема 1.2: Алгебра борелевских множеств

Борелевские множества $\mathcal{B}(M)$ любого топологического пространства M образуют σ -алгебру.

Доказательство. Очевидно. \square

Пример 2. \mathbb{R}^n является топологическим пространством (с топологией, порождённой стандартной евклидовой метрикой, с базой из открытых шаров всевозможных ненулевых радиусов и со всевозможными центраторами). Его борелевская сигма-алгебра порождается (при помощи операций, определяющих сигма-алгебру) этими же открытыми шарами.

Как правило, целевое множество является линейным пространством $(\mathbb{R}^n, +)$. Это позволяет определять интегральные характеристики случайных величин, например, математическое ожидание.

Определение 1.12: Функция случайной величины

Функцией \mathcal{X} -значной случайной величины $f : \Omega \rightarrow \mathcal{X}$, где (Ω, σ, P) - вероятностное пространство, называется измеримое отображение $g : \mathcal{X} \rightarrow \mathcal{Y}$ в измеримое пространство \mathcal{Y} . Функция такой случайной величины задаёт \mathcal{Y} -значную случайную величину $g \circ f$ на вероятностном пространстве (Ω, σ, P) . Это действительно случайная величина, так как композиция измеримых отображений измерима.

Теорема 1.3: Расширение - естественное преобразование

Расширение вероятностных пространств $\pi : \Omega' \rightarrow \Omega$ переводит функции случайных величин на Ω в функции случайных величин на Ω' , сопоставляя случайной величине $f : \Omega \rightarrow \mathcal{X}$ случайную величину $f \circ \pi : \Omega' \rightarrow \mathcal{X}$. Вероятности всех реализаций \mathcal{X} при этом сохраняются.

Доказательство. Аналогично теореме о сохранении вероятности расширениями. \square

Определение 1.13: Произведение сигма-алгебр

Пусть $(\Omega_1, \sigma_1), (\Omega_2, \sigma_2)$ - измеримые пространства. Произведением σ -алгебр σ_1 и σ_2 , обозначаемым $\sigma_1 \otimes \sigma_2$, называется σ -алгебра на декартовом произведении $\Omega_1 \times \Omega_2$, порождённая декартовым произведением σ_1 и σ_2 как наборов подмножеств Ω_1 и Ω_2 , то есть минимальная по включению σ -алгебра на $\Omega_1 \times \Omega_2$, содержащая $\sigma_1 \times \sigma_2$ как подмножество.

Определение 1.14: Распределение случайной величины

Распределением случайной величины $f : \Omega \rightarrow \mathcal{X}$, заданной на вероятностном пространстве (Ω, σ, P) и принимающей значения в (\mathcal{X}, Σ) , называется вероятностная мера p , дополняющая пару (\mathcal{X}, Σ) до вероятностного пространства, и определённая соотношением

$$\forall S \in \Sigma : p(S) := P(f^{-1}(S))$$

В дальнейшем мы будем обычно обозначать меру на исходном ве-

роятностном пространстве и на пространстве значений одной и той же буквой P . Говорят, что случайная величина f индуцирует меру $f^*P = p$ на \mathcal{X} , или что p является пушфорвардом P при помощи f .

Определение 1.15: Совместное распределение

Если две случайные величины f_1, f_2 заданы на одном и том же вероятностном пространстве (Ω, σ, P) , и принимают значения в измеримых пространствах $(\mathcal{X}_1, \Sigma_1)$, $(\mathcal{X}_2, \Sigma_2)$, то определено их совместное распределение, то есть определена случайная величина (f_1, f_2) на вероятностном пространстве (Ω, σ, P) , принимающая значения в измеримом пространстве $(\mathcal{X}_1 \times \mathcal{X}_2, \Sigma_1 \otimes \Sigma_2)$, по правилу

$$(f_1, f_2)(\omega \in \Omega) = (f_1(\omega), f_2(\omega))$$

$$P(X_1 \times X_2 \subset \mathcal{X}_1 \times \mathcal{X}_2) = P(f_1^{-1}(X_1) \cap f_2^{-1}(X_2))$$

Замечание 3. Если (Ω, σ, P) - вероятностное пространство, $f_1 : \Omega \rightarrow R_1, f_2 : \Omega \rightarrow R_2$ - случайные величины, а $g : R_1 \times R_2 \rightarrow R$ измерима, то композиция $g \circ (f_1, f_2)$ - случайная величина на (Ω, σ, P) . Например, если $R_1, R_2, R = \mathbb{R}$, и $g(x_1, x_2) = x_1 + x_2$, таким образом определена сумма случайных величин $f_1 + f_2$.

Определение 1.16: Функция распределения

Функцией распределения вещественнозначной случайной величины $f : \Omega \rightarrow \mathbb{R}$, где $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ - структура измеримого пространства на \mathbb{R} , называется функция

$$F(x) := p((-\infty, x])$$

Определение 1.17: Плотность распределения

Плотностью распределения случайной величины $f : \Omega \rightarrow \mathbb{R}$ называется такая функция $\rho(x) : \mathbb{R} \rightarrow \mathbb{R}_+$, что для всякого $S \in \mathcal{B}(\mathbb{R})$

$$p(S) = \int_S \rho(x) dx$$

(если она существует).

Замечание 4. Если производная от $F(x)$ по x существует, она равна $\rho(x)$ почти всюду относительно меры P на области своего определения.

1.1.2 Независимость случайных величин

Определение 1.18: Независимость

Две случайные величины $f_1 : \Omega \rightarrow R_1, f_2 : \Omega \rightarrow R_2$, определённые на одном и том же вероятностном пространстве (Ω, σ, P) , со значениями в $(R_1, \Sigma_1), (R_2, \Sigma_2)$ соответственно, называются независимыми, если для всякого измеримого подмножества $R_1 \times R_2$ вида $X_1 \times X_2 \in \Sigma_1 \times \Sigma_2$ выполняется

$$P(X_1 \times X_2) = P(X_1) \cdot P(X_2)$$

где $P(X_1 \times X_2)$ вычислено согласно совместному распределению f_1 и f_2

1.1.3 Моменты случайных величин и другие функционалы распределений

Определение 1.19: Медиана

Вещественное число $M(f)$ называется медианой случайной величины $f : \Omega \rightarrow \mathbb{R}$, если одновременно

$$P(f > M(f)) \leq 1/2$$

$$P(f < M(f)) \leq 1/2$$

то есть одновременно не более чем в половине случаев случайная величина принимает значения меньше медианы, и не более чем в половине случаев - больше медианы.

Теорема 1.4: О существовании медианы

Медиана случайной величины всегда существует. Если функция распределения $f, F(x)$, непрерывна, то медиана также единственная.

Определение 1.20: Матожидание

Если \mathcal{X} - подмножество \mathbb{R}^n , $f : \Omega \rightarrow \mathcal{X}$ - случайная величина, $p = f^*P$ - индуцированная ей мера на \mathcal{X} , то лебегов интеграл

$$\mathbb{E}[f] := \int_{\mathcal{X}} x dp(x)$$

называется математическим ожиданием f . В случае, если плотность распределения f , $\rho(x)$, существует, то

$$\mathbb{E}[f] = \int_{\mathcal{X}} x \rho(x) dx$$

Замечание 5. Если $f : \Omega \rightarrow \mathbb{R}^n$ - случайная величина, то $\forall c \in \mathbb{R} : cf : \Omega \rightarrow \mathbb{R}^n$ - также случайная величина.

