Конспект лекций по теории вероятностей Б.Б. Демешева

ПМИ ФКН ВШЭ, 2 курс, основной поток
2024-2025

Составители:

Смирнов Г.А.

Хромотов А.А.

Маркотенко А.Е.

Антипович В.О.

Жемуков А.А.

Исходный код

Содержание

Лекция 1. Основные понятия

2 сентября 2024 г.

1.1 Событие

Определение: Ω – множество всех исходов эксперимента.

Пример: Подбрасывание монетки до первого орла:

$$\Omega = \{T, HT, HHT, HHHT, \ldots\}$$

Орёл – T *(tail)*
Решка – H *(head)*

Определение: Событие A – подмножество Ω.

Обозначение: события традиционно обозначаются буквами A, B, C, D, и т.д.

Обозначение:

$$A^c = \overline{A} = \Omega \setminus A = \text{«не } A\text{»}$$

Поскольку обозначение $\overline{A}$ сильно перегружено, предпочтительнее использование обозначения A^c.

Пример: «Орёл не выпал при первом броске» = $\{HT, HHT, \ldots\}$

1.2 Вероятность и её свойства

Обозначение: $\mathbb{P}(A)$ – вероятность A.

Аксиомы:

1. $\mathbb{P}(\Omega) = 1$
2. $\forall A : \mathbb{P}(A) + \mathbb{P}(A^c) = 1$
3. Если $\forall i \neq j : A_i \cap A_j = \varnothing$, то $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Следствия:
1. $\mathbb{P}(A) \in [0; 1]$
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
3. $\mathbb{P}\left(\bigcup_{i=0}^{\infty} B_i\right) \leq \sum_{i=0}^{\infty} \mathbb{P}(B_i)$

1.3 Случайная величина и математическое ожидание

Определение: X – случайная величина $\overset{\text{def}}{\iff} X : \Omega \to \mathbb{R}$.

Иногда подразумевается $X : \Omega \to \mathbb{R} \cup \{\infty\}$.

Обозначение: Случайные величины традиционно обозначают как попало в зависимости от традиции, области применения и т.д.

Пример: Рассмотрим эксперимент, в котором подбрасывается монетка, пока не выпадет орёл. Пусть X – количество выпавших при этом решек. Тогда:

$$X(T) = 0$$

$$X(HHHT) = 3$$

$$X(HHHHH...) = \infty$$

Определение: Медиана случайной величины X – это любое число m, удовлетворяющее условию $P(X \geq m) \geq \frac{1}{2}$ и $P(X \leq m) \geq \frac{1}{2}$.

Пример:

t	1	2	3
$P(X = t)$	0.1	0.7	0.2

Медиана – 2.

Стоит отметить, что в различной литературе вы можете столкнуться с тем, что медианой случайной величины считают самое большое или самое маленькое подходящее число, но в рамках курса мы будем использовать отрезок.

Обозначение: $\mathbb{E}(X)$ – математическое ожидание.

Строгое определение математического ожидания будет в конце курса. Пока что попробуем дать определение для дискретных случаев и будем полагаться на интуицию в остальных.

Частичное определение. Пусть у величины X счётное или конечное множество значений у X.

Тогда

$$\mathbb{E}(X) = \sum_x x * \mathbb{P}(X = x),$$

предполагая, что $\infty * 0 = 0$.

Строго говоря, это означает, что

$$\mathbb{E}(X) = \sum_{x \in \operatorname{Range} X} x * \mathbb{P}(X^{-1}(x))$$

Пример: X – значение на брошенном кубике.

$$\mathbb{E}(X) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{7}{2}$$

1.4 Свойства математического ожидания

1. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
2. c – константа, X – случайная величина, тогда $\mathbb{E}(c * X) = c * \mathbb{E}(X)$.

1.5 Интуиция математического ожидания

- **Долгосрочное среднее**, т.е. среднее значение при бесконечно большом количестве повторений эксперимента.
- **Центр масс** прямой, если в токах, соответствующих значениям случайной величины, поместить грузики с массами, равными вероятностям этих значений.

1.6 Перестановочный тест

Пусть есть две гипотезы:

- H_0: нулевая гипотеза, «статус-кво».
- H_1: альтернативная гипотеза, «так жить больше нельзя».

Мы хотим понять, какая их этих гипотез более правдоподобна.

Будем объяснять на примере.

Кафе «Груша» рассматривает возможность не просто говорить студентам «Доброе утро!», но и улыбаться.

- H_0 – цена покупки не зависит от того, улыбаются ли студентам.
- H_1 – если улыбаться студентам, то они будут покупать больше.

Проводится эксперимент – в части случаев студентам говорят «Доброе утро!» как раньше, а в части кроме этого ещё и улыбаются.

y_i (цена покупки студента)	x_i (используемая гипотеза)
300	A
200	A
400	B
300	B
200	B

A – «Доброе утро!»
B – «Доброе утро!» + улыбка

S – целевой показатель, насколько больше студенты покупают, если им улыбаться.

$$S = \overline{y}_B - \overline{y}_A, \quad \overline{y} - \text{среднее по вектору.}$$

$$S = 300 - 230 = 50$$

Т.е. в среднем студенты платили на 50 больше, если им улыбаться.

Уточним определения для общего случая.

- Гипотеза H_0: x_i и y_i независимы
- Гипотеза H_1: $\mathbb{E}(y_B) > \mathbb{E}(y_A)$

Для того, чтоб интерпретировать результаты эксперимента, используется *p-значение (p-value)*:

p-значение = $\mathbb{P}(S^{\text{new}} \geq S \mid S, H_0)$

То есть p-значение – это вероятность того, что целевой показатель мог случайно оказаться таким или больше, каким оказался, если на самом деле верна гипотеза H_0.

В примере: $\mathbb{P}(S^{\text{new}} \geq 50 \mid H_0) = \frac{5}{C_5^2} = \frac{1}{2}$.

Чтоб решить, отвергать ли гипотезу H_0, p-значение сравнивают с заранее выбранной величиной α – *порогом редкости*.

Если p-value < α, то H_0 отвергается. Если p-value > α, то H_0 не отвергается.

Лекция 2. Методы решения задач

9 сентября 2024 г.

2.1 Разложение в сумму

Задача. У Маши 27 пар туфель. Пёс Шарик утащил из них случайно выбранные 13 туфель.

X – количество пар туфель, доступных Маше для прогулки, т.е. пар туфель, из которых Шарик не утащил ни одну.

Требуется зайти:

1. $\mathbb{E}(X)$
2. $\mathbb{P}(X = 16)$

Решение: Считать математическое ожидание по определению в данном случае не очень удобно, поэтому поэтому мы применим следующий приём.

Обозначим за y_i величину, равную 1, если i-я пара доступна для прогулки, и 0 иначе. Тогда $\mathbb{P}(y_i = 1)$ – вероятность того, что Шарик не съел ни одну туфлю из i-ой пары, и равна $\frac{\binom{25}{13}}{\binom{27}{13}}$, так как всего возможных равновероятных исходов существует $\binom{54}{13}$, а $\binom{54}{13}$ из них благоприятные. То есть

$$\mathbb{P}(y_i = 1) = \frac{\binom{52}{13}}{\binom{54}{13}} = \frac{820}{1431}$$

Тогда вследствие линейности математического ожидания получаем:

$$\begin{aligned}\mathbb{E}(X) = \mathbb{E}(y_1 + \ldots + y_{27}) = \mathbb{E}(y_1) + \ldots + \mathbb{E}(y_{27}) = 27 * (1 * \mathbb{P}(y_i = 1) + 0 * \mathbb{P}(y_i = 0)) = \\ = 27 * \frac{820}{1431} = \frac{820}{53}\end{aligned}$$

Теперь найдём $\mathbb{P}(X = 16)$. Снова воспользуемся тем, что $X = y_1 + \ldots + y_{27}$.

Пусть туфли с номерами s_1, ..., s_{16} доступны Маше для прогулки, а туфли с номерами s_{17}, ..., s_{27} – нет. Вероятность такого события равна

$$\begin{aligned}\mathbb{P}\Big(A\Big(\{s_i\}_{i=1}^{16}\Big)\Big) &= \mathbb{P}\Big(y_{s_1} = 1\Big) * \ldots * \mathbb{P}\Big(y_{s_{16}} = 1\Big) * \mathbb{P}\Big(y_{s_{17}} = 0\Big) * \ldots * \mathbb{P}\Big(y_{s_{27}} = 0\Big) = \\ &= \mathbb{P}(y_i = 1)^{16} * \mathbb{P}(y_i = 0)^{11} = \mathbb{P}(y_i = 1)^{16} * (1 - \mathbb{P}(y_i = 1))^{11} = \\ &= \left(\frac{820}{1431}\right)^{16} * \left(1 - \frac{820}{1431}\right)^{11}\end{aligned}$$

Но это вероятность для какого-то одного набора $s_1, \ldots, s_{16}$, а всего таких наборов $\binom{27}{16}$. Так как такие события взаимоисключают друг друга, получаем, что

$$\binom{27}{16} * \mathbb{P}\Big(A\Big(\{s_i\}_{i=1}^{16}\Big)\Big) = \binom{27}{16} * \left(\frac{820}{1431}\right)^{16} * \left(1 - \frac{820}{1431}\right)^{11}$$

Великодушный читатель наверняка простит нам то, что вычислять значение этого выражения мы не будем.

Как можно было увидеть на этом примере, суть метода разложения в сумму заключается в том, чтоб разложить сложную величину в сумму простых, для которых всё нужное нам вычисляется проще.

2.2 Метод первого шага

Задача. Алиса и Боб подбрасывают правильную монетку неограниченное число раз.

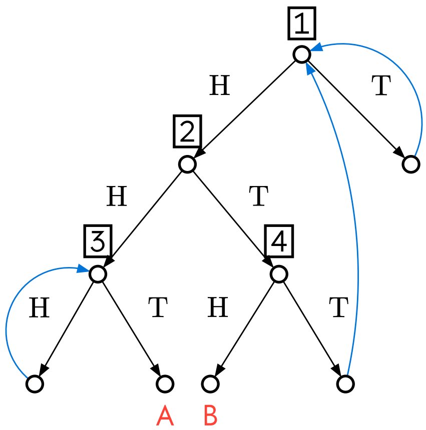
- Алиса выигрывает, если HHT выпадет раньше.
- Боб выигрывает, если HTH выпадет раньше.

Обозначим за A событие победы Алисы, а за B – событие победы Боба.

Требуется найти:

1. $\mathbb{P}(A)$
2. $\mathbb{P}(B)$
2. $\mathbb{E}(N)$, где N – продолжительность игры

Решение: Построим конечный автомат игры.



Синие рёбра ведут в эквивалентное состояние выше в графе.

Теперь выразим вероятности победы Алисы при старте из пронумерованных вершин. Обозначим вероятность победы Алисы при старте в вершине i как a_i.

$$\begin{cases} a_1 = \frac{1}{2} * a_1 + \frac{1}{2} * a_2 \\ a_2 = \frac{1}{2} * a_3 + \frac{1}{2} * a_4 \\ a_3 = \frac{1}{2} * a_3 + \frac{1}{2} * 1 \\ a_4 = \frac{1}{2} * 0 \ + \frac{1}{2} * a_1 \end{cases}$$

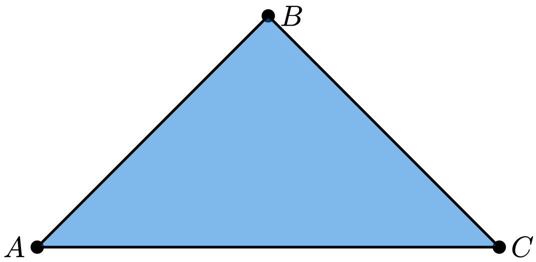
Эту систему линейных уравнений можно без труда решить. Заметим, что:

- $\mathbb{P}(A) = a_1$
- $\mathbb{P}(B) = 1 - \mathbb{P}(A) = 1 - a_1$

Не доделано

2.3 Респект Архимеду

Задача.




```
/* Опять не успел, зато чуть-чуть разобрался с cetz */
```

Не доделано

Лекция 3. Условные вероятности и ожидания

16 сентября 2024 г.

Идея: новая информация приводит к обновлению вероятностей.

Интуитивный пример:

x	1	2	3
$\mathbb{P}(X=x)$	0.2	0.2	0.6
$\mathbb{P}(X=x \mid \text{Info})$	0	0.25	0.75

$$\text{Info} = \{X > 1\}$$

Интуитивный алгоритм действия:

1. Занулить невозможножные вероятности.
2. Отмасштабировать оставшиеся.

Обозначение: $\mathbb{P}(A \mid B)$ – «вероятность A при условии B».

Определение (формальное):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$$

Следствия:

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) * \mathbb{P}(B)$
2. *Формула полной вероятности:*

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$$
$$\mathbb{P}(A) = \mathbb{P}(A \mid B) * \mathbb{P}(B) + \mathbb{P}(A \mid B^c) * \mathbb{P}(B^c)$$

$$\forall i \neq j : B_i \cap B_j = \varnothing, \ \bigcup_{i=1}^{\infty} B_i = \Omega, \ \forall i : \mathbb{P}(B_i) > 0 \quad \Rightarrow \mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i) * \mathbb{P}(B_i)$$

Все эти формулы называют формулой полной вероятности, причём их смысл по сути совпадает, либо обобщает друг друга.

3. *Формула Байеса:*

> **Важный факт:** Портрет Байеса, который вы найдёте в интернете – это не Байес.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A) * \mathbb{P}(A)}{\mathbb{P}(B \mid A) * \mathbb{P}(A) + \mathbb{P}(B \mid A^c) * \mathbb{P}(A^c)}$$

Определение: I — *индикатор события* $B \overset{\text{def}}{\Leftrightarrow} \quad I(w) = \begin{cases} 1, & w \in B \\ 0, & w \in B^c \end{cases}$

Определение (Условное математическое ожидание):

$$\mathbb{E}(X \mid B) = \frac{\mathbb{E}(X * I_B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$$

Свойства условного математического ожидания:

- **Линейность.**

$$\mathbb{E}(X + Y \mid B) = \mathbb{E}(X \mid B) + \mathbb{E}(Y \mid B)$$
$$\mathbb{E}(c * X \mid B) = c * \mathbb{E}(X \mid B)$$

- $\mathbb{E}(X) = \mathbb{E}(X \mid B) * \mathbb{P}(B) + E(X \mid B^c) * \mathbb{P}(B^c)$

Определение: События A и B называются независимыми, если $\mathbb{P}(A \cap B) = \mathbb{P}(A) * \mathbb{P}(B)$.

Мотивация: такое определение обеспечивает свойство:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) * \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$
$$\mathbb{P}(B \mid A) = \mathbb{P}(B)$$

Пример (1): Рассмотрим бросок правильного шестигранного кубика:

$$\Omega = \{1, 2, ..., 6\}$$

$$\mathbb{P}(i) = \frac{1}{6}$$

$$A = \{5\} = \{X = 5\}, \quad B = \{6\} = \{X = 6\}$$

$$\mathbb{P}(X = 5) = \frac{1}{6}, \quad \mathbb{P}(X = 6) = \frac{1}{6}$$

$$\{X = 5\} \cap \{X = 6\} = \varnothing$$

$$P(\varnothing) = 0 \neq \frac{1}{6} * \frac{1}{6} \quad \Rightarrow A \text{ и } B \text{ – зависимые события}$$

Пример (2): Рассмотрим два последовательных броска шестигранного кубика:

$$\Omega = \{(1, 1), (1, 2), ..., (6, 6)\}$$

$$\mathbb{P}((i, j)) = \frac{1}{36}$$

$$X_1 \text{ — результат первого броска}, \quad X_2 \text{ — результат второго броска}$$

$$A = \{X_1 = 5\}, B = \{X_2 = 6\}$$

$$\mathbb{P}(A) = \frac{1}{6}, \quad \mathbb{P}(B) = \frac{1}{6}$$

$$\mathbb{P}(A \cap B) = \mathbb{P}((5, 6)) = \frac{1}{36} = \frac{1}{6} * \frac{1}{6} \quad \Rightarrow A \text{ и } B \text{ – независимые события}$$

Обозначение: $A \perp B \overset{\text{def}}{\Leftrightarrow}$ «A и B независимы»

Следствие:

$$A \perp B \Leftrightarrow \begin{cases} A \ \perp B^c \\ A^c \perp B \\ A^c \perp B^c \end{cases}$$

Пусть есть набор событий $\mathcal{A}$.

Определение: События набора $\mathcal{A}$ называются *независимыми в совокупности (mutually independent)*, если любое событие из набора независимо с любым пересечением оставшихся событий.

Следствие: Набор событий $\mathcal{A}$ независим в совокупности тогда и только тогда, когда $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_{k_i}\right) = \prod_{i=1}^{\infty} \mathbb{P}\left(A_{k_i}\right)$.

Пример: $\mathcal{A} = \{A_1, A_2, A_3, ...\}, \quad A_2 \perp A_3 \cap A_{17} \cap A_{28} \cap A_{29} \cap ...$

Метатеорема: У любой теоремы (или определения) с двойными индексами есть более понятная переформулировка без них.

Это был странный вброс. Да.

Определение: Случайные величины X и Y *независимы*, если для любых множеств $A \subseteq \mathbb{R}$ и $B \subseteq \mathbb{R}$, на которых определены вероятности $\mathbb{P}(X \in A)$ и $\mathbb{P}(Y \in B)$, события $\{X \in A\}$ и $\{Y \in B\}$ независимы.

Теорема: Случайные величины X и Y независимы, если и только если события $\{X \leq x\}$ и $\{Y \leq y\}$ независимы для любых $x, y \in \mathbb{R}$.

Доказательство очевидно.

Теорема: Дискретные случайные величины X и Y независимы, если и только если события $\{X = x\}$ и $\{Y = y\}$ независимы для любых $x, y \in \mathbb{R}$.

Лекция 4. Производящие функции и распределения

23 сентября 2024 г.

4.1 Немного доказательств вам в лекцию

Напомним:

Определение (с прошлой лекции): Случайные величины X и Y независимы, если $\forall A \subseteq \mathbb{R}$, $\forall B \subseteq \mathbb{R}$ события $\{X \in A\}$ и $\{Y \in B\}$ независимые, если $P(X \in A)$ и $P(Y \in B)$ определены.

Теорема: Дискретные случайные величины X и Y независимы если и только если $\{X = x\}$ и $\{Y = y\}$ независимы $\forall x, y$.

Доказательство: Для доказательства воспользуемся метатеоремой:

Метатеорема (№2): Любое доказательство про дискретные случайные величины получается перестановкой слагаемых.

Пример:

1. «Только если»

Если дискретные случайные величины X и Y независимы, то события $\{X \in A\}$ и $\{Y \in B\}$ независимы даже при $A = \{x\}$ и $B = \{y\}$.

2. «Если»

$$\tilde{A} = A \cap \mathrm{Range}(X), \quad \tilde{B} = B \cap \mathrm{Range}(Y), \quad \left|\tilde{A}\right| = \left|\tilde{B}\right| = |\mathbb{N}|$$

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) = \mathbb{P}\left(X \in \tilde{A}, Y \in \tilde{B}\right) = \sum_{x \in \tilde{A}} \sum_{y \in \tilde{B}} \mathbb{P}(X = x, Y = y) = \\ = \sum_{x \in \tilde{A}} \sum_{y \in \tilde{B}} \mathbb{P}(X = x) * \mathbb{P}(Y = y) = \sum_{x \in \tilde{A}} \mathbb{P}(X = x) * \sum_{y \in \tilde{B}} * \mathbb{P}(Y = y) = \\ = \mathbb{P}\left(X \in \tilde{A}\right) * \mathbb{P}\left(Y \in \tilde{B}\right) = \mathbb{P}(X \in A) * \mathbb{P}(Y \in B)\end{aligned}$$

■

Вспомним, что для любый случайных величин X и Y, таких, что $\mathbb{E}(X)$ и $\mathbb{E}(Y)$ существуют, выполняется равенство:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

💡 *Этот факт выполняется даже для зависимых случайных величин.*

Теорема: Две случайные величины X и Y независимы тогда и только тогда, когда

$$\forall f : \mathbb{R} \to \mathbb{R}, \; g : \mathbb{R} \to \mathbb{R}$$

$$\mathbb{E}(f(X) * g(Y)) = \mathbb{E}(f(X)) * \mathbb{E}(g(Y)), \quad \text{если } \mathbb{E}(f(X)) \text{ и } \mathbb{E}(g(Y)) \text{ существуют}$$

Доказательство: Пусть X и Y — независимые дискретные случайные величины, а функции f и g таковы, что $\mathbb{E}(f(X))$ и $\mathbb{E}(g(Y))$ определены. Тогда $\mathbb{E}(f(X) * g(Y)) = \mathbb{E}(f(X)) * \mathbb{E}(g(Y))$.