Теорема 1.5: Линейность матожидания

Математическое ожидание линейно: если $f_1, f_2 : \Omega \rightarrow \mathcal{X} \subseteq \mathbb{R}^n$ - случайные величины, то

$$\mathbb{E}[c_1 f_1 + c_2 f_2] = c_1 \mathbb{E}[f_1] + c_2 \mathbb{E}[f_2]$$

Доказательство. Это просто линейность интеграла Лебега. \square

Теорема 1.6: Монотонность матожидания

Математическое ожидание монотонно: если для пары случайных величин $f, g : \Omega \rightarrow \mathcal{X} \subseteq \mathbb{R}$ выполняется $\forall \omega \in \Omega : f(\omega) \leq g(\omega)$, то $\mathbb{E}[f] \leq \mathbb{E}[g]$.

Доказательство. Опять свойство интеграла Лебега. \square

Теорема 1.7: Формула замены переменных

Если $f : \Omega \rightarrow \mathbb{R}$ - случайная величина, и $F : \mathbb{R} \rightarrow \mathbb{R}$ - измеримая абсолютно интегрируемая функция, то

$$\mathbb{E}[F(f)] = \int_{\mathbb{R}} F(x) dp(x) = \int_{\mathbb{R}} F(x) \rho(x) dx$$

Доказательство.

□

Теорема 1.8: Матожидание произведения независимых случайных величин

Матожидание произведения независимых случайных величин $f_1, f_2 : \Omega \rightarrow \mathbb{R}$ равно произведению их матожиданий:

$$\mathbb{E}[f_1 f_2] = \mathbb{E}[f_1] \mathbb{E}[f_2]$$

Доказательство. Следующее доказательство не очень строгое, зато понятное:

$$\begin{aligned} \mathbb{E}[f_1 f_2] &= \int_{\mathbb{R} \times \mathbb{R}} dp(x_1, x_2) = \int_{\mathbb{R} \times \mathbb{R}} p(dx_1 \times dx_2) = \\ &= \int_{\mathbb{R} \times \mathbb{R}} p(dx_1)p(dx_2) = \int_{\mathbb{R}} dp(x_1) \int_{\mathbb{R}} dp(x_2) = \mathbb{E}[f_1] \mathbb{E}[f_2] \end{aligned}$$

□

Определение 1.21: Моменты

n -ым моментом \mathbb{R} -значной случайной величины f , индуцирующей меру p на $\mathcal{X} \subseteq \mathbb{R}$, называется интеграл

$$M_k(f) := \mathbb{E}[f^n] = \int_{\mathcal{X}} x^n dp(x) = \int_{\mathcal{X}} x^n \rho(x) dx$$

если указанный интеграл существует.

Определение 1.22: Дисперсия

Дисперсией случайной величины $f : \Omega \rightarrow \mathbb{R}$ называется величина

$$\mathbb{D}[f] := \mathbb{E}[(f - \mathbb{E}[f])^2]$$

Теорема 1.9: Дисперсия - кумулянт 2 порядка

Константы выносятся из дисперсии с квадратом:

$$\forall c \in \mathbb{R} : \mathbb{D}[cf] = c^2 \mathbb{D}[f]$$

Доказательство.

$$\begin{aligned}\mathbb{E}[(cf - \mathbb{E}[cf])^2] &= \mathbb{E}[c^2f^2 - 2\mathbb{E}[cf]cf + \mathbb{E}[cf]^2] = \\ &= c^2\mathbb{E}[f^2 - 2\mathbb{E}[f]f + \mathbb{E}[f]^2] = c^2\mathbb{D}[f]\end{aligned}$$

□

Теорема 1.10: Формула дисперсии

$$\mathbb{D}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2$$

Доказательство. Продолжая выкладку в предыдущем доказательстве и пользуясь линейностью матожидания,

$$\begin{aligned}\mathbb{E}[f^2 - 2\mathbb{E}[f]f + \mathbb{E}[f]^2] &= \mathbb{E}[f^2] - 2\mathbb{E}[f]^2 + \mathbb{E}[f]^2 \\ &= \mathbb{E}[f^2] - \mathbb{E}[f]^2\end{aligned}$$

□

Теорема 1.11: Дисперсия суммы независимых случайных величин

Если $f_1, f_2 : \Omega \rightarrow \mathbb{R}$ - независимые случайные величины, то

$$\mathbb{D}[f_1 + f_2] = \mathbb{D}[f_1] + \mathbb{D}[f_2]$$

Доказательство. По формуле дисперсии

$$\begin{aligned}\mathbb{D}[f_1 + f_2] &= \mathbb{E}[(f_1 + f_2)^2] - \mathbb{E}[f_1 + f_2]^2 = \\ &= \mathbb{E}[f_1^2 + 2f_1f_2 + f_2^2] - (\mathbb{E}[f_1] + \mathbb{E}[f_2])^2 = \\ &= \mathbb{E}[f_1^2] + \mathbb{E}[f_2^2] + 2\mathbb{E}[f_1f_2] - \mathbb{E}[f_1]^2 - \mathbb{E}[f_2]^2 - 2\mathbb{E}[f_1]\mathbb{E}[f_2]\end{aligned}$$

Пользуясь свойством матожидания произведения

$$\begin{aligned}&= \mathbb{E}[f_1^2] + \mathbb{E}[f_2^2] + 2\mathbb{E}[f_1]\mathbb{E}[f_2] - \mathbb{E}[f_1]^2 - \mathbb{E}[f_2]^2 - 2\mathbb{E}[f_1]\mathbb{E}[f_2] = \\ &= (\mathbb{E}[f_1^2] - \mathbb{E}[f_1]^2) + (\mathbb{E}[f_2^2] - \mathbb{E}[f_2]^2) = \\ &= \mathbb{D}[f_1] + \mathbb{D}[f_2]\end{aligned}$$

□

Теорема 1.12: Дисперсия инвариантна относительно сдвигов

Для всякой случайной величины $f : \Omega \rightarrow \mathbb{R}$ и всякой константы (постоянной случайной величины) $c \in \mathbb{R}$

$$\mathbb{D}[f + c] = \mathbb{D}[f]$$

Доказательство.

$$\begin{aligned}\mathbb{D}[f + c] &= \mathbb{E}[(f + c)^2] - \mathbb{E}[f + c]^2 = \\ &= \mathbb{E}[f^2] + 2\mathbb{E}[fc] + \mathbb{E}[c^2] - \mathbb{E}[f]^2 - 2\mathbb{E}[c]\mathbb{E}[f] - \mathbb{E}[c]^2 \\ &= \mathbb{E}[f^2] + 2c\mathbb{E}[f] + c^2 - \mathbb{E}[f]^2 - 2c\mathbb{E}[f] - c^2 \\ &= \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \mathbb{D}[f]\end{aligned}$$

□

Определение 1.23: Стандартное отклонение

Стандартным отклонением σ случайной величины f , имеющей дисперсию $\mathbb{D}[f]$, называется её положительный корень $\sqrt{\mathbb{D}[f]}$

Из свойств дисперсии следует, что стандартное отклонение неотрицательно и линейно для независимых случайных величин.