$\Rightarrow$ По определению математического ожидания для дискретных случайных величин:

$$\mathbb{E}(f(X) * g(Y)) = \sum_x \sum_y f(x) * g(y) * \mathbb{P}(X = x, Y = y).$$

Из независимости X и Y следует, что:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) * \mathbb{P}(Y = y).$$

Подставляя это в формулу, получаем:

$$\mathbb{E}(f(X) * g(Y)) = \sum_x \sum_y f(x) * g(y) * \mathbb{P}(X = x) * \mathbb{P}(Y = y).$$

Разделим сумму на независимые части:

$$\mathbb{E}(f(X) * g(Y)) = \sum_x f(x) * \mathbb{P}(X = x) * \sum_y g(y) * \mathbb{P}(Y = y).$$

Поскольку выражения в скобках — это $\mathbb{E}(f(X))$ и $\mathbb{E}(g(Y))$, имеем:

$$\mathbb{E}(f(X) * g(Y)) = \mathbb{E}(f(X)) * \mathbb{E}(g(Y)).$$

$\Leftarrow$ Пусть $\mathbb{E}(f(X) * g(Y)) = \mathbb{E}(f(X)) * \mathbb{E}(g(Y))$ для любых функций f и g. Докажем, что из этого условия следует независимость случайных величин X и Y.

Рассмотрим индикаторные функции:

$$f(X) = \mathbb{I}(X \in A), \quad g(Y) = \mathbb{I}(Y \in B),$$

где A и B — произвольные измеримые множества.

Тогда:

$$\mathbb{E}(f(X)) = \mathbb{P}(X \in A), \quad \mathbb{E}(g(Y)) = \mathbb{P}(Y \in B),$$

и

$$\mathbb{E}(f(X) * g(Y)) = \mathbb{P}(X \in A, Y \in B).$$

Подставляя эти значения в условие, получаем:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) * \mathbb{P}(Y \in B),$$

что соответствует определению независимости случайных величин X и Y.

■

4.2 Производящая функция

Эксперимент: Подбрасываем монетку n раз. Монетка выпадает орлом (H) с вероятностью p, решкой (T) с вероятностью $1 - p$. Броски независимы.

X – количество H, Y – количество T.

$$n = 5, \quad p = \frac{1}{2}, \quad \Omega = \{HHHHH, HHHHT, ..., TTTTT\} \qquad \text{(язык множеств)}$$

$$\mathrm{reg} = \texttt{[HT]\{5\}} \qquad \text{(язык регулярных выражений)}$$

$$f(H,T) = HHHHH + HHHHT + ... + TTTTT \qquad \text{(язык функций)}$$

$$f(H,T) = (H+T)^5$$

Обозначение:

$$H, T - \text{некоммутативные буквы } (HT \neq TH)$$
$$h, t - \text{коммутативные буквы } (h * t = t * h)$$

Заметим, что если мы передадим в производящую функцию f вместо букв вероятность её появления, то мы получим уравнения баланса вероятностей, и значение функции будет равно единице.

Примеры:

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = 1 f\left(\frac{1}{3}, \frac{2}{3}\right) = 1$$

Воспользуемся производящими функциями для расчёта матожиданий и вероятностей. Обозначим за случайную величину X количество H в последовательности, а за Y – количество T. Рассмотрим следующее выражение:

$$\frac{\partial^5 f\left(\frac{1}{2}h, \frac{1}{2}t\right)}{(\mathrm{d}h)^3(\mathrm{d}t)^2}\bigg|_{\substack{t=0\\h=0}} * \frac{1}{3!2!}$$

Очевидно, что в таком выражении дифференцирование занулит все слагаемые (каждое из которых представляет один конкретный исход), в которых меньше трёх h или меньше двух t, т.е. все, кроме слагаемых вида h^3t^2, а их обратит в $3!2!$, на которое мы потом поделим и получим количество исходов $\{X = 3, Y = 2\}$. Таким образом, что, что мы получим в итоге, будет вероятностью $\{X = 3, Y = 2\}$.

Попробуем воспользоваться производящей функцией для рассчёта матожидания. Например, попробуем найти $\mathbb{E}(X)$. Рассмотрим выражение:

$$\frac{\mathrm{d}f\left(\frac{1}{2}h, \frac{1}{2}\right)}{\mathrm{d}h}\bigg|_{h=1}$$

$f\left(\frac{1}{2}h, \frac{1}{2}\right)$ – это выражение, в котором каждому исходу соответствует значение вида $h^k * \left(\frac{1}{2}\right)^5$, где k – количество H в слове. Когда мы его продифференцируем, мы получим $k * h^{k-1} * \left(\frac{1}{2}\right)^5$ для слов, содержащих H, и занулим не содержащие. После того, как мы подставим $h = 1$, от слагаемых останется $k * \left(\frac{1}{2}\right)^5$, где k – значение X при данном исходе, а $\left(\frac{1}{2}\right)^5$ – вероятность исхода, и очевидно, что сумма таких слагаемых будет матожиданием X.

Теперь попробуем найти $\mathbb{E}(X^2)$. Посмотрим на выражение:

$$\frac{\partial^2 f\left(\frac{1}{2}e^h, \frac{1}{2}\right)}{(\partial h)^2}\bigg|_{h=0}$$

Слагаемые будут принимать вид $k^2 * \left(\frac{1}{2}\right)^5$, где k – значение X на исходе. Очевидно, что это равно $\mathbb{E}(X^2)$. Воспользовавщись тем же приёмом, найдём $\mathbb{E}(X)$ другим способом:

$$\mathbb{E}(X) = \frac{\partial f\left(\frac{1}{2}e^h, \frac{1}{2}\right)}{\partial h}\Big|_{h=0}$$

И даже так:

$$\mathbb{E}(X^2 * Y) = \frac{\partial^3 f\left(\frac{1}{2}e^h, \frac{1}{2}e^t\right)}{(\partial h)^2 * \partial t}\Big|_{\substack{h=0\\t=0}}$$

> Подбрасываем монетку до трёх H (не обязательно подряд).
>
> $$p_H = \frac{1}{3}, \quad p_T = \frac{2}{3}, \quad X - \text{количество } H, \quad Y - \text{количество } T$$
>
> $$\mathbb{E}(Y^2) = ?$$

Решение:

$$\mathrm{reg} = \texttt{\textbackslash(T* H\textbackslash)\{3\}}$$

$$f(T, H) = \left((1 + T + T^2 + T^3 + \ldots) * H\right)^3 = \left(\frac{1}{1 - T} * H\right)^3$$

$$\mathbb{E}(Y^2) = \frac{\partial^2 f\left(\frac{1}{3}e^h, \frac{2}{3}e^t\right)}{\partial t^2}\Big|_{\substack{h=0\\t=0}} =$$

4.3 Формальное определение

Определение: Говорим, что случайная величина X имеет *биномиальное распределение*, если $\mathbb{P}(X = k)$ равна вероятности получить ровно k успехов в серии независимых испытаний с вероятностью успеха p.

Обозначение: $X \sim \mathrm{Bin}(n, p)$

Определение: Говорим, что случайная величина X имеет *геометрическое распределение*, если $\mathbb{P}(X = k)$ равна вероятности получить ровно k неудач до первого успеха в серии независимых испытаний с вероятностью успеха p.

Обозначение: $X \sim \mathrm{Geom}(p)$

Определение: Говорим, что случайная величина X имеет *отрицательное биномиальное распределение*, если $\mathbb{P}(X = k)$ – вероятность получить ровно k неудач получения r-го успеха по счёту.

Обозначение: $X \sim \mathrm{NBin}(r, p)$

Определение: Если X принимает целые неотрицательные значения, то $g(t)$ – функция, производящая вероятности, если

$$g(t) = \sum_x \mathbb{P}(X = x) * t^x = \mathbb{E}(t^X)$$

Следствие:

$$g'(1) = \mathbb{E}(X)$$

Определение: $m(t)$ - функция, производящая моменты, если

$$m(t) = \sum_{x} \mathbb{P}(X = x) * e^{xt} = \mathbb{E}(e^{Xt})$$

Следствие:

$$m'(0) = \mathbb{E}(X)$$
$$m''(0) = \mathbb{E}(X^2)$$
$$\dots$$

Определение: Пусть случайные величины X, Y принимают неотрицательные значения. Тогда $g(t, s)$ – производящая функция вероятностей, если

$$g(t, s) = \sum_{x,y} \mathbb{P}(X = x, Y = y) * t^x * s^y) = \mathbb{E}(t^X * s^Y)$$

Следствие:

$$\frac{1}{3!2!} \frac{\partial^5 g(t, s)}{(\partial t)^3 (\partial s)^2} = \mathbb{P}(X = 3, Y = 2)$$

Определение: $m(t, s)$ – производящая моменты функция, если

$$m(t, s) = \sum_{x,y} \mathbb{P}(X = x, Y = y) e^{xt} * e^{ys} = \mathbb{E}(e^{Xt+Ys})$$

Лекция 5. Распределения, функция плотности и квантильная функция

30 сентября 2024 г.

Определение (Гипергеометрическое распределение): Всего есть N яблок, из них K – вкусные. Выбираем случайно (равновероятностно, без повторов) n яблок.

X – количество вкусных яблок.

Обозначение: $X \sim \mathrm{HGeom}(N, K, n)$.

Следствия:

$$P(X = k) = \frac{\binom{K}{k} * \binom{N-K}{n-k}}{\binom{N}{n}}$$

$$E(X) = E(Y_1 + ... + Y_n) = n * E(Y_1) = n * \frac{K}{N},$$

где Y_i — было ли i-е извлечённое яблоко вкусным

<u>Эксперимент:</u>

Ω – полукруг:

$$\Omega = \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1, y \geq 0\right\}$$

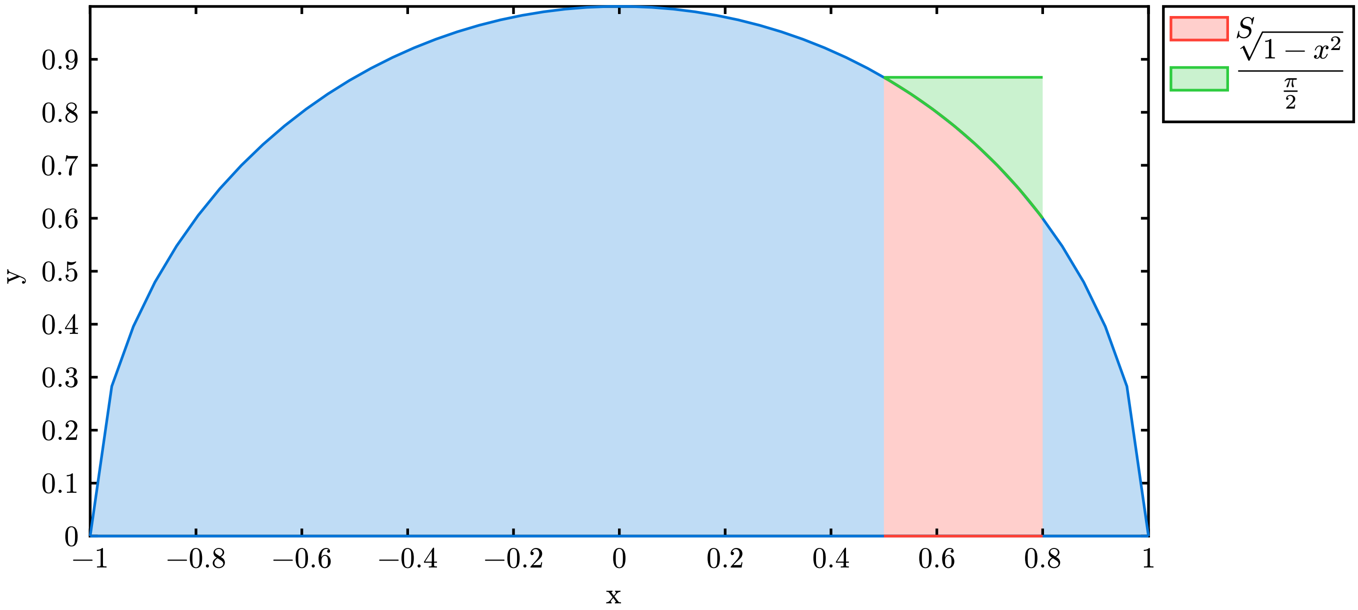
Равномерно выбираю точку $\binom{X}{Y}$ на Ω.

Уточним, что подразумевается под равномерным выбором:

$$P\left(\begin{pmatrix} X \\ Y \end{pmatrix} \in A\right) = \frac{S_A}{S_\Omega} \quad \text{(если } A \subseteq \Omega\text{)}$$

> Найдём $P(X \in [x; x + \Delta])$ с точностью до $o(\Delta)$, $[x; x + \Delta] \subseteq [-1; 1]$.

$$S_\Omega = \frac{\pi}{2}$$



$$o(\Delta) + \frac{\sqrt{1-x^2}}{\frac{\pi}{2}} * \Delta = P(X \in [x; x+\Delta]) = \frac{S}{S_\Omega}$$

Определение: Случайная величина X имеет функцию плотности $f(x)$, если

$$P(X \in [x; x+\Delta]) = f(x) * \Delta + o(\Delta)$$

Обозначение: $f(x)$ – `pdf` (probability density function)

Пример: X – абсцисса точки, равномерной выбранной на полукруге.

$$f(x) = \begin{cases} \dfrac{2\sqrt{1-x^2}}{\pi}, & x \in [-1; 1] \\ 0, & \text{иначе} \end{cases}$$

Теорема (Определение интеграла): Для случайной величины X с функцией плотности $f(x)$

$$P(X \in [a; b]) = \int\limits_a^b f(x)\,\mathrm{d}x.$$

Комментарий: $f(x)$ плохо интерпретируется сама по себе (может быть, что $f(x) > 1$).

Интерпретация:

$$f(x) * \Delta = P(X \in [x; x+\Delta]) + o(\Delta)$$

Обозначение: Вероятностно-дифференциальной формой функции плотности f называют $f(x) * \mathrm{d}x$.

Нюанс: Функция плотности определена неоднозначно.

Пример:

$$f_1(x) = \begin{cases} 1, & x \in [0; 1] \\ 0, & \text{иначе} \end{cases}$$

$$f_2(x) = \begin{cases} 1, & x \in \left(0; \frac{1}{2}\right) \cup \left(\frac{1}{2}; 1\right) \\ 2, & x = \frac{1}{2} \\ 0, & \text{иначе} \end{cases}$$

$$f_1 \neq f_2$$

$$\int\limits_a^b f_1(x)\,\mathrm{d}x = \int\limits_a^b f_2(x)\,\mathrm{d}x = P(X \in [a;b])$$

Определение: X *равномерно* распределена на отрезке $[a;b]$, если

$$P(X \in [c;d]) = \frac{|d-c|}{|b-a|} \quad \text{при } [c;d] \subseteq [a;b].$$

Обозначение: $X \sim U[a;b]$
$X \sim \mathrm{Unif}[a;b]$

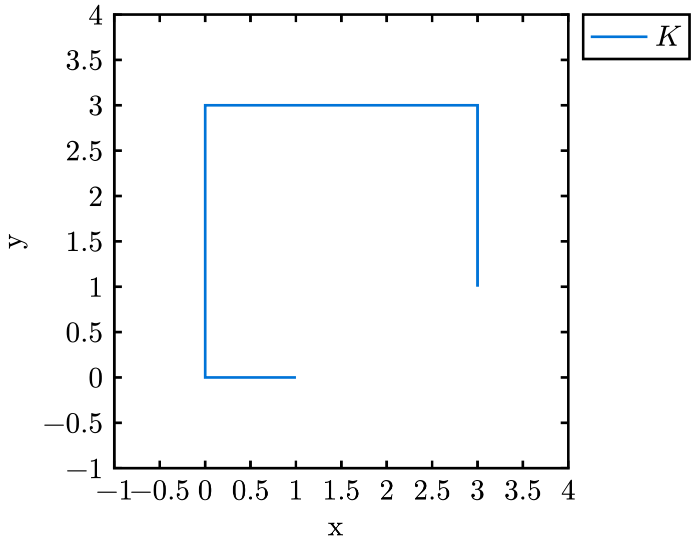
> Найдём функцию плотности $f(x)$ для $X \sim U[2;2024]$.

Для отрезка $[x;x+\Delta] \subseteq [2;2024]$ имеем

$$P(X \in [x;x+\Delta]) = \frac{|x+\Delta-x|}{2022} = \frac{1}{2022} \cdot \Delta = f(x) \cdot \Delta$$

$$f(x) = \begin{cases} \frac{1}{2022}, & x \in [2, 2024] \\ 0, & \text{иначе} \end{cases}$$

Пример:



Ломанная $K = \mathrm{AB} \cup \mathrm{BC} \cup \mathrm{CD} \cup \mathrm{DE}$. $A(1,0), B(0,0), C(0,3), D(3,3), E(3,1)$.

$\binom{X}{Y}$ – случайно выбранная точка на K.

$$P\left(\begin{pmatrix} X \\ Y \end{pmatrix} \in \mathrm{AB}\right) = \frac{|\mathrm{AB}|}{|\mathrm{K}|} = \frac{|\mathrm{AB}|}{|\mathrm{AB}| + |\mathrm{BC}| + |\mathrm{CD}| + |\mathrm{DE}|} = \frac{1}{9}$$

Пусть случайная величина X есть абсцисса этой точки. Она не дискретна – принимает действительные значения от 0 до 3. Однако и функции плотности у нее нет, это не дает представить вероятность в виде $f(x) \cdot \Delta + o(\Delta)$.

Как описать такую величину?

Определение: Случайная величина X называется *смешанной (mixed)*, если

$$P(X \in [x; x+\Delta]) = f(x) * \Delta + \sum_{x_i \in [x; x+\Delta]} P(X = x_i) + o(\Delta)$$

В примере выше нас интересуют точки $\{0, 3\}$, для остальных $P = 0$ – у точек на горизонтальных отрезках нет длины. Тогда в нашем случае получаем

$$P(X \in [x; x+\Delta]) = f(x) \cdot \Delta + \frac{|\mathrm{BC}|}{|\mathrm{K}|} * I(0 \in [x, x+\Delta]) + \frac{|\mathrm{DE}|}{|\mathrm{K}|} * I(3 \in [x, x+\Delta]) =$$
$$= f(x) \cdot \Delta + \frac{1}{3} * I(0 \in [x, x+\Delta]) + \frac{2}{9} * I(3 \in [x, x+\Delta])$$

$$f(x) = \begin{cases} \frac{2}{9}, & x \in [0, 1] \\ \frac{1}{9}, & x \in (1, 2) \\ 0, & \text{иначе} \end{cases}$$

Итого, вероятность смешанной величины это сумма интеграла (для континуальной части с.в.) и дискретной суммы (через Σ)

Теорема: Для смешанной случайной величины X

$$P(X \in [a; b]) = \int_a^b f(x)\,\mathrm{d}x + \sum_{x_i \in [a; b]} P(X = x_i).$$

Теорема (LOTUS, law of the uncosctous statistician):

1. *Для дискретных случайных величин:*

$$E(g(X)) = \sum_x g(x) * P(X = x)$$

2. *Для случайных величин с функцией плотности $f(x)$:*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x$$

3. *Для смешанных случайных величин:*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x + \sum_{x \,\mid\, P(X=x)>0} g(x) * P(X = x)$$

> $X \sim \mathrm{Unif}[a;b]$. $E(X^2) =?$

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a;b] \\ 0, & \text{иначе} \end{cases}$$

$$E(X^2) = \int\limits_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \frac{1}{b-a}\int\limits_{a}^{b} x^2\,\mathrm{d}x = \frac{1}{b-a} * \frac{x^3}{3}\Big|_a^b =$$
$$= \frac{b^3-a^3}{3(b-a)}$$

5.1 Функция распределения и квантильная функция

Определение: *Функций распределения («cumulative distribution function, CDF»)* случайной величины X называется $F(x) = P(X \le x)$.