Определение 1.24: Факториальная производящая функция

Факториальной производящей функцией моментов случайной величины $f : \Omega \rightarrow \mathbb{R}$ называется функция

$$M(t) = \mathbb{E} \left[\frac{1}{1 - tf} \right] = \sum_{k=0}^{\infty} t^k M_k(f)$$

Определение 1.25: Производящая функция моментов

Производящей функцией моментов случайной величины $f : \Omega \rightarrow \mathbb{R}$

называется функция

$$G(t) = \mathbb{E}[e^{tf}] = \sum_{k=0}^{\infty} \frac{t^k}{k!} M_k(f)$$

Определение 1.26: Производящая функция кумулянтов

Производящей функцией кумулянтов случайной величины $f : \Omega \rightarrow \mathbb{R}$ называется функция

$$\Gamma(t) := \ln G(t) = \sum_{k=1}^{\infty} \frac{t^k}{k!} K_k(f)$$

где величина $K_k(f)$ называется k -ым кумулянтом f .

Пример 3. $K_1(f) = \mathbb{E}[f]$, $K_2(f) = \mathbb{D}[f]$

Замечание 6. Кумулянты иногда называют полуинвариантами. На языке комбинаторики факториальная производящая функция моментов - это обыкновенная производящая функция, производящая функция - экспоненциальная производящая функция. Физики знают производящую функцию кумулянтов как производящий функционал связных диаграмм Фейнмана; в общем случае в комбинаторике кумулянты соответствуют связным, в некотором смысле, обЪектам.

Определение 1.27: Характеристическая функция

Характеристической функцией случайной величины $X : \Omega \rightarrow \mathbb{R}$ называется «преобразование Фурье» её плотности:

$$\chi(t) := \mathbb{E}[\exp(itX)] \tag{1}$$

1.1.4 Вероятностные неравенства

Теорема 1.13: Неравенство Маркова

Если случайная величина $f : \Omega \rightarrow \mathbb{R}$ принимает только неотрица-

тельные значения, то

$$\forall \lambda \geq 0 : P(f \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}[f]$$

Доказательство. Рассмотрим случайную величину $\lambda \mathbb{I}(f \geq \lambda)(\omega)$, которая равна нулю на тех $\omega \in \Omega$, на которых $f(\omega) < \lambda$, и λ на тех ω , на которых $f(\omega) \geq \lambda$. В силу линейности математического ожидания

$$\mathbb{E}[\lambda \mathbb{I}(f \geq \lambda)] = \lambda \mathbb{E}[\mathbb{I}(f \geq \lambda)] = \lambda P(f \geq \lambda)$$

В силу монотонности математического ожидания,

$$\lambda P(f \geq \lambda) \leq \mathbb{E}[f]$$

□

Теорема 1.14: Экспоненциальная граница

Если случайная величина имеет k -ый абсолютный момент, то есть $\mathbb{E}[|f|^k] \leq C \in \mathbb{R}$ то

$$\forall \lambda > 0 : P(|f| \geq \lambda) \leq C \lambda^{-k}$$

Доказательство. Из неравенства Маркова

$$P(|f| \leq \lambda) = P(|f|^k \leq \lambda^k) \leq \frac{1}{\lambda^k} \mathbb{E}[|f|^k]$$

□

Теорема 1.15: Неравенство Чебышёва

Пусть $\hat{\sigma}(f) := |f - \mathbb{E}[f]|$ - модуль отклонения случайной величины от её среднего. Тогда

$$P(\hat{\sigma}(f) \geq \lambda) \leq \frac{\mathbb{D}[f]}{\lambda^2}$$

Доказательство. Это неравенство Маркова для второго абсолютного момента $\hat{\sigma}(f)$. □

Теорема 1.16: Неравенство Йенсена

Пусть $f : \mathbb{R} \rightarrow \mathbb{R}$ - выпуклая функция, и пусть X - ограниченная вещественнонозначная случайная величина (т.е. вещественнонозначная случайная величина с ограниченным носителем). Тогда

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) \quad (2)$$

Если же функция f вогнутая, то при тех же условиях

$$\mathbb{E}[f(X)] \leq f(\mathbb{E}[X]) \quad (3)$$

1.1.5 Линейная и евклидова структура пространств случайных величин

Два способа складывать случайные величины

На множествах случайных величин существует две *разные* структуры вещественного векторного пространства. Говоря более точно, разные структуры имеют множество случайных величин на заданном пространстве и множество всех случайных величин, полученных из данного вероятностного пространства. Чтобы понять разницу, заметим, что в использованном нами выше смысле сумма двух одинаково распределённых случайных величин, заданных на одном и том же вероятностном пространстве, не равна ни одной из них, взятой двухкратно. Скажем, рассмотрим бернуlliевскую случайную величину ξ_p , принимающую значения 0 и 1 с вероятностью p и $1 - p$. Её можно рассматривать как тождественную функцию из множества $\{0, 1\}$ в $\{0, 1\} \subset \mathbb{R}$. Используемая нами сигма-алгебра - $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$, а вероятностная мера - функция, заданная таблицей

$$P(\emptyset) = 0, \quad P(\{1\}) = p, \quad P(\{0\}) = 1 - p, \quad P(\{0, 1\}) = 1$$

Мы можем сложить поточечно две одинаковые случайные величины ξ_p с коэффициентами α и β как измеримые функции $\{0, 1\} \rightarrow \mathbb{R}$, и получить новую измеримую функцию $\{0, 1\} \rightarrow \mathbb{R}$:

$$\alpha\xi_p + \beta\xi_p : \{0, 1\} \rightarrow \{0, \alpha + \beta\} = (\alpha + \beta)\xi_p$$

Её математическое ожидание равно, очевидно, $(\alpha + \beta)p$.

Аналогично, произвольная измеримая функция из $\{0, 1\}$ в \mathbb{R} имеет вид

$$f_{\alpha, \beta} : \{0, 1\} \rightarrow \{\alpha, \beta\}$$

Она может принимать не более двух значений, и нетрудно видеть, что такие случайные величины образуют двумерное вещественное линейное пространство с базисом $\{f_{1,0}, f_{0,1}\}$. Математическое ожидание для случайной величины $f_{\alpha, \beta}$ равно, очевидно, $(1-p)\alpha + p\beta$. При этом случайная величина $f_{0,0} : \{0, 1\} \rightarrow \{0, 0\}$, всегда принимающая значение 0, является нулем этого линейного пространства.

Сложим две теперь две *независимые* случайные величины ξ_{p1} и ξ_{p2} , имеющие одинаковое распределение, с коэффициентами α и β . Для этого нам сначала нужно определить их совместное распределение. Заметим, что независимые одинаково распределённые случайные величины не могут быть совпадать как измеримые функции. Действительно, чтобы сложить две случайные величины, нужно сначала определить их совместное распределение. Мы конструируем случайную величину (ξ_{p1}, ξ_{p2}) по правилу

$$(\xi_{p1}, \xi_{p2})(\omega \in \Omega) := (\xi_{p1}(\omega), \xi_{p2}(\omega))$$

т.е. $(\xi_{p1}, \xi_{p2}) : \{0, 1\} \rightarrow (\{0, 0\}, \{1, 1\})$. Вычислим $P(\{0, 1\})$:

$$P(\{0, 1\}) = P(\xi_{1p}^{-1}(\{0\}) \cap \xi_{2p}^{-1}(\{1\})) = P(\{0\} \cap \{1\}) = 0$$