Исторический нюанс: В СССР и их наследниках распространено альтернативное определение:

$$F(x) = P(X < x).$$

> $X \sim U[2;2024]$
>
> Функция распределения $F(x) =?$

$P(X \le 2) = 0$ – там нет значений величины. $P(X \le 2024) = 1$. На $(2, 2024)$ ее значения равномерно распределены (возрастают), т.е. расположены на прямой.

$$F(x) = \begin{cases} 0, & x \le 2 \\ \frac{x-2}{2022}, & x \in (2, 2024) \\ 1, & x \ge 2024 \end{cases}$$

> $X_1, X_2, ..., X_{10}$ независимы и $\mathrm{Unif}[0;1]$.
>
> Упорядочим их по возрастанию.
>
> $$X_{(1)} \le X_{(2)} \le ... \le X_{(10)}$$
>
> Найдём функцию плотности у $X_{(3)}$.

$$P\Big(X_{(3)} \in [x; x+\Delta]\Big) = o(\Delta) = \binom{10}{2} * \left(\frac{x}{1}\right)^2 * 8 * \Delta * (1-x)^7$$

$$f(x) = \binom{10}{2} * 8 * x^2 * (1-x)^7$$

Определение: *Квантильная функция («quantile function», «percentile point function», PPF)* – функция, обратная к функции распределения.

$$Q(p) = \inf_x \{x \mid F(x) \geq p\}$$

Лекция 6. Геометрия для случайных величин и событий

7 октября 2024 г.

6.1 Первая геометрия

Как известно, основным понятием, определяющим геометрию, является скалярное произведение. Попробуем определить скалярное произведение для случайных величин.

Возможны разные способы определить скалярное произведение для случайных величин. Начнём с более очевидного и менее удачного. Определим его так:

$$<X,Y>_0 = \mathbb{E}(X * Y)$$

Вспомним свойства скалярного произведения:

Свойство	Выполняется ли для нашего определения?
$< v, v >= 0 \Leftrightarrow v = 0$	с оговоркой на понимание нуля
$<v,v> \geq 0$	✅
$<v,w> = <w,v>$	✅
$<a+b,v> = <a,v> + <b,v>$	✅

Теперь мы можем определить скалярное произведение для событий как скалярное произведение их индикаторов:

$$<A,B>_0 = \mathbb{E}(I_A, I_B)$$

$$A \perp_0 B \Leftrightarrow \text{события } A \text{ и } B \text{ никогда не происходят одновременно}$$

$$\|A\|_0 = \|I_A\|_0 = \sqrt{\mathbb{E}(I_A * I_A)} = \sqrt{\mathbb{E}(I_A)} = \sqrt{\mathbb{P}(A)}$$

А что если попробовать брать проекцию на $\mathbb{R}$?

$$\min_c \|X - c\|_0^2 = \min_c \mathbb{E}((X-c)^2) = \min_c (\mathbb{E}(X^2) + c^2 - 2\mathbb{E}(X) * c)$$

$$\Rightarrow \text{минимум достигается при } c = \mathbb{E}(A)$$

Таким образом, проекция случайной величины на прямую $\mathbb{R}$ – это её математическое ожидание.

Теорема (Пифагора):

$$\|X\|_0^2 = \|E(X)\|_0^2 + \|X - \mathbb{E}(X)\|_0^2$$

Доказательство:

$$\|X\|_0^2 = <X,X> = \mathbb{E}(X^2) = \mathbb{E}(X)^2 + \mathbb{E}[(X - \mathbb{E}(X))^2] = \|X\|_0^2 + \|X - \mathbb{E}(X)\|_0^2$$

■

Определение: *Дисперсия (variance)* случайной величины X – это $\|X - \mathbb{E}(X)\|_0^2$.

Обозначение:
- В русскоязычной литературе: $D(X)$
- В европейской литературе: $\mathbb{V}\mathrm{ar}(X)$

Следствия:
- $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right]$
- $\mathbb{V}\mathrm{ar}(cX) = c^2\,\mathbb{V}\mathrm{ar}(X), \quad \forall c \in \mathbb{R}$
- $\mathbb{V}\mathrm{ar}(X) = \mathbb{V}\mathrm{ar}(-X)$

6.2 Вторая геометрия

Рассмотрим подпространство, ортогональное $\mathbb{R}$. Заметим, что это $\{W \mid \mathbb{E}(W) = 0\}$.

Обозначим величины X на это подпространство как X_c (X центрированный). Заметим, что $X_c = X - \mathbb{E}(X)$.

Определение (Ещё одно скалярное произведение):

$$<X,Y> = <X_c,Y_c>_0$$

Обозначение: Такое скалярвное произведение называют **ковариацией** X и Y.

$$\mathbb{C}\mathrm{ov}(X,Y) = <X,Y> = \mathbb{E}[(X - \mathbb{E}(X)) * (Y - \mathbb{E}(Y))]$$

Определим также ковариацию для событий.

> Традиционно ковариацию для событий никто в здравом уме не определяет, но мы всё равно это сделаем.

$$\begin{aligned} <A,B> &:= <I_A,I_B> = \mathbb{E}[(I_A - \mathbb{P}(A)) * (I_B - \mathbb{P}(B))] = \\ &= \mathbb{E}(I_A I_B) - \mathbb{P}(A) * \mathbb{P}(B) = \mathbb{P}(A \cap B) - \mathbb{P}(A) * \mathbb{P}(B) \end{aligned}$$

$$A \perp B \Leftrightarrow <A,B> = 0 \Leftrightarrow A \text{ и } B \text{ независимы}$$

Вспомним **критерий независимости случайных величин:**

$$X \text{ и } Y \text{ независимы} \Leftrightarrow \forall f,g : \mathbb{E}(f(x) * g(Y)) = \mathbb{E}(f(Y)) * \mathbb{E}(g(Y))$$

Переформулируем это в терминах ковариации:

$$\begin{aligned} \mathbb{C}\mathrm{ov}(X,Y) = <X,Y> &= \mathbb{E}[(X - \mathbb{E}(X)) * (Y - \mathbb{E}(Y))] = \\ &= \mathbb{E}[XY - \mathbb{E}(X) * Y - X * \mathbb{E}(Y) + \mathbb{E}(X) * \mathbb{E}(Y)] = \\ &= \mathbb{E}(XY) - \mathbb{E}(X) * \mathbb{E}(Y) \end{aligned}$$

Теорема (Критерий независимости СВ в терминах геометрии):

$$X \text{ и } Y \text{ независимы} \Leftrightarrow \forall f,g : f(X) \perp g(Y)$$

Определение: *Стандартное отклонение* случайной величины X – это $\sqrt{\mathbb{V}\mathrm{ar}(X)}$.

Обозначение:

$$\sigma_X$$

Следствие: $\sigma_X = \|X\|$.

Определение: *Корреляция* случайных величин X и Y – это $\dfrac{<X,Y>}{\|X\| * \|Y\|}$.

Обозначение: $\mathbb{C}\mathrm{orr}(X, Y)$

Следствие:

$$\mathbb{C}\mathrm{orr}(X, Y) = \frac{<X,Y>}{\|X\| * \|Y\|} = \frac{\mathbb{C}\mathrm{ov}(X, Y)}{\sqrt{\mathbb{V}\mathrm{ar}\, X} * \sqrt{\mathbb{V}\mathrm{ar}(Y)}}$$

Обратите внимание, что корреляция случайных величин в геометрическом смысле эквивалентна косинусу угла между ними.

Теорема: Если $\mathbb{C}\mathrm{orr}(X, Y)$ существует, то $\mathbb{C}\mathrm{orr}(X, Y) \in [-1; 1]$.

Определение: Случайные величины X и Y *некоррелированы*, если $\mathbb{C}\mathrm{ov}(X, Y) = 0$.

Теорема: Если случайные величины независимы, то они некореллированы.

Доказательство:

$$\forall f, g : f(X) \perp g(Y) \Rightarrow X \perp Y \Rightarrow \mathbb{C}\mathrm{ov}(X, Y) = 0$$

■

Так как наша новая геометрия случайных величин – геометрия, для неё можно переформулировать многие знакомые нам теоремы евклидовой геометрии. Приведём пример:

Теорема (косинусов):

$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2\|X\| * \|Y\| * \cos(X, Y)$$

Теорема (косинусов для случайных величин):

$$\mathbb{V}\mathrm{ar}(X - Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) - 2\sigma_X * \sigma_Y * \mathbb{C}\mathrm{orr}(X, Y)$$

Или, что то же самое:

$$\mathbb{V}\mathrm{ar}(X - Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) - 2\, \mathbb{C}\mathrm{ov}(X, Y)$$

Теорема: Если X и Y независимы, то $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X - Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$

Доказательство: Тривиально следует из теоремы Пифагора.

6.3 Неравенства

Теорема (Неравенство Маркова): Если $X \geq 0$ и $a > 0$, то $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}(X)}{a}$.

Доказательство:

$$X = X * I(X \geq a) + X * I(X < a)$$

$$\underline{X} \coloneqq a * I(X \geq a) + 0 * I(X < a)$$

$$\mathbb{E}(X) \geq \mathbb{E}(\underline{X}) = a * \mathbb{P}(X \geq a)$$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

■

Теорема (Неравенство Чебышёва):

$$\mathbb{P}(|A - \mathbb{E}(X)| \geq \Delta) \leq \frac{\mathbb{V}\mathrm{ar}\, X}{\Delta^2}$$

Доказательство:

$$Y \coloneqq (X - \mathbb{E}(X))^2$$

$$Y \geq 0$$

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}(Y)}{a}$$

$$\mathbb{P}\big(|A - \mathbb{E}(X)| \geq \sqrt{a}\big) \leq \frac{\mathbb{V}\mathrm{ar}\, X}{a}$$

$$\Delta \coloneqq \sqrt{a}$$

$$\mathbb{P}(|A - \mathbb{E}(X)| \geq \Delta) \leq \frac{\mathbb{V}\mathrm{ar}\, X}{\Delta^2}$$

■

Теорема (Неравенство Йенсена): Если f – выпуклая функция (надграфик f – выпуклое множество, f имеет форму $\cup$), то

$$\mathbb{E}(f(x)) \geq f(\mathbb{E}(X))$$

для любой случайной величины X.

Идея неравенства Йенсена: если на выпуклом блюде расположены грузы, то их центр масс тоже лежит внутри блюда.

Доказательство будет в следующий раз.

Лекция 7. Энтропия, экспоненциальное распределение

14 октября 2024 г.

Определение: Функция $f(x)$ называется *выпуклой*, если через любую точку $P = (a, f(a))$ можно провести прямую $l = \{(x, y) \mid y = kx + b\}$ такую, что $f(x) \geq kx + b$.

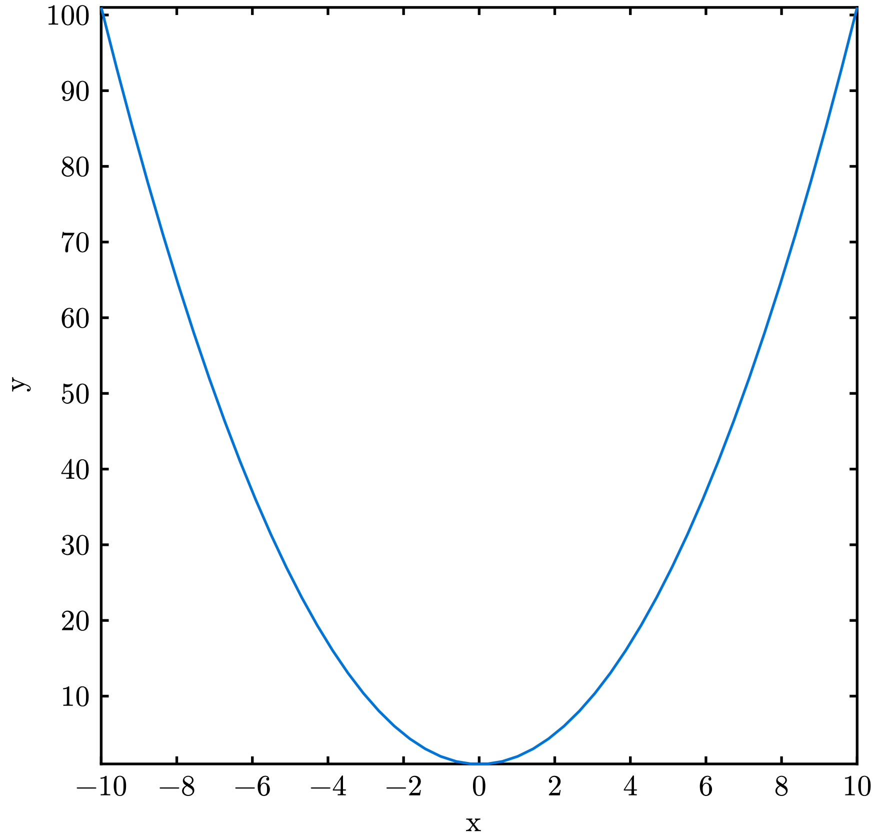
Докажем неравенство Йенсена для такого определения выпуклости.

Проведём прямую (как в определении выпуклости) l, заданную уравнением $y = kx + b$, через точку $(\mathbb{E}(X), f(\mathbb{E}(X)))$.

$$\forall x : f(x) \geq kx + b$$

$$\Rightarrow f(X) \geq k * X + b$$

$$\mathbb{E}(f(X)) \geq \mathbb{E}(k * X + b) = k * \mathbb{E}(X) + b = f(\mathbb{E}(X))$$



7.1 Энтропия случайной величины

Интуиция

Энтропия – это минимальное среднее число бинарных вопросов, за которое можно угадать значение X.

Пример: Загадаем число x:

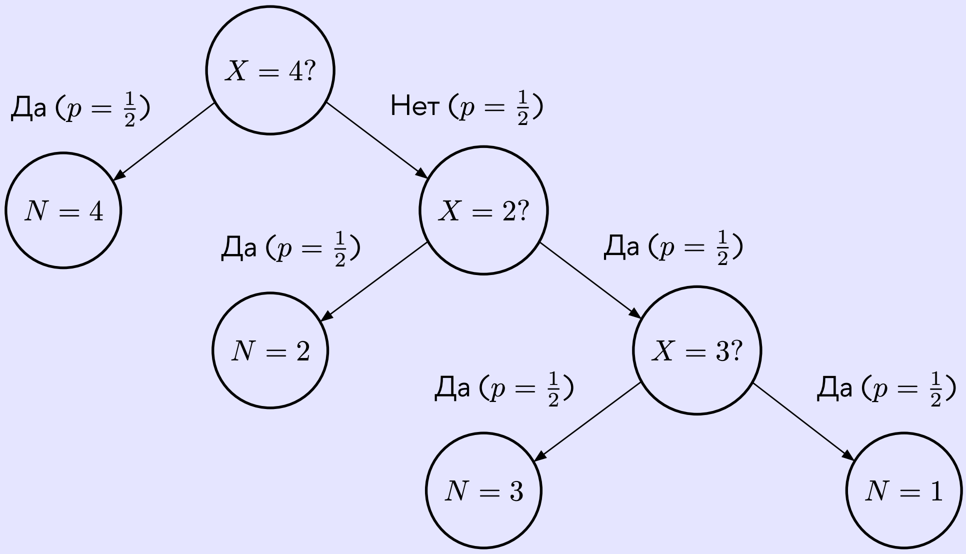
x	1	2	3	4
$\mathbb{P}(X = x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$

На лекции Б. Демешев загадал число, написал его на обратной стороне доски, и мы задавали ему вопросы, пытаясь угадать это число.

Попытаемся его угадать, задавая бинарные вопросы:

- Q_1: Правда ли, что $X = 4$?
- A_1: Нет.
- Q_2: Правда ли, что $X = 2$?
- A_2: Да.

Число было угадано за два вопроса. Но это не всегда будет так. Нас интересует именно среднее количество вопросов для оптимальной стратегии.



Формальное определение

Определение (Энтропия для дискретной СВ): $H(X) = \sum \mathbb{P}(X = x) * \log_{1/2} \mathbb{P}(X = x)$

Энтропия – это также количество бит, которым при оптимальной кодировке можно закодировать значение случайной величины. Вот альтернативная формула энтропии в битах:

$$H(X) = -\sum \mathbb{P}(X = x) * \log_2 \mathbb{P}(X = x)$$

Можно также считать энтропию не в битах, а в *натах*:

$$H(X) = -\sum \mathbb{P}(X = x) * \ln \mathbb{P}(X = x)$$

Определение: Для случайной величины с функцией плотности $f(y)$

$$H(Y) = -\int\limits_{-\infty}^{\infty} f(y) * \ln f(y)\,\mathrm{d}y \quad \text{[нат]}$$

$$H(Y) = -\int\limits_{-\infty}^{\infty} f(y) * \log_2 f(y)\,\mathrm{d}y \quad \text{[бит]}$$

$$\mathrm{CE}(f \parallel g) = -\int\limits_{-\infty}^{\infty} f(y) \ln g(y)\,\mathrm{d}y \quad \text{[нат]}$$

$$\mathrm{CE}(f \parallel g) = -\int\limits_{-\infty}^{\infty} f(y) \log_2 g(y)\,\mathrm{d}y \quad \text{[бит]}$$

7.2 Кросс-энтропия

Интуиция

Пусть есть случайная величина x. Обозначим за $\mathbb{P}(X = x)$ вероятность того, что $X = x$, а за $\mathbb{Q}(X = x)$ «ошибочную вероятность» того, что $X = x$ (т.е. вероятность в какой-то модели, отличной от реальности, то есть от модели, в которой мы считаем $\mathbb{P}$).

Тогда $\mathrm{CE}(p \parallel q)$ («из q в p») – минимальное ожидаемое число вопросов, чтоб узнать значение X, если истинное распределение – это p, а оптимизируем стратегию вопросов мы под Q.

Строгое определение

$$\mathrm{CE}(p \parallel q) = \sum \mathbb{P}(X = x) * \log_{1/2} \mathbb{Q}(X = x) \quad \text{[бит]}$$

$$\mathrm{CE}(p \parallel q) = -\sum \mathbb{P}(X = x) * \log_2 \mathbb{Q}(X = x) \quad \text{[бит]}$$

$$\mathrm{CE}(p \parallel q) = -\sum \mathbb{P}(X = x) * \ln \mathbb{Q}(X = x) \quad \text{[нат]}$$

Свойства

Теорема: $\mathrm{CE}(p \parallel q) \geq H(p)$.

Доказательство: Положим $p_i = \mathbb{P}(X = x_i), q_i = \mathbb{Q}(X = x_i)$. Заметим, что $x - 1 \geq \ln x$, так как функция $f(x) = \ln x$ – выпуклая, поэтому верно следующие неравенство

$$\frac{q_i}{p_i} - 1 \geq \ln \frac{q_i}{p_i} \Longleftrightarrow q_i - p_i \geq p_i \ln q_i - p_i \ln p_i$$

$$\Downarrow$$

$$\underbrace{\sum_i q_i}_{1} - \underbrace{\sum_i p_i}_{1} \geq \underbrace{\sum_i p_i \ln q_i}_{\mathrm{CE}(p \parallel q)} - \underbrace{\sum_i p_i \ln p_i}_{H(p)}$$

$$\Updownarrow$$

$$\mathrm{CE}(p \parallel q) \geq H(p)$$

■

7.3 Экспоненциальное распределение

Интуиция

Определение: Случайная величина обладает *отсутствием памяти*, если

$$\mathbb{P}(Y \in [y; y + \Delta] \mid Y \geq y) = \mathbb{P}(Y \in [0; \Delta])$$

Предположим, что у случайной величины Y есть функция плотности $f(y)$ и отсутствие памяти. Найдём $f(y)$.

По определению:

$$\frac{\mathbb{P}(Y \in [y; y + \Delta], Y \geq y)}{\mathbb{P}(Y \geq y)} = \mathbb{P}(Y \in [0; \Delta])$$

$$\mathbb{P}(Y \in [y; y + \Delta]) = \mathbb{P}(Y \in [0; \Delta]) * \mathbb{P}(Y \geq y)$$

$$f(y) * \Delta + o(\Delta) = (f(0) * \Delta + o(\Delta)) * \mathbb{P}(Y \geq y)$$

$$f(y) * \Delta = f(0) * \Delta * \mathbb{P}(Y \geq y) + o(\Delta)$$

$$/\Delta \quad \text{и} \quad \Delta \to 0$$

$$f(y) = f(0) * \mathbb{P}(Y \geq y) + 0$$

$$\frac{\mathrm{d}\mathbb{P}(Y \leq y)}{\mathrm{d}y} = f(0) * \mathbb{P}(Y \geq y)$$

$$\frac{\mathrm{d}(\mathbb{P}(Y \geq y))}{\mathrm{d}y} = -f(0) * \mathbb{P}(Y \geq y)$$

$$f(0) =: \lambda$$

$$\frac{\mathrm{d}(\mathbb{P}(Y \geq y))}{\mathrm{d}y} = -\lambda * \mathbb{P}(Y \geq y)$$

$$\mathbb{P}(Y \geq y) * c * e^{-\lambda y}$$

$$1 = \mathbb{P}(Y \geq 0) = c$$

$$\mathbb{P}(Y \geq y) = e^{-\lambda y}$$

$$\mathbb{P}(Y \leq y) = \begin{cases} 1 - e^{-\lambda y}, & y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

$$f(y) = \begin{cases} \lambda * e^{-\lambda y}, & y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Формальное определение

Определение: Случайная величина имеет *экспоненциальное (показательное)* распределение с интенсивностью λ, если её функция плотности равна

$$f(y) = \begin{cases} \lambda * e^{-\lambda y}, & y \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Следствие:

$$\mathbb{E}(Y) = \int\limits_0^\infty y * f(y)\,\mathrm{d}y = \int\limits_0^\infty y * \lambda e^{-\lambda y}\,\mathrm{d}y = \frac{1}{\lambda}$$

$$H(Y) = \int\limits_0^\infty \lambda e^{-\lambda y} * (\ln\lambda - \lambda y)\,\mathrm{d}y = -\int\limits_0^\infty \lambda e^{-\lambda y} \ln\lambda\,\mathrm{d}y + \lambda\int\limits_0^\infty y * \lambda e^{-\lambda y}\,\mathrm{d}y =$$
$$= 1 - \ln\lambda$$

Теорема: Среди неотрицательных величин с заданным ожиданием максимальную энтропию имеет экспоненциально распределённая.