Поскольку для независимых случайных величин

$$P_{(\xi_1, \xi_2)}(A \cap B) = P_{\xi_1}(A)P_{\xi_2}(B)$$

и ни один из сомножителей не равен нулю, приходим к противоречию. Отсюда вывод: сложение одинаково распределённых случайных величин требует расширения вероятностного пространства, на котором они определены. Итак, вместо пространства $(\{0, 1\}, 2^{\{0,1\}}, P)$ мы используем пространство $(\{0, 1\}^{\times 2}, 2^{\{0,1\}^{\times 2}}, P \times P)$, и определяем на нём величину (ξ_{p1}, ξ_{p2}) как

$$\xi_{p1} \times \xi_{p2}((\omega_1, \omega_2)) = (\xi_{p1}(\omega_1), \xi_{p2}(\omega_2))$$

Теперь мы можем сложить их как $\alpha\xi_{p1} + \beta\xi_2$ (умножение на скаляр определено на каждом члене прямого произведения так же, как в первом

случае), и получить новую случайную величину на том же вероятностном пространстве,

$$\alpha\xi_{p1} + \beta\xi_{p2} : (\omega_1, \omega_2) \rightarrow \alpha\xi_{p1}(\omega_1) + \beta\xi_{p2}(\omega_2)$$

Эта случайная величина может принимать уже четыре значения - $\{0, \alpha, \beta, \alpha + \beta\}$ - с вероятностями $\{(1-p)^2, p(1-p), p(1-p), p^2\}$ соответственно. При этом, в силу линейности матожидания, оно у неё также будет равно $p(\alpha + \beta)$, что проверяется прямым вычислением:

$$(\alpha + \beta)(p(1-p) + p^2) = (\alpha + \beta)p(p + 1 - p) = (\alpha + \beta)p$$

Итак, существует две линейные структуры, связанные со множествами случайных величин. Разберёмся, как работает каждая из них.

Линейное случайных величин на одном вероятностном пространстве

Теорема 1.17: Структура линейного пространства распределений случайных величин

Пусть задано вероятностное пространство (Ω, σ, P) . Тогда случайные величины, являющиеся измеримыми функциями из Ω в \mathbb{R}^n , образуют вещественное линейное пространство относительно операций поточечного сложения и умножения на число.

Замечание 7. Вообще, более правильно говорить о структуре линейного пространства на классах эквивалентности измеримых функций - две функции называются эквивалентными, если различаются на множестве меры 0 (т.е. $P(\xi_1 \neq \xi_2) = 0$, т.е. две величины совпадают с вероятностью 1).

Доказательство. Сумма и линейная комбинация двух классов эквивалентности измеримых функций измеримы. Класс эквивалентности случайной величины, имеющей плотность распределения $\delta(x)$ (то есть целиком сконцентрированной в нуле) является нейтральным элементом относительно сложения. \square

Это пространство естественным образом является нормированным.

Определение 1.28: L_p -нормы на пространстве случайных величин

Пусть $p \in [1, \infty)$. Пусть X - случайная величина на пространстве (Ω, σ, P) . Тогда

$$\|X\|_p := (\mathbb{E}[|X|^p])^{1/p} \quad (4)$$

Нетрудно проверить, что указанные функционалы действительно являются нормами, а именно:

1. $\|X\|_p \geq 0$, и $\|X\|_p = 0 \iff P(X = 0) = 1$
2. $\forall \alpha \in \mathbb{R} : \|\alpha X\|_p = |\alpha| \|X\|_p$
3. $\forall X_1, X_2 : \|X_1 + X_2\|_p \leq \|X_1\|_p + \|X_2\|_p$ (неравенство треугольника)

Действительно, первое и второе свойство очевидны и являются следствием линейности интеграла Лебега, а третье является следствием неравенства Минковского. Альтернативный способ[4] доказательства неравенства треугольника состоит в использовании неравенства Йенсена: заметим, что $\mathbb{R}_+ \times \mathbb{R}_+$ - выпуклое множество, а функция

$$g(x, y) = (x^{1/p} + y^{1/p})^p$$

вогнута на этом множестве.

Тогда по неравенству Йенсена для неотрицательных случайных величин U, V

$$\mathbb{E}[(U^{1/p} + V^{1/p})^p] \leq (\mathbb{E}[U]^{1/p} + \mathbb{E}[V]^{1/p})^p$$

Подставляя в качестве $U, V |X|^p, |Y|^p$ соответственно, получаем

$$\mathbb{E}[(|X| + |Y|)^p] \leq (\mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|Y|^p]^{1/p})^p$$

Извлекая корень степени p из обеих частей и используя неравенство $|X| + |Y| \geq |X + Y|$, а также монотонность матожидания, получаем требуемый результат.

В силу неравенства треугольника сумма двух случайных величин конечной L_p -нормы является случайной величиной конечной L_p -нормы,

а потому такие случайные величины, заданные на данном вероятностном пространстве, образуют линейное пространство L_p , являющееся знакомым из функционального анализа лебеговым функциональным пространством.

Теорема 1.18: Неравенство Ляпунова

Пусть $1 \leq j \leq k$. Пусть X - случайная величина. Тогда

$$\|X\|_j \leq \|X\|_k$$

Доказательство. Рассмотрим функцию $x^{j/k}$. Она выпукла и неотрицательна на \mathbb{R}_+ . Применим неравенство Йенсена к $x = |X|^k$ и получим требуемое. \square

Как известно из функционального анализа, любое нормированное пространство является метризуемым: метрика задаётся при помощи выражения

$$\rho_p(X, Y) = \|X - Y\|_p \quad (5)$$

Таким образом, в частности, стандартное отклонение оказывается L_2 -расстоянием от случайной величины до вырожденной случайной величины, равной её матожиданию на множестве меры 1:

$$\sigma(X) = \|X - \mathbb{E}[X]\|_2 = (\mathbb{E}[(X - \mathbb{E}[X])^2])^{1/2}$$

Определение 1.29: RMSE

Среднеквадратичным отклонением (ошибкой) вещественнозначной случайной величины X от константы t (root mean squared error) называется функция

$$\rho_2(X, t) = \|X - t\|_2 \quad (6)$$

Теорема 1.19: Минимум среднеквадратичной ошибки достигается на среднем

Пусть X - вещественнозначная случайная величина. Тогда

$$\arg \min_t \rho_2(X, t) = \mathbb{E}[X]$$

и

$$\min_t \rho_2(X, t) = \sigma(X)$$

Доказательство.

$$\begin{aligned} \mathbb{E}[(X - t)^2] &= \mathbb{E}[X^2 - 2tX + t^2] = \\ &= \mathbb{E}[X^2] - 2t\mathbb{E}[X] + t^2 \end{aligned}$$

Дифференцируя по t , получаем, что экстремум достигается в точке

$$2(t - \mathbb{E}[X]) = 0$$

и очевидным образом является минимумом. Таким образом, при минимальном значении RMSE

$$\rho_2(X, t)^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{D}[X]$$

и $\rho_2(X, t) = \sigma(X)$. □

Определение 1.30: Среднее абсолютное отклонение

Средним абсолютным отклонением вещественнонозначной случайной величины X от величины t называется

$$\rho_1(X, t) := \|X - t\|_1$$

Теорема 1.20: Медиана минимизирует среднее абсолютное отклонение

Если X - вещественнонозначная случайная величина, то

$$\arg \min_t \rho_1(X, t) = M(X)$$

Поскольку медиана может не быть единственной, среднее абсолютное отклонение может иметь больше одного минимума.

Евклидова структура пространства случайных величин

Пространство L_2 , как известно из функционального анализа, является не только банаховым, но и гильбертовым (т.е. банаховым и евклидовым). В вероятностных терминах скалярное произведение на L_2 определяется следующим образом:

Определение 1.31: Скалярное произведение на L_2 \mathbb{R} -значных случайных величин

Для вещественнозначных случайных величин X, Y , заданных на вероятностном пространстве (Ω, σ, P) и принадлежащих L_2 , скалярное произведение определено как

$$\langle X, Y \rangle := \mathbb{E}[XY] \quad (7)$$

Это действительно скалярное произведение, т.к. функция $\langle \cdot, \cdot \rangle$ билинейна, симметрична и невырождена.

Случайные величины из L_2 , имеющие нулевой первый момент, очевидным образом образуют подпространство в L_2 (например, потому, что это ядро линейного отображения матожидания). Любую случайную величину X из L_2 можно **центрировать**, отняв от неё её матожидание, и превратив, таким образом, в случайную величину с нулевым матожиданием. Обозначим эту операцию

$$C(X) := X - \mathbb{E}[X]$$

Отображение центрирования линейно в силу линейности матожидания. При этом отображение центрирования не изменяет центрированную случайную величину. Значит, $C^2 = C$, и C является проектором на подпространство центрированных случайных величин. Из линейной алгебры известно, что L_2 , в таком случае, разлагается на прямую сумму

$$L_2 = \text{Ker } C \oplus \text{Im } C$$

При этом $\text{Ker } C$ - это одномерное подпространство случайных величин с дельта-распределением $\rho_X(x) = \delta(x - \mathbb{E}[X])$. Таким образом, любую случайную величину из L_2 можно однозначно представить в виде

$$X = C(X) + \mathbb{E}[X]$$

Тогда

$$\begin{aligned} \langle X, Y \rangle &= \langle C(X) + \mathbb{E}[X], C(Y) + \mathbb{E}[Y] \rangle \\ &= \langle C(X), C(Y) \rangle + \mathbb{E}[X]\langle 1, C(Y) \rangle + \mathbb{E}[Y]\langle 1, C(X) \rangle + \mathbb{E}[X]\mathbb{E}[Y]\langle 1, 1 \rangle \\ &= \langle C(X), C(Y) \rangle + \mathbb{E}[X]\mathbb{E}[C(Y)] + \mathbb{E}[Y]\mathbb{E}[C(X)] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \langle C(X), C(Y) \rangle + \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

Таким образом, L_2 оказывается не просто прямой суммой, а ортогональным разложением на подпространства констант и случайных величин с нулевым матожиданием. Можно рассматривать вырожденное «скалярное произведение», взяв только его компоненту на пространстве центрированных случайных величин. Такое скалярное произведение называется **ковариацией**.

Определение 1.32: Ковариация

Ковариация случайных величин X, Y , принадлежащих L_2 , определяется как

$$\text{cov}(X, Y) := \langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle = \langle C(X), C(Y) \rangle \quad (8)$$

Из билинейности скалярного произведения следует также, что

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (9)$$

и, в частности,

$$\text{cov}(X, X) = \mathbb{D}[X] \quad (10)$$

Из этой формулы следует, что если пара случайных величин X, Y независима, то $\text{cov}(X, Y) = 0$. Действительно, в таком случае $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Если в пространстве задано скалярное произведение, то определён (при условии выполнения неравенства Коши-Буняковского-Шварца) угол между двумя векторами. Обычно определяется угол между центрированными случайными величинами. Косинус этого угла называется их корреляцией:

Определение 1.33: Корреляция случайных величин

Корреляцией вещественнозначных случайных величин $X, Y \in L_2$ называется величина

$$\text{cor}(X, Y) := \frac{\langle C(X), C(Y) \rangle}{\sqrt{\mathbb{D}[X]\mathbb{D}[Y]}} \quad (11)$$

Выполнение неравенства Коши-Буняковского-Шварца для L_2 -нормы и определённого выше скалярного произведения следует из того факта,

что по определению

$$\sqrt{\langle X, X \rangle} := \|X\|_2 \quad (12)$$

Таким образом, как известно из линейной алгебры, верна

Теорема 1.21: Неравенство КБШ

Для случайных величин $X, Y \in L_2$ выполняется

$$|\langle X, Y \rangle| \leq \|X\|_2 \|Y\|_2 \quad (13)$$

Из неравенства КБШ следует неравенство для ковариации

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y) \quad (14)$$

и корреляции:

$$-1 \leq \text{cor}(X, Y) \leq 1 \quad (15)$$

Очевидным следствием свойств скалярного произведения является

Теорема 1.22: Теорема Пифагора (равенство Парсеваля)

Пусть $\{X_i\}_{i=1}^n$ - набор попарно ортогональных случайных величин из L_2 . Тогда

$$\left\| \sum_{i=1}^n X_i \right\|_2^2 = \sum_{i=1}^n \|X_i\|_2^2$$

Структура скалярного произведения также позволяет определить проекции случайных величин на подпространства. Из линейной алгебры известно, что проекция вектора x на подпространство евклидова пространства, порождённое вектором y , даётся выражением

$$\pi_y(x) = \frac{\langle x, y \rangle}{\langle y, y \rangle} y \quad (16)$$

Аналогично можно определить проекцию случайной величины X на подпространство, порождённое случайной величиной Y :

$$\pi_Y(X) := \frac{\langle X, Y \rangle}{\mathbb{D}[Y]} Y \quad (17)$$

Проекция обладает тем свойством, что это наиболее близкий к X по L_2 -норме вектор в линейной оболочке Y . Аналогично и с сохранением

этого свойства, можно спроектировать случайную величину на произвольное подпространство случайных величин.

Определение 1.34: Матрица ковариации

Матрица Грама $\langle C(X_i), C(X_j) \rangle$ набора центрированных случайных величин $\{X_i\}_{i=1}^n$ называется **матрицей ковариации** этого набора.

Алгебра случайных величин

Определим теперь алгебру случайных величин со сложением во втором смысле.

Определение 1.35: Алгебра случайных величин

Пусть $\mathcal{A} = \{a_i\}$ - алфавит формальных символов. Пусть для каждого формального символа задано вероятностное пространство $(\Omega_{a_i}, \sigma_{a_i}, P_{a_i})$. Рассмотрим свободную коммутативную алгебру, порождённую \mathcal{A} .

Теорема 1.23: Структура алгебры случайных величин

Множество всех случайных величин на данном вероятностном пространстве образует линейное пространство (и, более того, бесконечно-порождённую ассоциативную алгебру) с операцией сложения, определённой следующим образом: если $f_1 : \Omega \rightarrow \mathbb{R}_1^n$, $f_2 : \Omega \rightarrow \mathbb{R}_2^n$, то $f_1 + f_2 : \Omega \rightarrow \mathbb{R}_3^n$ определено

$$f_1 + f_2 := + \circ (f_1, f_2) \quad (18)$$

1.1.6 Типы сходимости

Наличие метрики на пространстве распределений позволяет определить сходимость. Существует несколько типов сходимости случайных величин, связанных друг с другом нетривиальным образом.

Сходимость в среднем

L_p -метрики задают на пространстве распределений L_p сходимость в p -среднем, которая определяется стандартным для нормированных про-

странств образом: сходящимися называются последовательности, удовлетворяющие критерию Коши из матанализа.