Или, говоря более формально, если случайная величина $X \geq 0$ у X есть $f(x)$ и $\mathbb{E}(X) = \frac{1}{\lambda}$, то $H(X) \leq 1 - \ln\lambda$.

Доказательство: Пусть модель q описывает величину X как экспоненциальную, а на самом деле она не экспоненциальная.

$$\mathrm{CE}(p \parallel q) \geq H(p)$$

$$q(x) = \lambda e^{-\lambda x}$$

$$
\begin{aligned}
-\int_0^{\infty} p(y) * \ln\left(\lambda * e^{-\lambda y}\right) \mathrm{d}y = -\int_0^{\infty} p(y) * (\ln \lambda - \lambda * y) \mathrm{d}y = \\
= -\int_0^{\infty} p(y) \ln \lambda \, \mathrm{d}y + \lambda * \int_0^{\infty} y * p(y) \, \mathrm{d}y = \\
= -\ln \lambda \int_0^{\infty} p(y) \, \mathrm{d}y + \lambda * \int_0^{\infty} y * p(y) \, \mathrm{d}y = \\
= -\ln \lambda + \lambda \mathbb{E}(X) = -\ln \lambda + \lambda \cdot \frac{1}{\ln \lambda} = 1 - \ln \lambda \geq H(p)
\end{aligned}
$$

Определение (Совместная энтропия):

$$
H(X, Y) = -\sum_{v} \mathbb{P}(V = v) * \ln \mathbb{P}(V = v), \quad V = \begin{pmatrix} X \\ Y \end{pmatrix}
$$

$$
H(X, Y) = -\sum_{x, y} \mathbb{P}(X = x, Y = y) * \ln \mathbb{P}(X = x, Y = y)
$$

Теорема: Если X и Y независимы, то $H(X, Y) = H(X) + H(Y)$.

Доказательство: Идея в том, что $\ln(a * b) = \ln a + \ln b$, далее тривиально.

Определение (Условная энтропия):

$$
H(Y \mid X) = -\sum \mathbb{P}(X = x, Y = y) * \ln \mathbb{P}(Y = y \mid X = x)
$$

Интуиция: минимальное ожидаемое количество вопросов, чтобы узнать Y, если X известен.

Пасхалка от составителей конспекта: если бы вы пошли на курс по функциональному программированию, то задумались бы насчёт того, чему равно YH.

А чему?

Теорема: $H(X) + H(Y \mid X) = H(X, Y)$.

Лекция 8. Распределение случайного вектора

21 октября 2024 г.

8.1 Случайные матрицы и векторы

Определение: Если X – *случайная матрица* $[n \times k]$, то $\mathbb{E}(X) = \big(\mathbb{E}(X_{ij})\big)_{ij}$

Определение: Если X – *вектор случайных величин* $[n \times 1]$ или $[1 \times n]$, то

$$\mathbb{V}\mathrm{ar}(X) = \begin{bmatrix} \mathbb{V}\mathrm{ar}(X_1) & \mathbb{C}\mathrm{ov}(X_1, X_2) & \ldots \\ \mathbb{C}\mathrm{ov}(X_2, X_1) & \ddots & \\ \vdots & & \mathbb{V}\mathrm{ar}(X_n) \end{bmatrix}$$

Эта матрица называется *ковариационной матрицей.*

Определение: Если X и Y – случайные векторы, то

$$\mathbb{C}\mathrm{ov}(X, Y) = A$$

$$A_{ij} = \mathbb{C}\mathrm{ov}\big(X_i, Y_j\big)$$

Теорема:

$$\mathbb{E}(A * X + B * Y + C) = A * \mathbb{E}(X) + B * \mathbb{E}(Y) + C$$

$$X, Y \text{ — случайные матрицы}$$

$$A, B, C - \text{константные матрицы}$$

$$\mathbb{E}(X * A) = \mathbb{E}(X) * A$$

Определение: *Совместная функция распределения* $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ – это

$$F(x_1, ..., x_n) = \mathbb{P}(X_1 \le x_1, ..., X_n \le x_n)$$

Определение: У вектора $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ есть совместная функция плотности $f(x_1, ..., x_n)$ если

$$F(x_1, ..., x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, ..., t_n) \mathrm{d}t_1 ... \mathrm{d}t_n$$

Теорема (Критерий независимости случайных величин): Случайные величины независимы тогда и только тогда, когда

$$F(x_1, ..., x_n) = F_1(x_1) * ... * F_n(x_n),$$

что эквивалентно тому, что

$$\mathbb{P}(X_1 \le x_1, ..., X_n \le x_n) = \mathbb{P}(X_1 \le x_1) * ... * \mathbb{P}(X_n \le x_n).$$

Если у случайных величин есть функция плотности, то это эквивалентно тому, что

$$f(x_1, ..., x_n) = f_1(x_1) * ... * f_n(x_n)$$

8.2 Аналоги условной вероятности

Определение (Условная функция плотности):

$$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Определение (Условная вероятность):

$$\mathbb{P}(Y \in A \mid X = x) = \underset{y \in A}{\int \cdots \int} f_{Y \mid X}(y \mid x) \mathrm{d}y_1 * \ldots * \mathrm{d}y_n$$

Определение (Условное математическое ожидание):

$$\mathbb{E}(Y \mid X = x) = \int\limits_{\mathbb{R}} y * f(y \mid x) \, \mathrm{d}y$$

Определение (Условная дисперсия):

$$\mathbb{V}\mathrm{ar}(Y \mid X = x) = \mathbb{E}(Y^2 \mid X = x) - (\mathbb{E}(Y \mid X = x))^2$$

Определение (Условная ковариация):

$$\mathbb{C}\mathrm{ov}(Y, W \mid X = x) = \mathbb{E}(Y * W \mid X = x) - \mathbb{E}(Y \mid X = x) * \mathbb{E}(W \mid X = x)$$

В последних двух определениях X мог быть как скаляром, так и вектором, а вот Y и X – скаляры. Если же предположить, что Y и W – векторы $n \times 1$ и $1 \times n$ соответственно, то получаем немного другие формулы... которые я не успел записать, так что ждите, пока Андрей их сюда допишет.

Определение (Условная функция распределения):

$$F_{Y \mid X} = \int\limits_{-\infty}^{y_1} \cdots \int\limits_{-\infty}^{y_n} f(t_1, \ldots, t_n) \mathrm{d}t_1 * \ldots * \mathrm{d}t_n$$

Теорема (LOTUS): Если X – вектор случайных величин с функцией плотности $f(x)$, $h : \mathbb{R}^n \to \mathbb{R}$, то

$$\mathbb{E}(h(X)) = \underset{\mathbb{R}^n}{\int \cdots \int} h(x) * f(x) \mathrm{d}x_1 * \ldots * \mathrm{d}x_n.$$

Если $h : \mathbb{R}^n \to \mathbb{R}$, у (X, Y) есть совместная плотность $f(x, y)$, то

$$\mathbb{E}(h(y) \mid X = x) = \underset{\mathbb{R}^n}{\int \cdots \int} h(y) * f(y \mid x) \, \mathrm{d}y.$$

8.3 Переход в новые координаты

Определение (классическое): Если $h : \mathbb{R}^n \to \mathbb{R}^n$ дифференцируема, взаимнооднозначная, определитель её матрицы Якоби J не равен нулю и не уходит на бесконечность и $y = h(x)$, то

$$f_Y(y_1, ..., y_n) = f_X(h^{-1}(y)) * |\det J|$$

Пример:

Упражнение.

$$f(x_1, x_2) := \begin{cases} 2x_1 + 3x_2^2, & x_1 \in [0; 1], x_2 \in [0; 1] \\ 0, & \text{иначе} \end{cases}$$

$$S := X_1 + X_2$$

$$f_{X_1, S} = ?$$

Решение:

$$X_1 = X_1$$

$$X_2 = S - X_1$$

$$J = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

$$|\det J| = 1$$

$$f_{X_1, S} = 1 * \begin{cases} 2x_1 + 3(s - x_1)^2, & x_1 \in [0; 1], s \in [x_1, x_1 + 1] \\ 0, & \text{иначе} \end{cases}$$

8.4 Снижение размерности

Теорема: Пусть есть векторы случайных величин X и Y. Тогда:

$$F_X(x) = \lim_{\substack{y_1 \to \infty \\ y_2 \to \infty \\ \vdots \\ y_k \to \infty}} F_{X,Y}(x, y)$$

$$f_X(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x, y) \mathrm{d}y_1 * ... * \mathrm{d}y_k$$

8.5 Переход в другие координаты в терминах дифференциальных форм

Определение (Дифференциальная форма):

$$f(x_1, ..., x_n) * \mathrm{d}x_1 \wedge ... \wedge \mathrm{d}x_n$$

Правила работы с ней

- $\mathrm{d}a \wedge \mathrm{d}a = 0$
- $\mathrm{d}a \wedge \mathrm{d}b = -\mathrm{d}b \wedge \mathrm{d}a$

Правила перехода в другие координаты

Просто подставьте новые координаты в старую форму.

Пример:

$$f(x_1, x_2)\mathrm{d}x_1 \wedge \mathrm{d}x_2 = (2x_1 + 3x_2)\mathrm{d}x_1 \wedge \mathrm{d}x_2$$

$$\begin{cases} x_1 = x_1 \\ x_2 = s - x_1 \end{cases}$$

$$\mathrm{d}x_1 \wedge \mathrm{d}x_2 = \mathrm{d}x_1 \wedge (\mathrm{d}s - \mathrm{d}x_1) = \mathrm{d}x_1 \wedge \mathrm{d}s - \underbrace{\mathrm{d}x_1 \wedge \mathrm{d}x_1}_{0} = \mathrm{d}x_1 \wedge \mathrm{d}s$$

$$f(x_1, x_2)\mathrm{d}x_1 \wedge \mathrm{d}x_2 = (2x_1 + 3x_2)\mathrm{d}x_1 \wedge \mathrm{d}x_2 = (2x_1 + 3(s - x_1))\mathrm{d}x_1 \wedge \mathrm{d}s$$

8.6 Формула свёртки

Теорема: Если X_1 и X_2 независимы и имеют функции плотности $f_1(x_1)$ и $f_2(x_2)$ и $S = X_1 + X_2$, то

$$f_S(S) = \int\limits_{-\infty}^{\infty} f_1(x_1) * f_2(s - x_1)\mathrm{d}x_1$$

Доказательство:

$$f(x_1, x_2) = f_1(x_1) * f_2(x_2)$$

Перейдём в координаты x_1, s. $|\det J| = 1$.

$$f_{X_1,S}(x_1, s) = f_1(x_1) * f_2(s - x_1) * 1$$

$$f_S(s) = \int\limits_{-\infty}^{\infty} f_1(x_1) * f_2(s - x_1)\mathrm{d}x_1$$

■

8.7 Граф вероятностной модели / байесовская сеть

Определение: Ориентированный граф без циклов, в котором вершины ассоциированны со случайными величинами, а рёбра – «влияние» случайной величины на другую, называется *графом вероятностной модели*, или *байесовской сетью*.

Если есть граф вероятностной модели с вершинами $X_1, ..., X_n$, то

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i \mid \mathrm{Parents}(X_i)),$$

где $\mathrm{Parents}(X_i)$ – вершины, из которых есть рёбра в X_i.

8.8 Копула

Идея

- **Шаг 1.** Выразим все $X_1, ..., X_n$ через равномерно распределённые $U_1, ..., U_n$.
- **Шаг 2.** Зададим зависимости $U_1, ..., U_n, \quad U_i \sim \mathrm{Unif}[0; 1]$.

Определение

Копула – функция распределения $U_1, ..., U_n$, где $U_i \sim U[0; 1]$.

Теорема: Любая функция распределения $F(x_1, ..., x_n)$ представима в виде копулы

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)),$$

где C – копула.

Лекция 9. Пуассоновский поток (процесс). Распределение Пуассона.

11 ноября 2024 г.

9.1 Аксиомы и свойства

Построим процесс *во времени*, которым удобно будет моделировать некоторые реальные события. Пусть на непрерывном луче времени (t) происходят некоторые независимые происшествия (события). Введём случайные величины:

X_t - Количество происшествий за период $[0, t]$

Y_n - Время между $(n-1)$ и n-ым событием

Договоренность. СВ X_t непрерывна справа.

Строя новую теорию, необходимо принять новые постулаты (предпосылки):

1. $X_0 = 0$
2. Приращение $X_t - X_s (s \leq t)$ зависит только от $t - s$ (то есть конкретное время событий не играет роли, только временной промежуток)
3. Приращение на непересекающихся интервалах независимы:
$\forall t_1 \leq ... \leq t_k$ независимы величины $\left(X_{t_2} - X_{t_1}\right), \left(X_{t_4} - X_{t_3}\right), ..., \left(X_{t_n} - X_{t_{n-1}}\right)$
4. $\mathbb{P}(X_\Delta = 1) = \lambda \cdot \Delta + o(\Delta); \mathbb{P}(X_\Delta \geq 2) = o(\Delta)$ - На малом интервале вероятность совершения ровно одного события примерно пропорциональна длине интервала; вероятность же совершения более одного события мала.

Теорема/упражнение.

$$\mathbb{P}(X_t = 0) -? \ \mathbb{P}(X_t = 1) -? \ \mathbb{P}(X_t = k) -?$$

$$f_{Y_1}(y), f_{Y_2}(y) -?$$

Решение:

$$\mathbb{P}(X_{t+\Delta} = 0) = \mathbb{P}(X_t = 0, X_{t+\Delta} - X_t = 0) = \mathbb{P}(X_t = 0) \cdot \mathbb{P}(X_{t+\Delta} - X_t = 0) =$$
$$= \mathbb{P}(X_t = 0) \cdot \mathbb{P}(X_\Delta = 0) = \mathbb{P}(X_t = 0) \cdot (1 - \lambda\Delta + o(\Delta))$$

$$\frac{\mathbb{P}(X_{t+\Delta} = 0) - \mathbb{P}(X_t = 0)}{\Delta} = -\lambda\mathbb{P}(X_t = 0) + \frac{o(\Delta)}{\Delta}$$

$$\frac{dP(X_t = 0)}{dt} = -\lambda\mathbb{P}(X_t = 0)$$

Производная функции содержит искомую функцию. Это экспонента.

$$P(X_t = 0) = c \cdot \exp(-\lambda t)$$

$$f_{Y_1}(y) = \frac{dP(Y_1 \leq y)}{dy} = \frac{dP\left(X_y \geq 1\right)}{dy} = \frac{d\left(1 - \mathbb{P}\left(X_y = 0\right)\right)}{dy} = \frac{d(1 - \exp(-\lambda y))}{dy} = \lambda e^{-\lambda y}, y \geq 0$$

$$\mathbb{P}(X_{t+\Delta}=1)=\mathbb{P}(X_t=1)\cdot\mathbb{P}(X_{t+\Delta}-X_t=0)+\mathbb{P}(X_t=0)\cdot\mathbb{P}(X_{t+\Delta}-X_t=1)$$

$$\mathbb{P}(X_{t+\Delta}=1)=\mathbb{P}(X_t=1)\cdot\mathbb{P}(1-\lambda\Delta+o(\Delta))+e^{-\lambda t}(\lambda\Delta+o(\Delta))$$

$$\frac{\mathbb{P}(X_{t+\Delta}=1)-\mathbb{P}(X_t=1)}{\Delta}=-\lambda\mathbb{P}(X_t=1)+\lambda e^{-\lambda t}+\frac{o(\Delta)}{\Delta}$$

$$\frac{d\mathbb{P}(X_t=1)}{dt}=-\lambda\mathbb{P}(X_t=1)+\lambda e^{-\lambda t}$$

Такие диффуры можно решать в два шага:
1. Решеним уравниение попроще (в данном случае без $\lambda e^{-\lambda t}$)
2. Модифицируем (сдвинем)

So, лайт-версия:

$$\frac{d\mathbb{P}(X_t=1)}{dt}=-\lambda\mathbb{P}(X_t=1)\Rightarrow\mathbb{P}(X_t=1)=c\cdot e^{-\lambda t}$$

Модификация: домножаем на неизвестную функцию $h(t)$

$$\mathbb{P}(X_t=1)=h(t)\cdot e^{-\lambda t}$$

$$\frac{d\mathbb{P}(X_t=1)}{dt}=h'(t)\cdot e^{-\lambda t}+h(t)\cdot\left(-\lambda e^{-\lambda t}\right)$$

Подставляем вместо экспонент известную нам $\mathbb{P}$

$$\frac{d\mathbb{P}(X_t=1)}{dt}=-\lambda\mathbb{P}(X_t=1)+\lambda e^{-\lambda t}$$

Выражаем h'

$$h'(t)e^{-\lambda t}=\lambda e^{-\lambda t}$$

$$h'(t)=\lambda\Rightarrow h(t)=\lambda t+d$$

$$\mathbb{P}(X_t=1)=(\lambda t+d)\cdot e^{-\lambda t}$$

Константы найдем через предел (тут нам и нужна была непрерывность справа):

$$\lim_{t\to 0+}\mathbb{P}(X_t=1)=\lambda t+o(1)=0=d$$

Итого получаем общую формулу:

$$\mathbb{P}(X_t=k)=e^{-\lambda t}\cdot\frac{(\lambda t)^k}{k!}$$

Действительно,

$$\sum_{k=0}^{\infty}\mathbb{P}(X_t=k)=e^{-\lambda t}\cdot\underbrace{\left(1+\lambda t+\frac{\lambda t^2}{2!}+...\right)}_{\text{Тейлор экспоненты}}=e^{-\lambda t}\cdot e^{\lambda t}=1$$

Определение: СВ R имеет распределение Пуассона с интенсивностью λ, если

$$P(R = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Обозначение: $R \sim PP(\lambda)$ или $R \sim PPP(\lambda)$ - Poisson [Point] Process

Теорема/упражнение.

$$\mathbb{E}(X_t) = t\mathbb{E}(X_1)$$

$$\mathbb{V}\mathrm{ar}(X_t) = t\,\mathbb{V}\mathrm{ar}(X_1)$$

$$\mathbb{E}(X_1) = \sum_{k=0}^{\infty} k\mathbb{P}(X_1 = k) = \sum_{k=0}^{\infty} ke^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda \cdot \underbrace{\sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!}}_{\mathbb{P}(\Omega)=1} = \lambda$$

$$\mathbb{V}\mathrm{ar}(X_1) = \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 = \sum_{k=0}^{\infty}\left(k^2 e^{-\lambda} \cdot \frac{\lambda^k}{k!}\right) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$\lambda = \mathbb{E}(X_1)$ есть среднее количество событий за единицу времени [посетителей/час].

9.2 Сумма независимых пуассоновских потоков

Пусть есть два независимых потока посетителей в кофейню «Груша»: студентов (A) и сотрудников (B), с интенсивностью λ_A и λ_B: $X_t^A \sim PP(\lambda_A), X_t^B \sim PP(\lambda_B)$

$S_t = X_t^A + X_t^B \sim PP(\lambda_A + \lambda_B)$. Проверим постулаты, чтобы убедиться, что это действительно поток Пуассона:

1. $S_0 = X_0^A + X_0^B = 0 + 0 = 0$
2. $$P(S_{t+\Delta} - S_t = k) = \sum_{j=0}^{k} \underbrace{\mathbb{P}(X_{t+\Delta}^A - X_t^A = j)}_{\text{независим от t}} \cdot \underbrace{\mathbb{P}(X_{t+\Delta}^B - X_t^B = k - j)}_{\text{независим от t}}$$

3,4 проверяются по аналогии.

Теорема/упражнение. Даны два различных распределения (всё та же Груша)

$$Y^A \sim \text{Expo } (\lambda_A), Y^B \sim \text{Expo } (\lambda_B)$$

$$\mathbb{P}(Y^A < Y^B) - ?$$

Решение: Задача может быть решена классическим образом:

$$\iint\limits_{0<a<b} f(a,b)\,\mathrm{d}a\,\mathrm{d}b = \iint\limits_{0<a<b} \lambda_A e^{-\lambda_A \cdot a} \lambda_B e^{-\lambda_B \cdot b}\,\mathrm{d}a\,\mathrm{d}b = \ldots = \frac{\lambda_A}{\lambda_A + \lambda_B}$$

Того же ответа можно достичь через о-малые. Рассмотрим варианты событий (и их вероятности), которые могут произойти на малом отрезке $[0, \Delta]$:

1. Пришел студент: $o(\Delta) + \lambda_A \Delta(1 - \lambda_B \Delta)$

2. Пришел препод: $o(\Delta) + \lambda_B \Delta (1 - \lambda_A \Delta)$
3. Пришли оба: $1 - \lambda_A \Delta - \lambda_B \Delta + o(\Delta)$
4. Прочее: $o(\Delta)$

Собираем полную вероятность: $p = \lambda_A \Delta + (1 - \lambda_A \Delta - \lambda_B \Delta) p + o(\Delta)$

$$p = \underbrace{\frac{\lambda_A}{\lambda_A + \lambda_B}}_{\text{Не зависит от } \Delta \text{, как и } \mathbb{P}(Y^A < Y^B)} + \underbrace{\frac{o(\Delta)}{\Delta}}_{\to 0}$$

Лекция 10. Гамма-распределение. Бета-распределение.

18 ноября 2024 г.

Нам уже известно достаточно много дискретных распределений: равномерное, Bin, Bern, NBin, Geom, HGeom, PP; но очень мало непрерывных: Unif, Expo. Цель сегодняшней лекции узнать несколько новых, возникающих в задачах, касательных уже известных нам распределений.

Упражнение 1. Поток Пуассона: $Y_1, ..., Y_n \sim Expo(\lambda)$ - независимы; $X_t \sim PP(\lambda)$

$$S := \sum_{i=1}^{a} Y_i;\ \ f_S(s) -?$$

Решение: Мы не будем решать через индукцию и формулу свертки, применим соображения с o-малыми.

$$f_S(s)\,\mathrm{d}s = P(S \in [s, s+\mathrm{d}s]) + \underbrace{o(\mathrm{d}s)}_{\geq 2 \text{ клиентов}}$$

То есть где-то на отрезке $[s, s+\mathrm{d}s]$ пришел клиент a, и с точностью до o(dif s) $a-1$ клиентов пришли до s.

$\mathbb{P}(a$ -й клиент пришел в $[s, s+\mathrm{d}s]) = \mathbb{P}(X_s = a-1)\lambda\,\mathrm{d}s + o(\mathrm{d}s)$

Вспоминая предыдущую лекцию, $f_S(s) = \exp(-\lambda s)\frac{(\lambda s)^{a-1}}{(a-1)!}\lambda$

Упражнение на 10 секунд.

$$\int\limits_0^\infty \lambda\exp(-\lambda s)(\lambda s)^{a-1}\,\mathrm{d}s = ?$$

Решение:

$$\int\limits_0^\infty \lambda\exp(-\lambda s)(\lambda s)^{a-1}\,\mathrm{d}s = (a-1)!$$

Причем, отмасштабировав время $x = \lambda s$, имеем тот же результат:

$$\int\limits_0^\infty \exp(-x)x^{a-1}\,\mathrm{d}x = (a-1)!$$

Определение: Гамма-функция

$$\Gamma(a) = \int\limits_0^\infty \exp(-x)x^{a-1}\,\mathrm{d}x, x \in \mathbb{R}$$

Для $x \in \mathbb{N}$ верно $\Gamma(a) = (a-1)!$

Определение: СВ S имеет Гамма-распределение, если

$$f_S(s) = \begin{cases} \dfrac{\lambda \exp(-\lambda s)(\lambda s)^{a-1}}{\Gamma(a)}, s \geq 0 \\ 0, s < 0 \end{cases}$$

Обозначение: $S \sim Gamma(a, \lambda)$

> **Упражнение.** $\mathbb{E}(S)$ –?

Решение:

$$a \in \mathbb{N} \Rightarrow \mathbb{E}(Y_1 + \dots + Y_a) = a\mathbb{E}(Y_1) = \frac{a}{\lambda}$$

$$a > 0 \Rightarrow \int_0^\infty s\lambda \exp(-\lambda s)(\lambda s)^{a-1} \,\mathrm{d}s = \frac{1}{\Gamma(a)} \int_0^\infty \frac{x}{\lambda} \exp(-x) x^{a-1} \,\mathrm{d}x = \frac{\Gamma(a+1)}{\lambda\Gamma(a)}$$

> **Теорема/упражнение.** Дан поток $X_t \sim PP(\lambda)$, a клиентов пришло к моменту времени $S_a = Y_1 + \dots + Y_a$, еще b клиентов подошло к $S_{a+b} = Y_1 + \dots + Y_{a+b}$
>
> $$R := \frac{S_a}{S_{a+b}};\ f_R(r) \text{ –?}$$

Решение: СВ R есть отношение СВ с Гамма-распределениями. Сведём к интенсивности 1.

$$S \sim Gamma(a, \lambda), \mathbb{E}(S) = \frac{a}{\lambda} \Rightarrow S' = S\lambda \sim Gamma(a, 1), \mathbb{E}(S') = a$$

$$R = \frac{\lambda S'_a}{\lambda S'_{a+b}} \Rightarrow f_R(r) \text{ не зависит от } \lambda$$

Найдем $f_R(r)$ через $f_{R,S_{a+b}}(r,s)$. S_a, S_b зависимы, так что пойдем через $\begin{cases} S'_a = S_1 \\ S'_{a+b} - S_a = S_2{}' \end{cases}$ новые величины независимы.

$$f_{S_1,S_2}(s_1,s_2) = \underbrace{f_{S_1}(s_1)}_{Gamma(a,1)} \cdot \underbrace{f_{S_2}(s_2)}_{Gamma(a,1)} = \exp(-s_1)s_1^{a-1}\exp(-s_2)s_2^{b-1}$$

Совершим переход в новые координаты для последующего снижения размерности:

$$\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \to \begin{bmatrix} R = \frac{S_1}{S_1+S_2} \\ S = S_1 + S_2 \end{bmatrix};\ \begin{cases} S_1 = RS \\ S_2 = S - RS \end{cases}$$

Нахождение якобиана через дифференциальные формы:

$$\mathrm{d}S_1 \wedge \mathrm{d}S_2 = \mathrm{d}(RS) \wedge \mathrm{d}(S - RS) = \mathrm{d}(RS) \wedge \mathrm{d}S = (R\,\mathrm{d}s + S\,\mathrm{d}R) \wedge \mathrm{d}S = S\,\mathrm{d}R \wedge \mathrm{d}S$$

$$f_{R,S}(r,s) = s \cdot f_{S_1,S_2}(rs, s-rs) = s \cdot \exp(-rs)(rs)^{a-1}\exp(rs-s)(s-rs)^{b-1} =$$
$$= r^{a-1}(1-r)^{b-1}\exp(-s)s^{a+b-1} = f_R(r)f_S(s) \text{ (т.к. независимы)}$$

PDF случайной величины S мы знаем, т.к. это сумма Гамма-распределений:

$$f_S(s) = \frac{\exp(-s)s^{a+b-1}}{\Gamma(a+b)} \Rightarrow f_R(r) = \begin{cases} \dfrac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} r^{a-1}(1-r)^{b-1}, r \in [0,1] \\ 0 \text{ иначе} \end{cases}$$

Оказывается, эта СВ имеет распределение, важное для дальнейшего исследования.

Упражнение.

$$\int\limits_0^1 r^{2024}(1-r)^{2025}\,\mathrm{d}r - ?$$

Решение:

$$\int\limits_0^1 r^{2024}(1-r)^{2025}\,\mathrm{d}r = \frac{\Gamma(2025)\Gamma(2026)}{\Gamma(4051)} = \frac{2024!\cdot 2025!}{4050!}$$

Определение: Бета-функция

$$\mathrm{B}(a,b) = \int\limits_0^1 r^{a-1}(1-r)^{b-1}\,\mathrm{d}r = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Бета-функция играет роль нормирующей константы для Бета-распределения.

Определение: СВ R имеет Бета-распределение, если

$$f_R(r) = \begin{cases} \dfrac{r^{a-1}(1-r)^{b-1}}{\mathrm{B}(a,b)}, r \in [0,1] \\ 0 \text{ иначе} \end{cases}$$

Обозначение: $R \sim Beta(a,b)$

Упражнение.

$$\mathbb{E}(R) - ?$$

Решение:

$$a,b \in \mathbb{N} \Rightarrow \mathbb{E}(R) = \mathbb{E}\left(\frac{S_1}{S_1+S_2}\right) = \sum_{i=1}^{a}\mathbb{E}\left(\frac{Y_i}{Y_1+\ldots+Y_{a+b}}\right) = \frac{a}{a+b}$$

$$a, b > 0 \Rightarrow \mathbb{E}(R) = \int\limits_0^1 r f_R(r)\,\mathrm{d}r = \int\limits_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r \cdot r^{a-1} \cdot (1-r)^{b-1}\,\mathrm{d}r =$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int\limits_0^1 r^a \cdot (1-r)^{b-1}\,\mathrm{d}r = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b}$$

Лекция 11. Нормальное распределение

25 ноября 2024 г.

В 1600 году Тихо Браге измерял прямое восхождение Марса. Данные, которые он получил:

- $134° 23' 39''$
- $132° 27' 37''$
- $134° 23' 18''$
- $134° 29' 48''$

Астрономы разделились на тех, кто для вычисления примерного реального значения взял среднее, и тех, кто придумывали различные сложные рассуждения и способы вычисления, отличные от взятия среднего.

В 1810 году Карл Гаусс придумал весьма интересную идею на эту тему:

1. Предположим, что берущие среднее правы.
2. Предположим, что наблюдения независимы как случайные величины, представляющие из себя сумму реального значения и какого-то случайного отклонения.
3. Предположим, что распределения одинаково распределены.
4. Предположим, что ошибки в большую и в меньшую сторону равновероятны.

Говоря языком теории вероятности:

$$f(x_1, ..., x_n \mid \mu) \text{ — совместная функция распределения результатов измерений,}$$

$$x_1, ..., x_n \text{ — результаты измерений}$$

$$\mu \text{ — реальное значение}$$

Тогда вследствие описанных выше посылок

$$f(x_1, ..., x_n \mid \mu) \overset{2}{=} f_1(x_1 \mid \mu) * ... * f_n(x_n \mid \mu) \overset{3}{=} f(x_1 \mid \mu) * ... * f(x_n \mid \mu) \overset{4}{=}$$
$$\overset{4}{=} h(x_1 - \mu) * ... * h(x_n - \mu),$$

где h – чётная функция.

Далее Гаусс предполагает, что полученные результаты измерений – самые вероятные из возможных при данном μ. Так как h – функция плотности отклонения, то h будет такой, что $h(x_1 - \mu) * ... * h(x_n - \mu)$ максимально. При этом мы предполагаем, что реальное значение равно среднему, то есть $\mu^* = \overline{x}$.

Заметим, что

$$\max_{\mu} \ln \prod_{i=1}^{n} h(x_i - \mu) = \max_{\mu} \sum_{i=1}^{n} \underbrace{\ln h(x_i - \mu)}_{=: g(x_i - \mu)} = \max_{\mu} \sum_{i=1}^{n} g(x_i - \mu)$$

Заметим, что g – чётная функция. Тогда

$$\left(\sum_{i=1}^{n} g(x_i - \mu) \right)' = - \sum_{i=1}^{n} g'(x_i - \mu) = 0$$

Тогда для любых результатов измерений $x_1, ..., x_n$ выполняется

$$\sum_{i=1}^{n} g'(x_i - \overline{x}) = 0$$

Пример:

$$x = \begin{pmatrix} 0 \\ 0 \\ 3u \end{pmatrix}, \quad \overline{x} = u$$

$$\begin{pmatrix} x_1 - \overline{x} \\ x_2 - \overline{x} \\ x_3 - \overline{x} \end{pmatrix} = \begin{pmatrix} -u \\ -u \\ 2u \end{pmatrix}$$

$$g'(-u) + g'(-u) + g'(2u) = 0$$

$$-2g'(u) + g'(2u) = 0$$

$$\forall u : g'(2u) = 2 * g'(u)$$

Становится очевидно, что g' – это линейная функция.

$$g'(u) =: -\tau * u$$

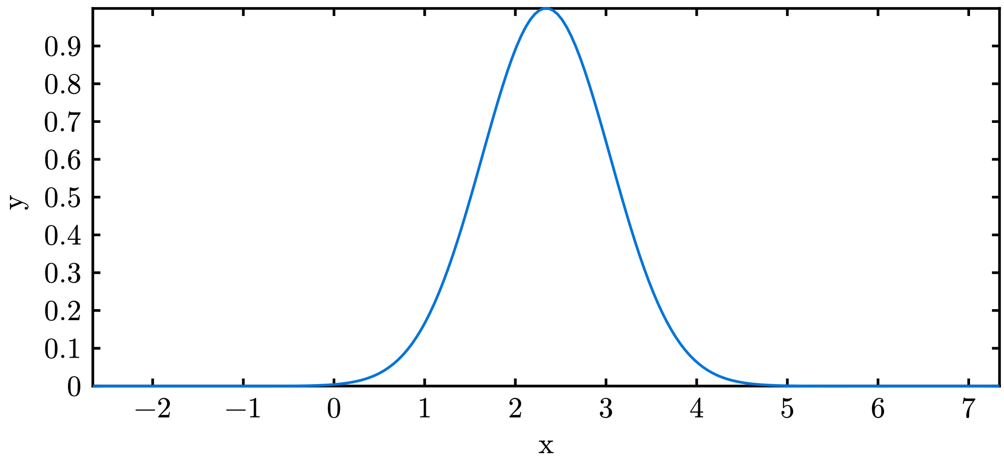
$$g(u) = -\tau * \frac{u^2}{2} + \text{const}$$

$$\Rightarrow f(x \mid \mu) = h(x - \mu) = \exp g(u) = \exp\left(-\frac{\tau * (x-\mu)^2}{2} + \text{const}\right) =$$

$$= \exp\left(-\frac{\tau * (x-\mu)^2}{2}\right) * \text{const}$$

Определение: Случайная величина x имеет **распределение Гаусса** с математическим ожиданием μ и точностью τ, если её функция плотности

$$f(x \mid \mu, \tau) = \exp\left(\frac{-\tau(x-\mu)^2}{2}\right) * \text{const}$$



Попробуем найти константу из функции плотности.

$$\int\limits_{-\infty}^{\infty} \exp\left(\frac{-\tau(x-\mu)^2}{2}\right) * \mathrm{const}\, \mathrm{d}x = 1$$

Посмотрим на единицы измерения:

$$\begin{array}{ll} x & [\text{градус}] \\ \mu & [\text{градус}] \\ \mathrm{d}x & [\text{градус}] \\ \tau & \left[\dfrac{1}{\text{градус}^2}\right] \\ \mathrm{const} & \left[\dfrac{1}{\text{градус}}\right] =: \sqrt{\tau} * c \end{array}$$

$$f(x \mid \tau, \mu) = c * \sqrt{\tau} * \exp\left(\frac{-\tau(x-\mu)^2}{2}\right)$$

Гораздо позже Уильям Гершель вывел эту же формулу, воспользовавшись не явными математическими построениями, а просто рассуждениями. Попробуем примерно повторить его рассуждение.

Предположим, что мы поймали случайно летящую молекулу кислорода и измерили компоненты её скорости. Для простоты предположим, что мы живём в плоском мире, и компоненты этой скорости оказались равны x_1 и x_2. Попытаемся найти совместную функцию плотности $f(x_1, x_2)$.

Сделаем несколько очевидных предположений:

1. Ветра нет, так что $f(x_1, x_2)$ не зависит от направления, потому что иначе все молекулы улетели бы в некотором фиксированном направлении.

   То есть $f(x_1, x_2)$ зависит только от $\|x\|$. $\quad f(x_1, x_2) =: h(x_1^2 + x_2^2)$.

   Распеределение инвариантно к повороту.

2. Компоненты скорости в любых перпендикулярных осях независимы.

Тогда

$$f(x_1, x_2) \overset{2}{=} f_1(x_1) * f_2(x_2)$$

Так как рапределение инвариантно относительно поворота, то выполнив замену переменных $x_1 \sim x_2$, $x_2 \sim -x_1$, получаем:

$$f(x_1, x_2) = f(x_1) * f(x_2)$$

$$\underbrace{h(x_1^2 + x_2^2)}_{f(x_1, x_2)} \overset{1}{=} \underbrace{g(x_1^2)}_{f(x_1)} * \underbrace{g(x_2^2)}_{f(x_2)}$$

$$\forall a, b \geq 0 : h(a + b) = g(a) * g(b)$$

$$g(a + b) = g(a) * g(b) * \mathrm{const}$$

$$\underbrace{\ln g}_{=:\varphi}(a + b) = \ln g(a) + \ln g(b) + \mathrm{const}$$

$$\varphi(a + b) = \varphi(a) + \varphi(b) + \mathrm{const}$$

$$\varphi'_b(a + b) = \varphi'_b(b)$$

$$b = 0 \Rightarrow \varphi'(a) = \mathrm{const}$$

Тогда φ – линейная функция:

$$\varphi(u) = -\frac{\tau}{2} * u + \mathrm{const}$$

$$g(u) = \exp\Big(-\frac{\tau}{2} * u + \mathrm{const}\Big)$$

$$f(x_1, x_2) = c * \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

Мыленно построим график этой функции. Он будет иметь форму горы – срез в любой плоскости $f(x_1, x_2) = v$ будет кругом. Площадь этого круга

$$S = \pi * (-2 \ln v)$$

$$\Rightarrow c = \frac{1}{\int\limits_{v=0}^{1} -2\pi \ln v \, \mathrm{d}v} = \frac{1}{2\pi * \int\limits_{-\infty}^{\infty} e^4 \, \mathrm{d}u} = \frac{1}{2\pi}$$

Таким образом, **по Гершелю-Максвеллу**

$$f(x_1, x_2) = \frac{1}{2\pi} * \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) = f(x_1) * f(x_2)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Определение (предаварительное): Случайная величина имеет **нормальное распределение** со средним μ и точностью τ, если её функция плотности равна

$$f(x, \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} * \exp\left(-\frac{\tau(x-\mu)^2}{2}\right)$$

Попробуем найти дисперсию нормального распределения. Предположим для простоты, что $\mu = 0$ – значение μ не должно влиять на дисперсию.

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) = \\
&= \int\limits_{-\infty}^{\infty})x^2 * \sqrt{\frac{\tau}{2\pi}} * \exp\left(\frac{-\tau x^2}{2}\right) \mathrm{d}x = \\
&= \underbrace{x * \exp\left(-\frac{\tau x^2}{2}\right) * x * \sqrt{\frac{\tau}{2\pi}} * \frac{1}{-\tau}\Big|_{-\infty}^{\infty}}_{0} + \int\limits_{-\infty}^{\infty} \frac{\exp\left(-\frac{\tau x^2}{2}\right)}{\tau} * \sqrt{\frac{\tau}{2\pi}} \, \mathrm{d}x = \\
&= \frac{1}{\tau} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{\tau x^2}{2}\right) * \sqrt{\frac{\tau}{2\pi}} \, \mathrm{d}x = \frac{1}{\tau} * \int\limits_{-\infty}^{\infty} f(x) \, \mathrm{d}x = \frac{1}{\tau}
\end{aligned}$$

$$\tau = \frac{1}{\mathbb{V}\mathrm{ar}(X)} = \frac{1}{\sigma_X^2}$$

Определение: Случайная величина X имеет **нормальное распределение** с $\mathbb{E}(X) = \mu$ и $\mathbb{V}\mathrm{ar}(X) = \sigma^2$, если её функция плотности равна

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

Лекция 12. Многомерное нормальное распределение

2 декабря 2024 г.

Определение: Говорят, что случайный вектор $X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix}$ имеет многомерное стандартное нормальное распределение, если X_i независимы, $\forall 1 \le i \le n \hookrightarrow X_i \sim N(0,1)$, т.е.

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

Обозначение: $X \sim N(0, E)$, 0 - нулевой вектор, E - единичная матрица.