Определение 1.36: Сходимость в p -среднем

Последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$, заданных на вероятностном пространстве (Ω, σ, P) , снабжённом L_p -нормой $\|\cdot\|_p$, называется сходящейся, если

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall n, m > N : \|X_n - X_m\|_p < \varepsilon \quad (19)$$

Определение 1.37: Сходимость к случайной величине в p -среднем

Говорят, что последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ сходится к случайной величине X в p -среднем, если

$$\lim_{n \rightarrow \infty} \rho_p(X_n, X) = 0 \quad (20)$$

Сходимость по L_1 -норме называется **сходимостью в среднем**, сходимость по L_2 -норме - **сходимостью в среднеквадратичном**.

Из неравенства Ляпунова следует, что сходимость в L_p влечёт сходимость в L_q при $q \leq p$. Больше того, из функционального анализа известно, что L_p - банаховы пространства, а значит, любая сходящаяся в L_p -норме случайная величина имеет предел с конечной L_p -нормой.

Сходимость почти наверное

Определение 1.38: Сходимость почти наверное

Последовательность случайных величин X_n называется **сходящейся почти наверное** к случайной величине X , если

$$P \left(\lim_{n \rightarrow \infty} X_n = X \right) = 1 \quad (21)$$

Сходимость почти наверное обозначается

$$X_n \xrightarrow{\text{a.s.}} X$$

(от английского almost surely).

Сходимость по вероятности

Определение 1.39: Сходимость по вероятности (по мере)

Пусть f - измеримая функция на пространстве с мерой (X, Σ, μ) , $\{f_n\}_{n=1}^{\infty}$ - последовательность таких функций. Говорят, что последовательность $\{f_n\}_{n=1}^{\infty}$ сходится к f по вероятности, если

$$\lim_{n \rightarrow \infty} \mu(\{x \in X : |f_n - f| \geq \varepsilon\}) = 0 \quad (22)$$

Теорема 1.24: Сходимость почти наверное сильнее сходимости по вероятности

...но в законе больших чисел не участвует. Говоря более точно, сходимость почти наверное влечёт сходимость по вероятности, однако обратное в общем случае неверно.

Теорема 1.25: Сходимость в среднем сильнее сходимости по вероятности

Пусть $\{X_n\}_{n=1}^{\infty}$ сходится в среднем к X . Тогда $\{X_n\}_{n=1}^{\infty}$ сходится по вероятности к X , однако обратное в общем случае неверно.

Доказательство. Используем неравенство Маркова (1.1.4). Пусть $\varepsilon > 0$. Тогда

$$0 \leq P(|X_n - X| > \varepsilon) \leq \mathbb{E}[|X_n - X|]/\varepsilon \rightarrow 0$$

при стремлении n к бесконечности. \square

Сходимость по распределению

Определение 1.40: Сходимость по распределению

Пусть (Ω, σ, P) - вероятностное пространство. Пусть $\{X_n\}_{n=1}^{\infty}$ - последовательность случайных величин $X_n : \Omega \rightarrow \mathbb{R}^m$. Говорят, что последовательность $\{X_n\}_{n=1}^{\infty}$ сходится по распределению к случайной величине X , если для всякого открытого множества $A \subseteq \mathbb{R}^n$

$$\lim_{n \rightarrow \infty} P(X_n \in A) = P(X \in A) \quad (23)$$

Сходимость по распределению - самый слабый тип сходимости из используемых нами. Из неё не следуют никакие другие сходимости, а она, в свою очередь, следует из остальных. Иными словами, верна следующая

Теорема 1.26: Сходимость по вероятности сильнее сходимости по распределению

Сходимость по вероятности влечёт за собой сходимость по распределению.

Таким образом, верны следующие взаимоотношения между разными типами сходимости:

- Сходимость в p -среднем и сходимость почти наверное - самые сильные типы сходимости. Сходимость в p -среднем влечёт сходимость в q -среднем при $p \geq q$, но ни для какого p сходимость в p -среднем не влечёт сходимость почти наверное, и наоборот.
- Сходимость по вероятности слабее и сходимости в среднем, и сходимости почти наверное. Для равномерно интегрируемых распределений сходимость по вероятности эквивалентна сходимости в среднем.
- Сходимость по распределению слабее сходимости по вероятности.

Список литературы

- [1] Terrence Tao, *Topics in Random Matrix Theory*. AMS, 2012, глава 1.
- [2] Joseph P.S. Kung, Gian-Carlo Rota, Catherine H. Yan, *Combinatorics: The Rota Way*. Cambridge University Press, 2009, глава 4.
- [3] G.-C. Rota, B.D. Taylor, *The Classical Umbral Calculus*. SIAM J. Math. Anal. Vol. 25, No. 2, pp. 694-711, March 1994.
- [4] Kyle Siegrist, Random: Probability, Mathematical Statistics, Stochastic Processes, <https://www.randomservices.org/random/index.html>

2 Оценивание

2.1 Лекция 1

2.1.1 Основные определения

Определение 2.1: Выборка из распределения

Выборкой X размера n из распределения μ называется набор независимых случайных величин $\{X_i\}_{i=1}^n$ со значениями в \mathcal{X} , каждая из которых имеет распределение μ .

Выборкой называют также конкретную реализацию такого набора случайных величин. Различие обычно ясно из контекста.

Определение 2.2: Статистика

Семейство μ -измеримых функций $f : \mathcal{X}^{\times n} \rightarrow \mathbb{R}$ такое, что для каждого натурального n в семействе есть одна и только одна функция, называется статистикой.

Будем обозначать \mathcal{P} семейство распределений $\{\mu_\theta\}_{\theta \in \Theta}$, где Θ - некоторое множество.

Определение 2.3: Генеральная совокупность

Генеральной совокупностью называется множество всех объектов, о которых мы делаем некоторые статистические выводы.

Генеральная совокупность - не столько математическое, сколько «физическое» понятие. Обычно мы моделируем генеральную совокупность при помощи некоторого распределения или семейства распределений. Например, мы хотим оценить усреднённые характеристики всех потенциальных клиентов компании. Множество всех потенциальных клиентов - генеральная совокупность. Генеральная совокупность, как правило, недоступна для наблюдения целиком (иначе все интересующие нас характеристики можно было бы вычислить в явном виде, не прибегая к статистике). Поэтому мы пытаемся моделировать множество всех потенциальных клиентов при помощи некоторого семейства распределений на множестве возможных характеристик потенциальных клиентов.

Определение 2.4: Выборка из генеральной совокупности

Выборкой из генеральной совокупности называется любое подмножество генеральной совокупности. Альтернативно, выборкой называется случайная величина со значениями во множестве подмножеств генеральной совокупности.

Иными словами, выборку из генеральной совокупности можно рассматривать как результат случайного процесса, выбирающего некоторые элементы из генеральной совокупности.

Определение 2.5: Эмпирическая функция распределения

Пусть $\{X_i\}_{i=1}^n$ - реализация выборки из некоторого распределения, т.е. элемент $\mathbb{R}^{\times n}$. Тогда функция

$$F(x) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^x \delta(x' - X_i) dx' = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x) \quad (24)$$

называется **эмпирической функцией распределения** для выборки $\{X_i\}_{i=1}^n$. Здесь $\delta(x)$ - дельта-функция Дирака, $\mathbb{I}(X_i \leq x)$ - индикаторная функция луча $[X_i, +\infty)$.

Аналогичным образом можно определить эмпирическую функцию распределения для выборки как последовательности случайных величин. В таком случае эмпирическая функция распределения сама будет случайной величиной со значением в функциональном пространстве.