Определение: Если $X \sim N(0, E), Y = \mu + A * X$, то Y имеет многомерное нормальное распределение: $Y \sim N(\mu, AA^T)$

$(X \in \mathrm{Mat}_{n\times 1}(\mathbb{R}), A \in \mathrm{Mat}_{k\times n}(\mathbb{R}), \mu \in \mathrm{Mat}_{k\times 1}(\mathbb{R}), Y \in \mathrm{Mat}_{k\times 1}(\mathbb{R}))$

Если $A \in \mathrm{Mat}_{n\times n}(\mathbb{R}), \det(A) \neq 0$, то Y имеет многомерное невырожденное распределение

> **Упражнение 1.** $\mathbb{E}(Y) -?\quad \mathbb{V}\mathrm{ar}(Y) -?$

Решение:

$$\mathbb{E}(Y) = \mathbb{E}(\mu + AX) = \mu + A\mathbb{E}(X) \text{ (по линейности)}$$

$$\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(\mu + AX) \underbrace{=}_{\mathbb{V}\mathrm{ar}(\mathrm{const})=0} \mathbb{V}\mathrm{ar}(AX) = A\underbrace{\mathbb{V}\mathrm{ar}(X)}_{=E}A^T = AA^T$$

> **Упражнение 2.** Найти функцию плотности $f_Y(y)$

Решение: Смена координат через Якобиан: $f_Y(y) = f_X(x) \cdot \left|\frac{\partial x}{\partial y}\right|$

$$X = A^{-1}(Y - \mu); f(x) = \frac{1}{\sqrt{(2\pi)^n}} \exp\Big(-\frac{1}{2}x^T x\Big); \left|\frac{\partial x}{\partial y}\right| = \det(A^{-1})$$

$$\rceil C = AA^T = \mathbb{V}\mathrm{ar}(Y) \Rightarrow \det(C) = (\det(A))^2$$

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}(y-\mu)^T \underbrace{(A^{-1})^T A^{-1}}_{C^{-1}}(y-\mu)\right)\underbrace{\det(A^{-1})}_{\det(C)^{-\frac{1}{2}}}$$

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\Big(-\frac{1}{2}(y-\mu)^T C^{-1}(y-\mu)\Big), Y \sim N(\mu, C)$$

> **Упражнение 3.** Найти производящие моменты для Y

Решение: Пусть $Y = \begin{bmatrix} Y_1 \\ ... \\ Y_n \end{bmatrix}, U = \begin{bmatrix} u_1 \\ ... \\ u_n \end{bmatrix}$

Функция, производящая моменты $m_Y(U) = \mathbb{E}(\exp(U^T Y)) = \mathbb{E}(\exp(u_1 Y_1 + ... + u_n Y_n))$

Выразим через X: Если $Y \sim N(\mu, C), C = AA^T$, то $A = C^{\frac{1}{2}}$ (обозначение решения уравнения $C = AA^T$ относительно A) ,$Y = \mu + C^{\frac{1}{2}} X \Rightarrow X = C^{-\frac{1}{2}}(Y - \mu)$

$$m_Y(U) = \mathbb{E}\Big(\exp\Big(U^T\Big(\mu + C^{\frac{1}{2}}\Big)\Big)\Big) = \exp(U^T \mu)\mathbb{E}\Big(\exp\Big(U^T C^{\frac{1}{2}} X\Big)\Big) =$$

$$= \exp(U^T\mu) \int\limits_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\Big(-\frac{1}{2}\Big(x^T x - U^T C^{\frac{1}{2}} x - x^T C^{\frac{1}{2}} U + U^T C U - U^T C U\Big)\Big)\,\mathrm{d}x =$$

$$= \exp\Big(U^T\mu + \frac{1}{2} U^T C U\Big) \int\limits_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2}\left(\underbrace{x^T x - U^T C^{\frac{1}{2}} x - x^T C^{\frac{1}{2}} U + U^T C U}_{(x-\nu)^T(x-\nu) = x^T x - \nu^T x - x^T \nu + \nu^T \nu}\right)\right)\mathrm{d}x$$

Функция под интегралом получается как функция плотности некоего другого нормального распределения $\Rightarrow$ под интегралом на всём $\mathbb{R}^n$ значение равно 1. Итого,

$$m_Y(U) = \exp\Big(U^T\mu + \frac{1}{2} U^T C U\Big)$$

Упражнение 4. $\mathbb{E}(Y_1), \mathbb{E}(Y_1 Y_2), \mathbb{E}(Y_1 Y_2 Y_3), ... -?$

Решение: 1) $\mathbb{E}(Y_1) = \mu_1$ из матрицы мат. ожиданий μ. Получим тот же ответ через производящую функцию: $\mathbb{E}(Y_1) = \frac{\partial m(U)}{\partial u_1}\Big|_{U=0}$

$$m_Y(U) = \exp(Q(U)), Q(U) = U^T\mu + \frac{1}{2}U^T C U$$

При дифференцировании по u_1 от первого слагаемого останется только коэффциент при u_1; для вычисления второго посмотрим на двумерный случай:

$$\frac{1}{2}[u_1 \ \ u_2]\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{1}{2}(c_{11}u_1^2 + c_{12}u_1u_2 + c_{21}u_2u_1 + c_{22}u_2^T)$$

Должно остаться $\frac{1}{2}(2c_{11}u_1 + 2c_{12}u_2)$. Суммируя:

$$g_1 = \frac{\partial Q}{\partial u_1} = \mu_1 + C_1^T u; g_2 = \frac{\partial Q}{\partial u_2} = c_{12}$$

$$\frac{\partial m(U)}{\partial u_1}\Big|_{U=0} = \exp(Q(U)) g_1 = 1 \cdot \mu_1$$

$$\mathbb{E}(Y_1 Y_2) = \frac{\partial^2 m(U)}{\partial u_1 \partial u_2}\Big|_{U=0} = \frac{\partial(\exp(Q(U))g_1)}{\partial u_2} = \exp(Q(U)) \cdot (c_{12} + g_1 g_2) \overset{u=0}{=} 1 \cdot (c_{12} + \mu_1\mu_2)$$

$$\mathbb{E}(Y_1 Y_2 Y_3) = \exp(Q(U))(0 + c_{13}g_2 + c_{23}g_1 + c_{12}g_3 + g_1g_2g_3) = c_{13}\mu_2 + c_{23}\mu_1 + c_{12}\mu_3 + \mu_1\mu_2\mu_3$$

Короче говоря, мат ожидание произведения k элементов случайного вектора есть сумма всевозможных мономов из элементов C и Y, чтобы в каждом все k индексов встречались по 1 разу.

> **Задача.** Найти условное распределение $(Y_1 \mid Y_2)$ для СВ из векторов $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$
> $Y \sim N(\mathrm{M}, C) = N\left(\begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\right)$ (все матрицы блочные и невырожденные)

Решение:

$$f_{Y_1 \mid Y_2} = \frac{f_{Y_1, Y_2}(y_1, y_2)}{f_{Y_2}(y_2)}$$

Избавимся от M : $\begin{cases} Y_1 = \mathrm{M}_1 + X_1 \\ Y_2 = \mathrm{M}_2 + X_2 \end{cases}$, $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbb{O} \\ \mathbb{O} \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\right)$ ($\mathbb{O}$ - нулевой вектор, а не просто 0)

Теперь будем искать распределение $(X_1 \mid X_2)$

$$f_{X_1 \mid X_2} = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{\mathrm{const} \exp\left(-\frac{1}{2} x^T C^{-1} x\right)}{\mathrm{const} \exp\left(-\frac{1}{2} x_2^T C_{22}^{-1} x_2\right)} = \mathrm{const} \exp\left(-\frac{1}{2} (\ldots)\right)$$

Вновь получается некоторое нормальное распределение. Его параметры можно найти через выделение полных квадратов и прочей алгебры. Мы же воспользуемся вероятностным подходом.

$$\rceil X_1 = A X_2 + R, \mathbb{C}\mathrm{ov}(X_2, R) = 0; X_1, R \in \mathrm{Mat}_{k_1 \times 1}(\mathbb{R}), X_2 \in \mathrm{Mat}_{k_2 \times 1}(\mathbb{R}), A \in \mathrm{Mat}_{k_1 \times k_2}(\mathbb{R})$$

$$\mathbb{C}\mathrm{ov}(X_1, X_2) = \mathbb{C}\mathrm{ov}(A X_2 + R, X_2)$$

$$C_{12} = \mathbb{C}\mathrm{ov}(A X_2 + R, X_2) = A\, \mathbb{C}\mathrm{ov}(X_2, X_2) \Rightarrow A = C_{12} C_{22}^{-1}$$

$$\mathbb{V}\mathrm{ar}(X_1) = \mathbb{V}\mathrm{ar}(A X_2 + R) \underbrace{=}_{\mathrm{Cov}=0} \mathbb{V}\mathrm{ar}(A X_2) + \mathbb{V}\mathrm{ar}(R)$$

$$C_{11} = A C_{22} A^T + \mathbb{V}\mathrm{ar}(R)$$

$$C_{11} = C_{12} C_{22}^{-1} C_{22} C_{22}^{-1} C_{21} + \mathbb{V}\mathrm{ar}(R)$$

$$C_{\mathbb{R}} := \mathbb{V}\mathrm{ar}(R) = C_{11} - C_{12} C_{22}^{-1} C_{21} \Rightarrow$$

$$(X_1 \mid X_2) \sim N(C_{12} C_{22}^{-1} X_2, C_{11} - C_{12} C_{22}^{-1} C_{21})$$

Осталось восстановить закон распределения $(Y_1 \mid Y_2)$, сдвинув на M

$$(Y_1 \mid Y_2) \sim N(\mathrm{M}_1 + C_{12} C_{22}^{-1} (Y_2 - \mathrm{M}_2), C_{11} - C_{12} C_{22}^{-1} C_{21})$$

Лекция 13. Сходимость по распределению

9 декабря 2024 г.

Сходимость по распределению – не единственный вид сходимости в теории вероятностей, но при этом самый слабый, то есть следует из всех остальных видов сходимости.

Определение: Последовательность случайных величин R_n **сходится по распределению** к случайной величине R, если $\mathbb{P}(R_n \le x) = F_n(x) \to \mathbb{P}(R \le x) = F(x)$ в точках непрерывности $F(x)$.

Обозначение:

$$R_n \xrightarrow{\text{dist}} R$$

Пример (1):

$$R_n \sim \mathrm{Unif}\left[0; 2 + \frac{1}{n}\right] \xrightarrow{\text{dist}} R \sim \mathrm{Unif}[0; 2]$$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & x \in [0; 2] \\ 1, & x > 2 \end{cases}$$

$$F_n(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2 + \frac{1}{n}}x, & x \in \left[0; 2 + \frac{1}{n}\right] \\ 1, & x > 2 + \frac{1}{n} \end{cases}$$

$$F_n \longrightarrow F \quad - \quad \text{очевидно.}$$

Пример (2):

$$R_n \sim \mathrm{Unif}\left[0; \frac{1}{n}\right] \xrightarrow{\text{dist}} R = 0$$

$$F_n(x) = \begin{cases} 0, & x < 0 \\ nx, & x < \frac{1}{n} \\ 1, & x \ge \frac{1}{n} \end{cases}$$

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$$

$$F_n \longrightarrow F$$

Утверждение: $R_n \xrightarrow{\text{dist}} R$ если и только если $\mathbb{E}(h(R_n)) \longrightarrow \mathbb{E}(h(R))$ для достаточно широкого (но не слишком широкого) класса функций h.

Рассмотрим некоторые подходящие классы для функции h:

- $H = \{h(r) \mid \exists x : h(r) = I(r \leq x)\}$
- H – все непрерывные ограниченные функции
- H – все бесконечно дифференцируемые функции, у которых каждая производная ограничена.

Пол доказательства: $H = \left\{\begin{matrix} \text{все бесконечное количество раз дифференцируемые} \\ \text{функции с ограниченными производными} \\ \end{matrix}\right\}$

$$\daleth h \in H, \quad \mathbb{E}(h(R_n)) \to \mathbb{E}(h(R))$$

Цель: доказать, что $\mathbb{E}(I(R_n \leq x)) \longrightarrow \mathbb{E}(I(R \leq x))$.

Пусть x – некоторая точка непрерывности F. Тогда

$$\forall \varepsilon > 0 : \exists s > 0 : \forall t \in [x - s; x] : F(t) \geq F(x) - \varepsilon.$$

$$\exists h \in H : \forall r : I(r \leq x - s) \leq h(r) \leq I(r \leq x)$$

$$I(R_n \leq x) \geq h(R_n)$$

$$F_n(x) = \mathbb{E}(I(R_n \leq x)) \geq \mathbb{E}(h(R_n)) \overset{n\to\infty}{\longrightarrow} \mathbb{E}(I(R \leq x - s)) = F(x - s) \geq F(x) - \varepsilon$$

$$\Rightarrow F_n(x) \geq F(x) - 2\varepsilon$$

Повторив это доказательство для $x + s$ вместо $x - s$, получим, что

$$F_n(x) \leq F(x) + 2\varepsilon$$

Что доказывает, что $F_n \longrightarrow F$.

■

Теорема (Центральная предельная теорема): Если случайные величины Q_1, Q_2 независимы и одинаково распределены с $\mathbb{E}(Q_i) = \mu$ и $\mathbb{V}\mathrm{ar}(Q_i) = \sigma^2$, то

$$R_n = \frac{\sum_{i=1}^{n} Q_i - \mathbb{E}(\sum Q_i)}{\sqrt{\mathbb{V}\mathrm{ar}(\sum Q_i)}} \overset{\text{dist}}{\longrightarrow} R \sim N(0;1)$$

Неформальное объяснение: если сложить достаточно много случайных величин, то полученное распределение будет примерно нормальным.

Доказательство:

$$\begin{aligned} R_n &= \frac{Q_1 - \mu + Q_2 - \mu + ... + Q_n - \mu}{\sqrt{n * \sigma^2}} = \\ &= \underbrace{\frac{Q_1 - \mu}{\sigma\sqrt{n}}}_{X_1} + \underbrace{\frac{Q_2 - \mu}{\sigma\sqrt{n}}}_{X_2} + ... + \underbrace{\frac{Q_n - \mu}{\sigma\sqrt{n}}}_{X_n} = \\ &= X_1 + ... + X_n \end{aligned}$$

$$X_i = \frac{Q_i - \mu}{\sigma\sqrt{n}} \text{ — независимы и одинаково распределены}$$

$$\mathbb{E}(X_i) = 0, \quad \mathbb{V}\mathrm{ar}(X_i) = \frac{\sigma^2}{\sigma^2 n} = \frac{1}{n}$$

$$Y_i \sim N\left(0; \frac{1}{n}\right) \text{ и независимы}$$

Далее будем выполнять подмену слагаемых. Предположим, что $n = 5$, тогда это будет выглядеть как-то так:

$$\overbrace{X_1 + X_2 + X_3 + X_4 + X_5}^{Z_{5,5}} \longrightarrow \overbrace{X_1 + X_2 + X_3 + X_4 + 0}^{S_{5,5}} \longrightarrow \overbrace{X_1 + X_2 + X_3 + X_4 + Y_5}^{Z_{5,4}} \longrightarrow$$
$$\longrightarrow \underbrace{X_1 + X_2 + X_3 + 0 + Y_5}_{S_{5,4}} \longrightarrow \underbrace{X_1 + X_2 + X_3 + Y_4 + Y_5}_{Z_{5,3}} \longrightarrow \ldots$$
$$\ldots \to \underbrace{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}_{Z_{5,0}}$$

Нам надо доказать, что $\mathbb{E}\big(h(Z_{n,n})\big) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\big(h(Z_{n,0})\big) \sim N(0;1)$, $h \in H$.

Индекс n, обозначающий общее число слагаемых, писать не будем, потому что и так понятно.

Для доказательства посмотрим на ошибку на каждом шаге (на каждой комбинации двух полушагов) с подменой и покажем, что общая ошибка $\left|\underbrace{E\big(h(Z_{n,n})\big) - E\big(h(S_{n,n})\big) + E\big(h(S_{n,n})\big) - E\big(h(Z_{n,n-1})\big)}_{*} + \ldots\right| \overset{n\to\infty}{\longrightarrow} 0$, т.е. что по определению предела $* < \frac{\varepsilon}{n}$

$$h(Z_i) = h(S_i) + h'(S_i) * X_i + \underbrace{\frac{h''(C)}{2!} * X_i^2}_{\substack{\text{остаточный член в форме Лагранжа,} \\ C \text{ между } S_i \text{ и } Z_i}}$$

$$h(Z_{i-1}) = h(S_i) + h'(S_i) * Y_i + \frac{h''(D)}{2!} Y_i^2, \quad D \text{ между } S_i \text{ и } Z_{i-1}$$

$$S_i \text{ не зависит от } X_i \text{и} Y_i$$

$$h(Z_i) - h(S_i) = h'(S_i) * X_i + \frac{h''(S_i)}{2!} X_i^2 + \frac{h''(C) - h''(S_i)}{2!} X_i^2$$

$$\mathbb{E}(h(Z_i) - h(S_i)) = \mathbb{E}\left(\underbrace{h'(S_i) * X_i}_{\text{независимы}}\right) + \mathbb{E}\left(\frac{h''(S_i)}{2!} * X_i^2\right) + \mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} X_i^2\right) =$$
$$= \mathbb{E}(h'(S_i)) * \underbrace{\mathbb{E}(X_i)}_{0} + \mathbb{E}\left(\frac{h''(S_i)}{2!}\right) * \mathbb{E}(X_i^2) + \mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} X_i^2\right) =$$
$$= \mathbb{E}\left(\frac{h''(S_i)}{2}\right) * \frac{1}{n} + \mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} X_i^2\right)$$

$$\mathbb{E}(h(Z_{i-1}) - h(S_i)) = 0 + \mathbb{E}\left(\frac{h''(S_i)}{2}\right) * \frac{1}{n} + \mathbb{E}\left(\frac{h''(D) - h''(S_i)}{2!} Y_i^2\right)$$

Так как это слагаемое идёт с минусом в $*$, то $\mathbb{E}\left(\frac{h''(S_i)}{2}\right) * \frac{1}{n}$ сократятся

Далее нам достаточно доказать, что

$$\forall \varepsilon > 0 : \exists N : \forall n > N : \mathbb{E}\left(\left|\frac{h''(C) - h''(S_i)}{2!}\right| X_i^2\right) < \frac{\varepsilon}{2^n}$$

Если мы докажем это для произвольного X_i, будет справедливо и для нормального Y_i в $\mathbb{E}\left(\left|\frac{h''(D)-h''(S_i)}{2!}\right| Y_i^2\right)$

Введём индикаторы «хорошего» и «плохого» случая значений X_i относительно некоторого δ, чтобы в хорошем случае $h''(C) - h''(S_i) < \frac{\varepsilon}{2}$:

$$B = \{|x_i| > \delta\}, G = \{|x_i| \leq \delta\} \Rightarrow |x_i| = |x_i| * I(B) + |x_i| * I(G)$$

Вспомним, что $X_i = Z_i - S_i$, тогда

$$\mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} X_i^2\right) = \underbrace{\mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} * X_i^2 * I(B)\right)}_{*} + \underbrace{\mathbb{E}\left(\frac{h''(C) - h''(S_i)}{2!} * X_i^2 * I(G)\right)}_{*},$$

Так как h''' ограничены, то h'' на коротком участке не могут сильно измениться, так что будет выполняться

$$* \leq \mathbb{E}\left(\frac{\varepsilon}{4} X_i^2\right) = \frac{\varepsilon}{4n}$$

Так как h'' ограничена, то

$$\left|\frac{h''(C) - h''(S_i)}{2!}\right| \leq M$$

$$* \leq M\mathbb{E}(X_i^2 * I(B))$$

Напомню, мы не знаем про X_i ничего, кроме как $\mathbb{E}(X_i) = 0, \mathbb{V}\text{ar}(X_i) = \frac{1}{n}$. Введем V_i

$$X_i = \frac{V_i}{\sqrt{n}}, \quad \mathbb{E}(V_i) = 0, \mathbb{V}\text{ar}(V_i) = 1$$

$$\mathbb{E}(X_i^2 I(B)) = \mathbb{E}\left(\frac{V_i^2}{n} * I\left(|V_i| > \sqrt{n}\delta\right)\right) = \frac{1}{n} * \mathbb{E}\left(V_i^2 * I\left(|V_i| > \delta * \sqrt{n}\right)\right)$$

Для окончания доказательства нам нужен доп факт.

Теорема (DCT - теорема Лебега о мажорируемой сходимости):

$$\begin{cases} L_n \to L \text{ поточечно} \\ \exists G : |L_n| < G \wedge E(G) < \infty (\forall \omega \in \Omega) \end{cases} \Rightarrow \mathbb{E}(L_n) \to E(L)$$

Доказывать мы её, конечно же, не будем. Просто применим, положив $L_n = V_i^2 \cdot I(|V_i| > \delta \cdot \sqrt{n})$, тогда $L_{n(\omega)} \xrightarrow{n\to\infty} 0$. Также ясно, что $L_n \leq V_i^2, E(V_i^2) = 1$

Короче говоря, все промежуточные слагаемые мы устремили к 0, тогда и их сумма попадает в требуемые пределы. ■

Лекция 14. Большая сила о-малых

16 декабря 2024 г.

Последняя лекция в семестре посвящена демонстрации элегантных способов решения задач на нахождение мат. ожидания с использованием о-малых вместо классических с диффурами или метода первого шага.

Упражнение 1.

$$\underbrace{X_1, X_2}_{\text{независимы}} \sim \mathrm{Unif}[0;1]$$

$$\mathbb{E}(|X_1 - X_2|) - ?$$

Решение: Рассмотрим несколько более общий случай:

$$\underbrace{X_1, X_2}_{\text{независимы}} \sim \mathrm{Unif}[0;a]$$

$$\mathbb{E}(|X_1 - X_2|) =: g(a)$$

Будем пользоваться идеей малого приращения интервала на h

$$g(a+h) = \underbrace{\left(\frac{a}{a+h}\right)^2 * g(a)}_{X_1, X_2 \in [0;a]} + \underbrace{\left(\frac{h}{a}+h\right)^2 * O(h)}_{X_1, X_2 \in [a;a+h]} + \underbrace{2 * \frac{a}{a+h} * \frac{h}{a+h} * \left(\frac{a}{2} + O(h)\right)}_{\substack{X_1 \in [0;a], X_2 \in [a;a+h] \\ X_2 \in [0;a], X_1 \in [a;a+h]}}$$

Во втором слагаемом при малом h можно принять $O(h) = o(1)$

$$\Rightarrow g(a+h) = \left(\frac{a}{a+h}\right)^2 * g(a) + \left(\frac{h}{a}+h\right)^2 * o(1) + 2 * \frac{a}{a+h} * \frac{h}{a+h} * \left(\frac{a}{2} + o(1)\right)$$

$$\frac{a}{a+h} = \frac{1}{1+\frac{h}{a}} = 1 - \frac{h}{a} + o(h) \text{ (как геом. прогрессия)}$$

$$g(a+h) = \left(1 - \frac{2h}{a}\right) * g(a) + 2 * \frac{h}{a} * \frac{a}{2} + o(h)$$

$$g(a+h) - g(a) = -\frac{2h}{a} * g(a) + h + o(h)$$

$$\frac{g(a+h) - g(a)}{h} = -\frac{2g(a)}{a} + 1 + o(1)$$

$$h \to 0 \quad \Rightarrow g'(a) = -\frac{2g(a)}{a} + 1$$

Требуется вычислить $g(a)$. Вместо интегрирования предлагается угадать, исходя из предположения, что функция похожа на линейную. Имеем $g'(a) = k \Rightarrow g(a) = ka + b$

$$k = -\frac{2(ka+b)}{a} + 1 \Rightarrow k = -2k + 1 - \frac{2b}{a} \forall a, \text{ отсюда подбираем } b = 0, k = \frac{1}{3}$$

Ответ: $g(a) = \mathbb{E}(|X_1 - X_2|) = \frac{1}{3}a$.

Упражнение 2.

$$\underbrace{X_1, X_2, ..., X_{100}}_{\text{независимы}} \sim \text{Unif}[0;1]$$

$$\mathbb{E}(\max(X_1, ..., X_{100})) -?$$

Решение: Обобщим до $\underbrace{X_1, X_2, ..., X_{100}}_{\text{независимы}} \sim \text{Unif}[0;a]$, так $\mathbb{E}(\max(...)) =: g(a)$

$$g(a+h) = \underbrace{\left(\frac{a}{a+h}\right)^{100} \cdot g(a)}_{\text{все СВ попали в [0, a]}} + \underbrace{100\left(\frac{a}{a+h}\right)^{99} \cdot \frac{h}{a+h} \cdot (a + o(1))}_{\text{ровно одна СВ попала в [a, a+h]}} + o(h)$$

Во втором слагаемом $a + o(1)$ - это значение $\mathbb{E}(\max(...))$, поскольку это будет значение той единственной СВ, которая попала в [a, a+h] при $h \to 0$

Оставшиеся варианты распределения ста величин (в отрезок приращения попало больше СВ) мы засунули в $o(h)$, т.к. при $h \to 0$ в силу большей степени h их порядок пренебрежимо мал.

$$\left(\frac{a}{a+h}\right)^{100} = \left(1 - \frac{h}{a} + ...\right)^{100} = 1 - 100 \cdot \frac{h}{a} + o(h)$$

$$\frac{g(a+h) - g(a)}{h} = \frac{g(a)}{h} \cdot \left(\left(\frac{a}{a+h}\right)^{100} - 1\right) + 100 \cdot \frac{a^{99}}{(a+h)^{100}} \cdot (a + o(1)) + \frac{o(h)}{h}$$

$$g'(a) = -100\frac{g(a)}{h} + 100$$

Вновь подбираем линейную функцию: $g(a) = ka \Rightarrow k = -100k + 100 \Rightarrow k = \frac{100}{101}$

Ответ: $\mathbb{E}(\max(X_1, ..., X_n)) = \frac{n}{n+1} \cdot a$.

Упражнение 2.1. Те же условия, но требуется найти $\mathbb{V}\text{ar}(\max(...))$

Решение:

$$\rbrack \max(...) = R \Rightarrow \mathbb{V}\text{ar}(R) = \mathbb{E}(R^2) - (E(R))^2, \mathbb{E}(R^2) = q(a)$$

Второе слагаемое уже посчитали, первое вычисляется по аналогии:

$$q(a+h) = \left(\frac{a}{a+h}\right)^{100} q(a) + 100 \cdot \frac{h}{a+h} \cdot \left(\frac{a}{a+h}\right)^{99} \cdot (a + o(1))^2 + o(h)$$

$$...$$

> **Упражнение 2.2.** Те же условия, другой способ

Решение: Рассмотрим другой эксперимент: $\underbrace{X_1, X_2, ..., X_{100}, X_{101}}_{\text{независимы}} \sim \mathrm{Unif}[c], c$ - окружность единичной длины. Пронумеруем точки против часовой стрелки, начав с X_{101} как нуля. Тогда СВ $X_1, ..., X_{100}$ распределены равномерно на $[0, 1]$, как и в исходном эксперименте. Теперь введём случайные величины Y_i - длины дуг от $(i-1)$-й до i-й точки. Тогда $\mathbb{E}(\max(...)) = \mathbb{E}(Y_1 + ... + Y_{100})$

Y_i зависимы между собой, но всегда $Y_1 + ... + Y_{101} = \mathbb{E}(Y_1 + ... + Y_{101}) = 1 \Rightarrow \mathbb{E}(Y_i) = \frac{1}{101}$

Ответ: $\mathbb{E}(\max(...)) = \frac{100}{101}$.

> **Упражнение 1.1.** Первое упражнение можно решить способом через окружность.

Решение: Расставим 3 точки на окружности, введём Y_i с тем же смыслом. Имеем $\mathbb{E}(|X_1 - X_2|) = \mathbb{E}(Y_2) = \frac{1}{3}$

Ответ: $\frac{1}{3}$.

> **Упражнение 3.** Имеется тщательно перемешанная колода из 52 карт. По одной (по очереди) из неё извлекаются карты до выпадения 2-го туза. СВ N - номер этого второго туза в колоде.
>
> $\mathbb{E}(N) - ?$

Решение: Решение с помощью окружности. Добавляем 53-ю карту (Джокера) и располагаем все карты на окружности. Вводим Y_i как кол-ва карт между i-1 и i-м тузом, то есть разность между номером i-го туза и $(i-1)$-го (Y_1 есть просто номер первого туза, или расстояние от него до джокера, Y_5 есть расстояние от последнего джокера до конца колоды)

Имеем $\mathbb{E}(Y_1 + ... + Y_5) = 53$ (сумма растояний всегда даёт всю колоду) $\Rightarrow \mathbb{E}(N) = \mathbb{E}(Y_1 + Y_2) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) = \frac{2}{5} \cdot 53$

Ответ: $\frac{106}{5}$.

> **Упражнение 3.1.** Используя все те же данные, найти $\mathbb{C}\mathrm{ov}(Y_1, Y_2)$

Решение: Для начала сразу можно интуитивно догадаться, что искомая ковариация имеет знак минус. Это поможет нам оценить адекватность полученного ответа.

С одной стороны, $\mathbb{V}\mathrm{ar}(Y_1 + ... + Y_5) = \mathbb{V}\mathrm{ar}(53) = 0$, с другой,

$$\mathbb{V}\mathrm{ar}(Y_1 + ... + Y_5) = \sum_{i=1}^{5} + 2 \cdot \sum_{1 \leq i < j \leq 5} \mathbb{C}\mathrm{ov}(Y_i, Y_j)$$

$$5\,\mathbb{V}\mathrm{ar}(Y_1) + 20\,\mathbb{C}\mathrm{ov}(Y_1, Y_2) = 0$$

$$\mathbb{C}\mathrm{ov}(Y_1, Y_2) = -\frac{1}{4}\,\mathbb{V}\mathrm{ar}(Y_1)$$

Из неизвестного остается $E(Y_1^2)$. Для вычисления есть варианты тупого суммирования, метода первого шага или использования удачного условия. Мы же применяем производящую функцию $h(t,k)$, первый аргумент отвечает за тузы, второй за всю колоду

$$h(t,k) = \frac{t}{k} \cdot 1^2 + \left(1 - \frac{t}{k}\right) \cdot \mathbb{E}\Big(\big(Y_{1,t,k-1} + 1\big)^2\Big)$$

$$h(t,k) = \frac{t}{k} + \left(1 - \frac{t}{k}\right) \cdot \left(h(t,k-1) + 1 + 2 \cdot \frac{1}{t+1} \cdot k\right)$$

Снова возникает функция $g(k)$ (t мы знаем и фиксируем), её необходимо найти, но эту задачу мы оставляем читателю в качестве упражнения, так как на лекции дорешать не успели:)

Лекция 15. Сходимость по вероятности

13 января 2025 г.

15.1 Часто задаваемые вопросы

«Зачем ещё одна сходимость?»

Из прошлых лекций нам знакомо понятие *сходимости по распределению*:

$$R_n \xrightarrow{\text{dist}} R \quad \Leftrightarrow \quad \lim \underbrace{F_n(t)}_{\text{функция распределения } R_n} = \overbrace{F(t)}^{\text{функция распределения } R} \quad \text{в точках непрерывности } F$$

Но сходимость по распределению обладает некоторыми недостатками, делающими её не всегда удобной в использовании. Например:

- $\lim a_n + \lim b_n = \lim(a_n + b_n)$ для числовых последовательностей, но для сходимости случайных величин по распределению аналогичное условие не всегда выполняется.

  Пример:

v	0	1
$\mathbb{P}(R_i = v)$	$\frac{1}{2}$	$\frac{1}{2}$

$$F_n(t) = \mathbb{P}(R_n \leq t) = \begin{cases} 0, & t \in (-\infty; 0) \\ \dfrac{1}{2}, & t \in [0; 1) \\ 1, & t \in [1; +\infty) \end{cases}$$

$$(R_n) \xrightarrow{\text{dist}} R_1$$

$$S_n := 1 - R_n$$

v	0	1
$\mathbb{P}(S_i = v)$	$\frac{1}{2}$	$\frac{1}{2}$

$$(S_n) \xrightarrow{\text{dist}} R_1$$

$$S_n + R_n \not\xrightarrow{\text{dist}} 2R_1$$

- $\left.\begin{matrix} \lim a_n = a \\ \lim a_n = b \end{matrix}\right\} \Rightarrow a = b$ для числовых последовательностей, но для сходимости случайных величин по распределению аналогичное условие также не всегда выполняется.

  Пример: Из предыдущего примера:

$$R_n \xrightarrow{\text{dist}} R_1$$

$$R_n \xrightarrow{\text{dist}} S_1$$

$$R_1 \neq S_1 = 1 - R_1$$

«О нет, какие ещё пределы в разных смыслах?»

На самом деле пределы в разном смысле есть даже в неслучайном математическом анализе.

Пример: Для числовых последовательностей есть такое понятие как *предел в смысле Чезаро*:

$$\operatorname{ceslim} a_n = a \quad \Leftrightarrow \quad \lim \frac{a_1 + \ldots + a_n}{n} = a$$

15.2 Обобщение сходимости по распределению на векторы

$$\exists (X_n) - \text{последовательность случайных величин из } \mathbb{R}^d$$

Определение: Будем говорить, что последовательность X_n **сходится по распределению** к X, если для любой непрерывной ограниченной функции h выполнено

$$\lim \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$$

Теорема (1): Новое определение сходимости по распределению при $d = 1$ эквивалентно старому определению.

Теорема (2): Если $\binom{X_n}{Y_n} \xrightarrow{\text{dist}} \binom{X}{Y}$, то $X_n + Y_n \xrightarrow{\text{dist}} X + Y$.

Доказательство:

$$\operatorname{add}(X_n, Y_n) \coloneqq X_n + Y_n$$

$$\operatorname{add} - \text{непрерывная функция}$$

Тогда

$$\forall h : h \text{ непрерывна и ограничена} \Rightarrow \tilde{h} = h \circ \operatorname{add} \text{ непрерывна и ограничена}$$

$$h(X_n + Y_n) = h(\operatorname{add}(X_n, Y_n)) = \tilde{h}\left(\begin{bmatrix} X_n \\ Y_n \end{bmatrix}\right)$$

$$\Rightarrow \mathbb{E}(h(X_n + Y_n)) = \mathbb{E}\left(\tilde{h}\left(\begin{bmatrix} X_n \\ Y_n \end{bmatrix}\right)\right) \longrightarrow \mathbb{E}\left(\tilde{h}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)\right)$$

∎

15.3 Сходимость по вероятности

Определение:

$$]R_n - \text{случайные векторы из } \mathbb{R}^d, \quad R - \text{случайный вектор из } \mathbb{R}^d$$

Будем говорить, что R_n **сходится по вероятности** к R, если

$$\lim_{n\to\infty} \mathbb{P}(\|R_n - R\| > \varepsilon) = 0$$

Обозначение:

$$R_n \xrightarrow{\mathbb{P}} R$$

$$R_n \xrightarrow{\text{prob}} R$$

$$\operatorname{plim} R_n = R$$

Пример:

> $$X \sim \text{Unif}[0;1]$$
>
> $$R_n = \frac{X}{n}$$
>
> $$\operatorname{plim} R_n - ?$$

Решение: Проверим, что $\operatorname{plim} = 0$

$$\mathbb{P}(\|R_n - 0\| > \varepsilon) = \mathbb{P}\left(\left\|\frac{X}{n} - 0\right\| > \varepsilon\right) = \mathbb{P}(X > n * \varepsilon) \xrightarrow{n\to\infty} 0$$

$$\Rightarrow \operatorname{plim} R_n = 0$$

15.4 Свойства сходимости по вероятности

Сходимость по вероятности обобщает сходимость по распределению

Теорема: Если $\operatorname{plim} R_n = R$, то $R_n \xrightarrow{\text{dist}} R$.

Доказательство:

$$\operatorname{plim} R_n = R$$

$$\Rightarrow \forall \varepsilon > 0 : \lim \mathbb{P}(\|\mathbb{R}_n - R\| > \varepsilon) = 0$$

Докажем, что для любой непрерывной ограниченной функции h выполняется $\operatorname{plim} h(R_n) = h(R)$.

$$\|h(R_n) - h(R)\| > \varepsilon \Rightarrow \|R_n - R\| > \delta$$

$$\Rightarrow \mathbb{P}(\|h(R_n) - h(R)\| > \varepsilon) \leq \mathbb{P}(\|R_n - R\| > \delta) \to 0 \quad \forall \delta$$

$$\Rightarrow \operatorname{plim} h(R_n) = h(R)$$

$$A \coloneqq \{|h(R_n) - h(R)| > \varepsilon\}$$

$$|\mathbb{E}(h(R_n)) - \mathbb{E}(h(R))| \underbrace{\le}_{\text{по неравенству Йенсена}} \mathbb{E}(|h(R_n) - h(R)|) = \mathbb{E}(|h(R_n) - h(R)| * (I_A + I_{A^c})) =$$
$$= \mathbb{E}(|h(R_n) - h(R)| * I_A) + \mathbb{E}(|h(R_n) - h(R)| * I_{A^c}) \le 2 * K * \underbrace{\mathbb{P}(A)}_{\xrightarrow{n\to\infty} 0} + \varepsilon * \mathbb{P}(A^c) < 2 * \varepsilon$$

$$\Rightarrow \mathbb{E}(|h(R_n) - h(R)|) \to 0$$

$$\Rightarrow R_n \xrightarrow{\text{dist}} R$$

■

Пример:

$$\rceil R_n \text{ — независимые случайные величины}$$

$$R_n \sim \operatorname{Bern}\left(\frac{1}{2}\right)$$

v	0	1
$\mathbb{P}(R_n = v)$	$\frac{1}{2}$	$\frac{1}{2}$

$$R_n \xrightarrow{\text{dist}} R_1$$

$$\rceil\varepsilon = 0.1$$

$$\mathbb{P}(|R_n - R_1| > \varepsilon) = \frac{1}{2} \not\to 0$$

	$R_1 = 0$	$R_1 = 1$
$R_{100} = 0$	$\frac{1}{4}$	$\frac{1}{4}$
$R_{100} = 1$	$\frac{1}{4}$	$\frac{1}{4}$

$$\Rightarrow R_n \overset{\text{prob}}{\not\to} R_1$$

Единственность предела по вероятности в вероятностном смысле

К сожалению, свойство единственности предела в смысле равенства случайных величин мы так и не получили. Но чуть менее сильное условие, тем не менее, выполняется.

Теорема: Если $\operatorname{plim} R_n = X$ и $\operatorname{plim} R_n = Y$, то $\mathbb{P}(X = Y) = 1$.

Доказательство:

$$|X - Y| \ge \varepsilon \Rightarrow |X - R_n| \ge \frac{\varepsilon}{2} \oplus |Y - R_n| \ge \frac{\varepsilon}{2}$$

$$\mathbb{P}(|X - Y| \geq \varepsilon) \leq \mathbb{P}\Big(|X - R_n| \geq \frac{\varepsilon}{2}\Big) + \mathbb{P}\Big(|Y - R_n| \geq \frac{\varepsilon}{2}\Big)$$

Оба слагаемых стремятся к 0, тогда и LHS$= 0$, т.е. $P(X = Y) = 1$■

15.5 Закон больших чисел (в слабой форме)

Теорема: Если X_n независимы и одинаково распределены, $\mathbb{E}(X_n) = \mu < \infty$, то

$$\operatorname{plim} \overline{X_n} = \mathbb{E}(X_n) = \mu$$

Доказательство: Докажем теорему частично, то есть для случая, когда $\mathbb{V}\mathrm{ar}(x_i) = \sigma^2 < \infty$.

$$\mathbb{P}\Big(\big|\overline{X_n} - \mu\big| \geq \varepsilon\Big) \underbrace{\leq}_{\text{неравенство Чебышёва}} \frac{\mathbb{V}\mathrm{ar}\big(\overline{X_n}\big)}{\varepsilon^2} = \frac{\mathbb{V}\mathrm{ar}\Big(\frac{X_1 + \ldots + X_n}{n}\Big)}{\varepsilon^2} = \frac{\mathbb{V}\mathrm{ar}(X_1 + \ldots + X_n)}{(n\varepsilon)^2} =$$

$$= \frac{\mathbb{V}\mathrm{ar}(X_1) + \ldots + \mathbb{V}\mathrm{ar}(X_n)}{(n\varepsilon)^2} = \frac{\mathbb{V}\mathrm{ar}(X_1)}{n\varepsilon^2} \xrightarrow{n \to \infty} 0$$

■

15.6 Ещё немного хороших теорем

Теорема (О непрерывном отображении): Если g – непрерывная функция, то:

- Если $X_n \xrightarrow{\text{dist}} X$, то $g(X_n) \xrightarrow{\text{dist}} g(X)$.
- Если $\operatorname{plim} X_n = X$, то $\operatorname{plim} g(X_n) = g(X)$.

Следствия: Если обе части определены, то

- $\operatorname{plim} X_n + \operatorname{plim} Y_n = \operatorname{plim}(X_n + Y_n)$
- $\operatorname{plim} X_n * \operatorname{plim} Y_n = \operatorname{plim}(X_n * Y_n)$
- $\frac{\operatorname{plim} X_n}{\operatorname{plim} Y_n} = \operatorname{plim}\Big(\frac{X_n}{Y_n}\Big)$

Теорема:

$$R_n \xrightarrow{\text{dist}} \underbrace{c}_{\text{const}} \Leftrightarrow \operatorname{plim} R_n = c$$

Лекция 16. Сходимость почти наверное. Сходимость в L^p

20 января 2025 г.