Свойства эмпирической функции распределения

1. Монотонность: ЭФР не убывает.
2. Границные условия: $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$.
3. Непрерывность справа: $\forall x_0 \in \mathbb{R} : \lim_{x \rightarrow +x_0} F(x) = F(x_0)$. При этом функция очевидным образом не является непрерывной ни при каком конечном (и даже счётном) размере выборки.

Теорема 2.1: Теорема Гливенко-Кантелли

Пусть $\{X_i\}_{i=1}^{\infty}$ - последовательность независимых одинаково распределённых \mathbb{R} -значных случайных величин с функцией распределения $F(x)$. Пусть $F_n(x)$ - эмпирическая функция распределения подпоследовательности $\{X_i\}_{i=1}^n$. Тогда последовательность случайных величин

$$D_n = \sup_x |F_n(x) - F(x)|$$

сходится к 0 почти наверное.

Замечание 8. Напомним, что из сходимости почти наверное следует сходимость по вероятности.

Определение 2.6: Эмпирическая характеристическая функция

Эмпирической характеристической функцией выборки $\{X_i\}_{i=1}^n$ называется функция

$$\chi_n(t) := \frac{1}{n} \sum_{i=1}^n e^{itX_i} \quad (25)$$

Теорема 2.2: Теорема Гливенко-Кантелли для характеристической функции

Пусть $\{X_i\}_{i=1}^{\infty}$ - последовательность независимых одинаково распределённых случайных величин с характеристической функцией $\chi(t)$. Тогда последовательность $\{\chi_n(t)\}_{n=1}^{\infty}$ эмпирических характеристических функций $\{X_i\}_{i=1}^n$ сходится к $\chi(t)$ почти наверное.

2.1.2 Задача точечной оценки параметра

Пусть $\mathcal{P} = \{P_{\theta}, \theta \in \Theta\}$ - семейство распределений. Пусть дана выборка $\{X_i\}_{i=1}^n$, про которую известно, что распределение, её порождающее, взято из семейства \mathcal{P} , но не известно значение θ . Задача оценки параметра сводится к получению информации о значении θ из $\{X_i\}_{i=1}^n$. Одним из способов получить информацию о значении θ является вычисление некоторой **точечной оценки** для θ , то есть такой статистики, которая при некоторых предположениях и в некотором вероятностном смысле сходится к θ с ростом n , или к некоторой функции этой статистики $\tau(\theta)$.

Определение 2.7: Несмешённая оценка

Точечная оценка $\hat{\theta}_n(\{X\}_{i=1}^n)$ параметра θ называется несмешённой, если

$$\forall n : \mathbb{E}[\hat{\theta}_n] = \theta \quad (26)$$

Определение 2.8: Выборочное среднее

Выборочным средним называется статистика

$$M(\{X_i\}_{i=1}^n) := \frac{1}{n} \sum_{i=1}^n X_i \quad (27)$$

Пример 4. Для выборки из любого распределения, имеющего матожидание, выборочное среднее является несмешённой оценкой математического ожидания, что немедленно следует из линейности последнего.

Определение 2.9: Риск

Риском для выборки X и оценки τ называется статистика

$$R(X) := \mathbb{E}[(\hat{\theta}(X) - \theta)^2] \quad (28)$$

Теорема 2.3: Риск несмешённой оценки

Для несмешённой оценки $\hat{\theta}$ её риск равен дисперсии случайной величины, из которой берётся выборка.

Доказательство. Пусть $\hat{\theta}$ - несмешённая оценка θ . Тогда

$$\begin{aligned} \mathbb{E}[R(X)] &= \mathbb{E}[(\hat{\theta}(X) - \theta)^2] = \\ &= \mathbb{E}[(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)] + \mathbb{E}[\hat{\theta}(X)] - \theta)^2] = \\ &= \mathbb{E}[(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)])^2 + (\mathbb{E}[\hat{\theta}(X)] - \theta)^2 + 2(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)])(\mathbb{E}[\hat{\theta}(X)] - \theta)] \end{aligned}$$

В силу несмешённости $\hat{\theta}$ имеем $\mathbb{E}[\hat{\theta}(X)] = \theta$, следовательно, второе и третье слагаемое в последней строчке зануляются. Остаётся

$$\mathbb{E}[(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)])^2]$$

а это по определению дисперсия $\hat{\theta}(X)$. □

Так называемый **bias-variance tradeoff** в машинном обучении - это способ интерпретации формулы для матожидания риска

$$\begin{aligned}\mathbb{E}[R(X)] &= \mathbb{E}[(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)])^2 + \\ &\quad (\mathbb{E}[\hat{\theta}(X)] - \theta)^2 + \\ &\quad 2(\hat{\theta}(X) - \mathbb{E}[\hat{\theta}(X)])(\mathbb{E}[\hat{\theta}(X)] - \theta)]\end{aligned}$$

в которой мы можем изменять характеристики $\hat{\theta}(X)$. Если оценка несмещённая (unbiased), вклад второго и третьего слагаемого, как мы видели выше, нулевой. Путём добавления в оценку смещения (bias) можно попытаться уменьшить дисперсию (variance), но в таком случае станут ненулевыми члены, зависящие от смещения.

2.1.3 Эффективные оценки

Пусть выбран некоторый класс оценок $C = \{\hat{\theta}_\nu(X)\}_{\nu \in \Lambda}$ (например, несмешённых) параметра θ для семейства распределений \mathcal{P} . При прочих равных мы хотим, чтобы дисперсия нашей оценки была минимальной, поэтому разумно использовать величину дисперсии в качестве критерия качества оценки.

Определение 2.10: Эффективная оценка

Эффективной оценкой $\hat{\theta}$ в классе C называется такая оценка, что для всякой другой оценки $\hat{\theta}'$ из класса C и для всякого размера n выборки X

$$\mathbb{E}[R(\hat{\theta}(X))] \leq \mathbb{E}[R(\hat{\theta}'(X))]$$

Определение 2.11: Информация Фишера

Пусть $\mathcal{P} = \{p_\theta\}_{\theta \in \mathbb{R}}$ - однопараметрическое семейство регулярных распределений. Информацией Фишера, содержащейся в выборке X из распределения из этого семейства, называется функция

$$\mathcal{I}_X(\theta) := \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln p_\theta(X) \right)^2 \right] = \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta} \ln p_\theta(X) \right)^2 p_\theta(X) dX \quad (29)$$

Теорема 2.4: Неравенство Рао-Крамера

Пусть $\hat{\theta}(X)$ - несмешённая оценка функции $\tau(\theta)$ параметра θ семейства регулярных распределений $\mathcal{P} = \{p_\theta\}$ со значениями в \mathbb{R} , и, в частности, $\hat{\theta}(x)$ - эта оценка для случая выборки из одного элемента. Пусть $\theta \in \mathbb{R}$. Пусть конечны для всякого размера выборки $\mathbb{E}[\hat{\theta}(X)]$ и $\mathbb{D}[\hat{\theta}(X)]$. Пусть также выполнены следующие *условия регулярности*:

1. Носитель распределения $p_\theta(x) \in \mathcal{P}$ не зависит от θ .
2. $\forall \theta \in \mathbb{R} : \forall X \in \text{supp}_x(p_\theta)^{\times n} : \exists \frac{\partial}{\partial \theta} \ln p_\theta(X) := \lambda(X, \theta)$
3. $\frac{\partial}{\partial \theta} \int p_\theta(X)dX = \int \frac{\partial}{\partial \theta} p_\theta(X)dX = 0$
4. $\frac{\partial}{\partial \theta} \int \hat{\theta}(X)p_\theta(X)dX = \int \hat{\theta}(X) \frac{\partial}{\partial \theta} p_\theta(X)dX$
5. $0 < \mathcal{I}(\theta) < \infty$

В таком случае выполняется следующее неравенство:

$$\mathbb{D}[\hat{\theta}(X)] \geq \frac{\left(\frac{\partial}{\partial \theta} \tau(\theta)\right)^2}{\mathcal{I}(\theta)} \quad (30)$$

Замечание 9. Если $\tau(\theta) = \theta$, то

$$\mathbb{D}[\hat{\theta}(X)] \geq \frac{1}{\mathcal{I}(\theta)} \quad (31)$$

Замечание 10.

$$\mathcal{I}_X(\theta) = \mathbb{E}[\lambda^2(X, \theta)]$$

Доказательство. Из условия регулярности 3. имеем

$$0 = \int_{\mathbb{R}} \frac{\partial}{\partial \theta} p_\theta(X)dX = \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta} \ln p_\theta(X) \right) p_\theta(X)dX = \mathbb{E}[\lambda(X, \theta)]$$

где $\lambda(X, \theta)$ - логарифмическая производная плотности $p_\theta(X)$ в точке X ,

$$\lambda(X, \theta) = \frac{\partial}{\partial \theta} \ln p(X, \theta)$$

Из условия регулярности 4 имеем

$$\frac{\partial \tau(\theta)}{\partial \theta} = \int_{\mathbb{R}} \hat{\theta}(X) \left(\frac{\partial}{\partial \theta} \ln p_{\theta}(X) \right) p_{\theta}(X) dx = \mathbb{E}[\hat{\theta}(X) \lambda(X, \theta)]$$

Вычтем из последнего выражения 0, представив его как

$$\mathbb{E}[\tau(\theta) \lambda(X, \theta)] = \tau(\theta) \mathbb{E}[\lambda(X, \theta)] = 0$$

используя условие регулярности 3 и тот факт, что $\tau(\theta)$ - не случайная величина. Получим

$$\mathbb{E}[\hat{\theta}(X) \lambda(X, \theta)] - \mathbb{E}[\tau(\theta) \lambda(X, \theta)] = \mathbb{E}[(\hat{\theta}(X) - \tau(\theta)) \lambda(X, \theta)]$$

Таким образом,

$$\frac{\partial \tau(\theta)}{\partial \theta} = \mathbb{E}[(\hat{\theta}(X) - \tau(\theta)) \lambda(X, \theta)]$$

Используем неравенство Коши-Буняковского-Шварца. Напомним, оно заключается в том, что для любого евклидова пространства, т.е. линейного пространства с невырожденной положительно определённой симметричной билинейной формой (скалярным произведением) $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ выполняется

$$\forall a, b \in V : |\langle a, b \rangle| \leq \sqrt{\langle a, a \rangle \langle b, b \rangle}$$

Случайные величины на одном и том же вероятностном пространстве образуют линейное пространство, в котором роль скалярного произведения играет функционал $X, Y \rightarrow \mathbb{E}[XY]$. Фактор-пространство этого пространства по подпространству постоянных случайных величин имеет в качестве скалярного произведения, унаследованного из исходного, ковариацию. В данном случае мы используем исходное скалярное произведение, из которого получаем

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]$$

Применяя его к $\frac{\partial \tau(\theta)}{\partial \theta}$, получаем

$$\begin{aligned} \left(\frac{\partial \tau(\theta)}{\partial \theta} \right)^2 &= \left(\mathbb{E}[(\hat{\theta}(X) - \tau(\theta)) \lambda(X, \theta)] \right)^2 \leq \\ &\leq \mathbb{E}[(\hat{\theta}(X) - \tau(\theta))^2] \mathbb{E}[\lambda(X, \theta)^2] \end{aligned}$$

Поскольку оценка $\hat{\theta}(X)$ несмешённая,

$$\mathbb{E}[(\hat{\theta}(X) - \tau(\theta))^2] = \mathbb{D}[\hat{\theta}(X)]$$

В то же время

$$\mathbb{E}[\lambda(X, \theta)^2] = \mathcal{I}(\theta)$$

Таким образом,

$$\mathbb{D}[\hat{\theta}(X)] \geq \frac{\left(\frac{\partial \tau(\theta)}{\partial \theta}\right)^2}{\mathcal{I}_X(\theta)}$$

что и требовалось. \square

Определение 2.12: Оптимальная оценка

Оценка $\hat{\theta}(X)$ называется **оптимальной**, если для неё в неравенстве Рао-Крамера достигается равенство.

Определение 2.13: Информация Фишера в выборке из одного распределения

$\mathcal{I}_1(\theta)$ обозначается информация, содержащаяся в выборке из данного распределения из одного элемента с параметром θ .

Теорема 2.5: Аддитивность информации Фишера

Если выборка состоит из независимых наблюдений, информация Фишера во всей выборке равна сумме информаций Фишера в каждом из её элементов.

Из аддитивности информации Фишера следует, что для выборки размера n

$$\mathbb{D}[\hat{\theta}(X)] \geq \frac{(\partial \tau(\theta)/\partial \theta)^2}{n \mathcal{I}_1(\theta)} \propto \frac{1}{n}$$

Таким образом, нижняя граница дисперсии обратно пропорциональна размеру выборки, а для оптимальной оценки это также и верхняя граница.

2.1.4 Экспоненциальное семейство распределений

Неравенство КБШ, которое мы использовали в доказательстве неравенства Рао-Крамера, обращается в равенство только при условии пропорциональности входящих в него векторов. Векторы, которые мы подставляли в неравенство - $\lambda(X, \theta)$ и $\hat{\theta}(X) - \tau(\theta)$. Предположим, эти векторы пропорциональны с коэффициентом пропорциональности a , возможно, зависящим от θ .

Тогда

$$\lambda(X, \theta) = \frac{\partial}{\partial \theta} \ln p_\theta(X) = \frac{\hat{\theta}(X)}{a(\theta)} - \frac{\tau(\theta)}{a(\theta)}$$

Проинтегрируем вторую и третью части равенства по θ . Получим

$$\ln p_\theta(X) = c(\theta)\hat{\theta}(X) + d(\theta) + S(X) \quad (32)$$

где

$$c(\theta) = \int \frac{1}{a(\theta)} d\theta \quad d(\theta) = - \int \frac{\tau(\theta)}{a(\theta)} d\theta$$

а $S(X)$ возникает из константы интегрирования. Экспоненцируя, получаем

Определение 2.14: Экспоненциальное семейство распределений

Распределения с плотностью вида

$$p_\theta(X) = \exp \left(c(\theta)\hat{\theta}(X) + d(\theta) + S(X) \right)$$

называются распределениями из экспоненциального семейства.

Теорема 2.6: Оптимальная статистика для экспоненциального семейства

Статистика $\hat{\theta}(X)$, участвующая в определении распределения из экспоненциального семейства, является оптимальной для соответствующего распределения.

2.2 Лекция 2

2.2.1 Достаточные статистики

Определение 2.15: Достаточная статистика

Статистика $\hat{\theta}(X)$ называется достаточной для параметра θ в случае, если условное распределение X при некотором фиксированном значении $\hat{\theta}(X)$ одно и то же для всех $\theta \in \Theta$.

Пример 5. Рассмотрим статистику выборочного среднего:

$$\hat{\theta}(X) = \frac{\sum_{i=1}^n X_i}{n} = t$$

3 Тестирование гипотез