На лекции было добавлено еще два (строго говоря, один и ещё $\aleph_0$) вида сходимости. Будет показано, что $X_n \xrightarrow{\text{as}} X \Rightarrow \operatorname{plim} X_n = X$ и ... $\Rightarrow X_n \xrightarrow{\text{L}^2} X \Rightarrow X_n \xrightarrow{\text{L}^1} X \Rightarrow X_n \xrightarrow{\text{prob}} X$

Определение: (стандартное) Говорят, что X_n сходится к X почти наверное, если

$$\mathbb{P}(\lim X_n = X) = 1$$

Обозначение: $X_n \xrightarrow{\text{as}} X$

Определение: (альтернативное) Говорят, что X_n сходится к X почти наверное, если

$$\forall \varepsilon > 0 \lim_{n\to\infty} \mathbb{P}\big(|X_n - X| > \varepsilon \vee \left|X_{n+1} - X\right| > \varepsilon \vee \left|X_{n+2} - X\right| > \varepsilon \vee ...\big) = 0$$

Пример: Показываем, что из сходимости по вероятности не следует почти наверное.

Пусть есть вазы, в k-й вазе k шаров, среди них 1 черный. Мы извлекаем все шары по очереди из каждой вазы без возвращений.

$X_i = \mathbb{1}(\{i\text{-й шар черный}\})$, тогда $X_1 = 1, \mathbb{P}(X_2 = 0) = \frac{1}{2}, ..., \mathbb{P}(X_7 = 1) = \frac{1}{4}$

Для некоторого $\varepsilon > 0$ $\mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(X_n > \varepsilon) \le \mathbb{P}(X_n = 1) \longrightarrow 0$

В последовательности X_n есть ceslim, но нет просто lim (всё портят единички). Тогда существует сходимость по вероятности, но не почти наверное. Это верно и по определению-1: $\forall \omega \in \Omega \mathbb{P}(X_n = 0) \ne 0$, и по определению-2: $\lim \mathbb{P}(\bigcup ...) = 0$

Теорема: Определения эквивалентны

Доказательство:

$$\rceil A_n = \{|X_n - X| > \varepsilon\} \cup \left\{\left|X_{n+1} - X\right| > \varepsilon\right\} \cup ... = \bigcup_{k\ge n} \{|X_k - X| > \varepsilon\}$$

$$\text{Тогда } A_n \supseteq A_{n+1}, A_\infty = \bigcap_{n=1}^{\infty} A_n$$

Рассмотрим все исходы $\omega \in \Omega$ (получающиеся двоичные последовательности) и $B := \{\omega \mid X_n \nrightarrow X\}$

$$\omega \notin B \Rightarrow \forall \varepsilon > 0\ \exists T : \forall n \ge T\ |X_n(\omega) - X(\omega)| < \varepsilon \Leftrightarrow \varepsilon > 0\ \exists T : \omega \notin A_T \Leftrightarrow \forall \varepsilon > 0\ \omega \notin A_\infty \Rightarrow B \subseteq A_\infty$$

Обратно,

$$\omega \in B \Rightarrow \exists \varepsilon > 0\ \forall T : \exists n \ge T :\ |X_n(\omega) - X(\omega)| \ge \varepsilon \Leftrightarrow \exists \varepsilon > 0\ \forall T\ \omega \in A_T \Leftrightarrow \omega \in A_\infty(\varepsilon)$$

Тогда доказать в обе стороны можно следующим образом.

$$(\Rightarrow)\ P(B) = 0 \Rightarrow \forall \varepsilon > 0\ \mathbb{P}(A_\infty) = 0, \lim(\mathbb{P}(A_i)) = 0$$

$$B = \bigcup_{\varepsilon = \frac{1}{k}} A_\infty(\varepsilon)$$

Если опр-2 $\Rightarrow$ опр-1 $B = \bigcup_{\varepsilon=\frac{1}{k}} A_\infty(\varepsilon); \mathbb{P}(A_\infty) = 0 \Rightarrow \mathbb{P}(B) = 0$■

Определение: Говорят, что X_n сходится к X в смысле L^p, если

$$1)\mathbb{E}(|X_n|^p) < \infty, \mathbb{E}(|X|^p) < \infty$$

$$2) \lim_{n \to \infty} \mathbb{E}(|X_n - X|^p) = 0$$

L^p это некоторое линейное пространство случайных величин, для которых верно условие 1 определения. Чаще всего полезным оказывается L^2, сходимость в таком случае называют иногда сходимостью в среднеквадратичном эллиптическом пространстве. (mean squared, обозначение: $X_n \overset{\text{ms}}{\longrightarrow} X$). Данная нотация подразумевается как единый знак, а не декартово произведение.

Утверждение: $X_n \overset{\text{L}^1}{\longrightarrow} X \Rightarrow \operatorname{plim} X_n = X$

Доказательство:

$$\mathbb{P}(|X_n - X| > \varepsilon) \underbrace{\leq}_{\text{Марков}} \frac{\mathbb{E}(|X_n - X|)}{\varepsilon} \overset{n \to \infty}{\longrightarrow} 0 \blacksquare$$

Пример (контрпример обратной импликации):

$$X_n = \begin{cases} n, \mathbb{P} = \dfrac{1}{n} \\ 0, \mathbb{P} = \dfrac{n-1}{n} \end{cases}$$

$$\operatorname{plim} X_n = 0 : \mathbb{P}(|X_n - 0| > \varepsilon) \leq \mathbb{P}(X_n = n) = \frac{1}{n} \longrightarrow 0$$

$$\mathbb{E}\Big(|X_n - 0|^1\Big) = \mathbb{E}(|X_n|) = 1 \nrightarrow 0$$

Утверждение:

$$\forall a \geq b \geq 1 \hookrightarrow X_n \overset{\text{L}^a}{\longrightarrow} X \Rightarrow X_n \overset{\text{L}^b}{\longrightarrow} X$$

Доказательство:

$$\rceil Y_n = |X_n - X|^b, |X_n - X|^a = Y_n^{a/b}$$

Неравенство Йенсена для выпуклой степенной функции:

$$\mathbb{E}\Big(Y_n^{a/b}\Big) \geq (\mathbb{E}(Y_n))^{a/b}$$

Тогда если левая часть $\to 0$, то и правая тоже. ■

Утверждение (Усиленный закон больших чисел (strong law of large numbers)): Пусть СВ $Y_1, Y_2, \ldots$, независимы и одинаково распределены с $\mathbb{E}(Y_i) = \mu < \infty$. Тогда

$$\overline{Y_n} \overset{\text{as}}{\longrightarrow} \mathbb{E}(Y_1) = \mu$$

Сильное заявление. Проверять я его, конечно же, не буду.

Лемма (Слуцкий): Для $c = \text{const}$ и случайных величин X, Y верны леммы:

$$1) X_n \xrightarrow{\text{dist}} c \Leftrightarrow \operatorname{plim} X_n = c$$

$$2) \begin{cases} X_n \xrightarrow{\text{dist}} c \\ \operatorname{plim} |X_n - Y_n| = 0 \end{cases} \Rightarrow Y_n \xrightarrow{\text{dist}} c$$

$$3) \begin{cases} X_n \xrightarrow{\text{dist}} c \\ Y_n \xrightarrow{\text{dist}} Y \end{cases} \Rightarrow \begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{\text{dist}} \begin{bmatrix} c \\ Y \end{bmatrix}$$

$$4) \begin{cases} X_n \xrightarrow{\text{dist}} c \\ Y_n \xrightarrow{\text{dist}} Y \end{cases} \Rightarrow \begin{cases} X_n + Y_n \xrightarrow{\text{dist}} c + Y \\ X_n \cdot Y_n \xrightarrow{\text{dist}} c \cdot Y \\ \dfrac{Y_n}{X_n} \xrightarrow{\text{dist}} \dfrac{Y}{c} \text{ если } \exists \text{ обе части} \end{cases}$$

Лекция 17. Характеристические функции

27 января 2025 г.

Напоминание о комплекснозначных СВ

Определение: Математическое ожидание комплекснозначной СВ $X : \Omega \to \mathbb{C}^d$ определяется как

$$\mathbb{E}(X) = \mathbb{E}(\operatorname{Re} X) + i\mathbb{E}(\operatorname{Im} X)$$

Определение: Функция производящая моментов (moments generating function) Для одномерных СВ ($X : \Omega \to \mathbb{R}$)

$$mgt(t) = \mathbb{E}(\exp(tX))$$

Для многомерных $X : \Omega \to \mathbb{R}^d, X = \begin{bmatrix} X_1 \\ ... \\ X_d \end{bmatrix}$

$$mgt(t) = \mathbb{E}(\exp(< t, X >))$$

У этой функции есть недостаток: она иногда не определена. Например для такой СВ:

t	2^2	4^2	8^2	16^2	...
$P(X = t)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...

Решение - характеристическая функция, которая существует всегда.

Определение: Характеристическая функция СВ X это

$$\varphi_X(t) = \mathbb{E}(\exp(itX)) = \mathbb{E}(\cos(tX) + i\sin(tX)) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$$

Тривиальные свойства:
- $\varphi_X(0) = 1$
- $|\varphi_X(t)| \leq \mathbb{E}(|\exp(itX)|) \underbrace{=}_{\text{модуль комплексного числа}} 1$

> **Упражнение на самостоятельное решение.** Доказать, что $\mathbb{E}(|Y|) \geq |E(Y)|$ (комплексный модуль)

- $Y = -X \Rightarrow \varphi_Y(t) = \varphi_X(t)$ откуда следует $-X \sim X \Rightarrow \mathbb{E}(\sin(tX)) = 0 \Leftrightarrow$ $\varphi_X(t)$ вещественная

Полезные свойства:
- По характеристическим функциям можно судить о распределениях:

$$F_X(t) = F_Y(t) \Leftrightarrow \varphi_Y(t) = \varphi_X(t)$$

- $X_n \xrightarrow{\text{dist}} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi_X(t) \forall t$

Доказательство ($\Rightarrow$): $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X)) \forall$ непрерывных, ограниченных g. Так как $g(u) = \cos(tu), g(u) = \sin(tu)$ непрерывны и ограничены, то и $\begin{bmatrix} \mathbb{E}(\cos(X_n)) \\ \mathbb{E}(\sin(X_n)) \end{bmatrix} \to \begin{bmatrix} \mathbb{E}(\cos(X)) \\ \mathbb{E}(\sin(X)) \end{bmatrix}$

- X, Y независимы $\Leftrightarrow \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$

Доказать можно разложив в произведение, но спустившись от комплексных в действительные функции, затем раскрыть сумму тригонометрических функций, воспользоваться линейностью $\mathbb{E}$

- Если существует конечный k-й момент: $\mathbb{E}\left(|X|^k\right) = c < \infty$ то φ_X раскладывается в ряд Тейлора:

$$\varphi_X(t) = 1 + (it)\frac{\mathbb{E}(X)}{1!} + \frac{(it)^2}{2!}\mathbb{E}(X^2) + ... + \frac{(it)^k}{k!}\mathbb{E}\left(X^k\right) + o\left(t^k\right)$$

Таким образом,

$$\varphi_X(0) = 1, \varphi_X'(0) = i\mathbb{E}(X), \varphi_X''(0) = i^2\mathbb{E}(X), ... \Rightarrow \mathbb{E}\left(X^k\right) = \frac{\varphi_X^k(0)}{i^k}$$

> **Упражнение.**
>
> $$X \sim N(0,1).\varphi_X(t) - ?$$

Решение (Честное):

$$\varphi_X(t) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)) = \int_{-\infty}^{\infty} \cos(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x + \int_{-\infty}^{\infty} \underbrace{\sin(tx)}_{\text{нечетная}}\underbrace{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x}_{\text{четная}} =$$

$$= \int_{-\infty}^{\infty} \cos(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x$$

Трюк Фейнмана: обе части дифференцируем по t

$$\varphi_X'(t) = \int_{-\infty}^{\infty} \underbrace{\sin(tx)}_{u}\frac{1}{\sqrt{2\pi}}\underbrace{\exp\left(-\frac{x^2}{2}\right)(-x)\,\mathrm{d}x}_{v'} =$$

Интегрируем по частям (по x)

$$= \frac{1}{\sqrt{2\pi}}\sin(tx)t\exp\left(-\frac{x^2}{2}\right)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \cos(tx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x$$

Первое слагаемое обнуляется, имеем $\varphi'(t) = -t\varphi(t)$ - дифференциальное уравнение с разделяющимися переменными.

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = -t\varphi \Rightarrow \frac{\mathrm{d}\varphi}{\varphi} = -t\,\mathrm{d}t \Rightarrow \mathrm{d}\varphi = c_0 - \frac{t^2}{2}$$

Константу восстанавливаем из условия $\varphi(0) = 1$

Ответ: $\varphi_X(t) = \exp\left(-\frac{t^2}{2}\right)$.

Решение (Читерское):

$$\varphi_X(t) = \int\limits_{-\infty}^{\infty} \exp(itx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}(x^2 - 2itx + (it)^2 - (it)^2)\Big)\,\mathrm{d}x =$$

$$\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-it)^2 - \frac{t^2}{2}\right)\mathrm{d}x = \exp\left(-\frac{t^2}{2}\right)\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}(x-it)^2\Big)\,\mathrm{d}x$$

Из соображений комплексного анализа + из схожести интеграла с интегралом функции плотности нормального распределения можем принять интеграл как 1.

Теорема (ЦПТ): $X_1, ..., X_n$ независимы и одинаково распределены $\mathbb{E}(X_i) = \mu, \mathbb{V}\mathrm{ar}(X_i) = \sigma^2$ Тогда

$$R_n = \frac{X_1 + ... + X_n - \mathbb{E}(X_1 + ... + X_n)}{\sqrt{\mathbb{V}\mathrm{ar}(X_1 + ... + X_n)}}, R_n \xrightarrow{\text{dist}} N(0,1)$$

Доказательство (стандартное):

$$Y_i := \frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}\mathrm{ar}(X_i)}} \nsim N$$

$$\mathbb{E}(Y_i) = \mathbb{E}\left(\frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}\mathrm{ar}(X_i)}}\right) = 0$$

$$\mathbb{V}\mathrm{ar}(Y_i) = \mathbb{V}\mathrm{ar}\left(\frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}\mathrm{ar}(X_i)}}\right) = \frac{1}{\mathbb{V}\mathrm{ar}(X_i)} = \mathbb{V}\mathrm{ar}(X_i - \mathbb{E}(X_i)) = 1$$

$$\mathbb{E}(Y_i^2) = 1 = \mathbb{V}\mathrm{ar}(Y_i) + (\mathbb{E}(Y_i))^2$$

$$\varphi_{Y_i}(t) = 1 + (it)\frac{\mathbb{E}(Y_i)}{1!} + \frac{(it)^2}{2!}\mathbb{E}\big((Y_i)^2\big) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2)$$

Используя наши новые СВ, перепишем

$$R_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i = \frac{Y_1}{\sqrt{n}} + ... + \frac{Y_n}{\sqrt{n}}$$

Найдём $\varphi_{R_n}(t)$. В силу независимости слагаемых выше, эта функция есть произведение характеристических функций этих слагаемых

$$\varphi_{R_n}(t) = \prod \varphi_{Y_i/\sqrt{n}}(t) = \big(\varphi_{Y_1/\sqrt{n}}(t)\big)^n$$

$$\varphi_{Y_1/\sqrt{n}}(t) = \mathbb{E}\left(\exp\left(it\frac{Y_1}{\sqrt{n}}\right)\right) = \mathbb{E}\left(\exp\left(i\frac{t}{\sqrt{n}}Y_1\right)\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$$

$$\varphi_{R_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{\text{dist}} \exp\left(-\frac{t^2}{2}\right) = \varphi_{N(0,1)}(t)\blacksquare$$

Упражнение.

$$X \sim \mathrm{Bin}(n, p);\ \ \varphi_X(t) =?$$

Решение: $X = Y_1 + \ldots + Y_n$ Y_k независимы, равны 1 с вероятностью p и 0 иначе.

$$\varphi_X(t) = \varphi_{Y_1+\ldots+Y_n}(t) = \left(\varphi_{Y_k}(t)\right)^n$$

$\exp(itY_k)$	1	$\exp(it)$
Y_k	0	1
$\mathbb{P}$	$1-p$	p

$$\varphi_{Y_k}(t) = \mathbb{E}(\exp(itY_k)) = (1-p) \cdot 1 + p \cdot \exp(it) = 1 - p + p\cos(t) + ip\sin(t)$$

Ответ:

$$(1 - p + p\cos(t) + ip\sin(t))^n$$

.

Лекция 18. Темки, или приложения тервера в финансах

3 февраля 2025 г.

Пример: $S_0 = 100$ - благосостояние. Каждый день (год/месяц) всё благосостояние я ставлю на рисковый проект. Доходность после n раз: $S_n = S_0 \cdot M_1 \cdot M_2 ... M_n$ (например, 1.3, 1.5, .., 0.8)

Предположим, что M_j независимы и одинаково распределены, проиграть каждый раз вероятность $\frac{1}{2}$

v	x3	0
$\mathbb{P}(M_j = v)$	1/2	1/2

$$\lim_{n \to \infty} \mathbb{E}(S_n) -?$$

$$\operatorname{plim} S_n -?$$

Решение: M_j независимы, разлагаем в произведение:

$$\mathbb{E}(S_n) = \mathbb{E}(S_0)\mathbb{E}(M_1)...\mathbb{E}(M_n) = 100(\mathbb{E}(M_1))^n = 100\left(\frac{3}{2}\right)^n$$

То есть $\lim = +\infty, \operatorname{plim} S_n = 0$ (рано или поздно в цепочке проскочит 0, и далее будет 0 всегда)

Мы понимаем, что огромный сулящий успех средний ожидаемый заработок никакой роли не играет. А как тогда выгоднее ставить, чтобы разбогатеть?

Пусть есть S_n - богатство в момент времени n. Долю u (переменную) я ставлю (т.е. uS_n денег), остальное сберегаем без риска и без процентов

При тех же условиях из примера, каков закон распределения $(S_{n+1} \mid S_n)$

Значение	$(1-u)S_n + 3uS_n$	$(1-u)S_n$
$\mathbb{P}$	1/2	1/2

Неформально, наша цель - заработать в долгосроке побольше денег. Формально:

$$\max \operatorname{med}(S_n) - S_n$$

То есть максимизируем медиану по стратегии. Продемонстрируем, как с $\mathbb{P} = 1$ обойти стратегии всех соперников (выбрать лучшую траекторию заработка от времени). Эта стратегия используется в фин. рынках, однако бутылочное горлышко её применения в реальном мире - знание распределения событий M_j.

Если применить монотонное преобразование к этому распределению, самая лучшая траектория всё еще будет выше всех. Докажем для преобразования. Возьмем логарифм как наиболее удобное преобразование. Переобозначим M_j как СВ, зависящую от u (первоначально M_j зависела как бы от $u = 1$)

v	3u + (1-u)	1-u
$\mathbb{P}\left(M_j(u) = v\right)$	1/2	1/2

$S_n = S_0 M_1(u) ... M_n(u)$

Мы гарантируем, что $\mathbb{P}(S_n(\text{optim}) \geq S_n(\text{alternative})) \longrightarrow 1$

$$\frac{\ln \frac{S_n}{S_0}}{n} = \frac{\ln(M_1) + .. + \ln(M_n)}{n} \overset{n \to \infty}{\longrightarrow} \mathbb{E}(\ln(M_1(u)))$$

То есть максимизировать по стратегии есть максимизировать $\forall$ alternative $\max_{\text{strat}} \lim\limits_{n\to\infty} \mathbb{P}(S_n(\text{strat}) \geq S_n(\text{alter}))$. И по ЗБЧ тогда эта задача оказывается простой

$$\max_u \mathbb{E}(\ln(M_1(u)))$$

$$g(u) = \mathbb{E}(\ln(M_1(u))) = \frac{1}{2} \ln(1-u) + \frac{1}{2} \ln(3u + 1 - u)$$

$$g'(u) = \frac{1}{2} - \frac{1}{1-u} + \frac{1}{2} \cdot 2 \cdot \frac{1}{1+2u} = 0$$

$$2 - 2u = 1 + 2u$$

$$u = 0.25$$

Максимизация этого логарифма тесно связана с энтропией. Погуглите Критерий Келли.

> **Игрушечная постановка.** Кубик подбрасывается 100 раз. Наблюдаем результат очередного броска и хотим уйти из игры на последней 6 (выигрываем если вышли и шестерка реально оказалась последней). Какая оптимальная стратегия?

Решение: Дождемся $\frac{100}{6} = 16$ шестерок и сразу выйдем - стратегия неплохая, но неоптимальная.

Как насчет игнорировать k бросков, а выйдем по первой шестерке? $\exists l = 100 - k$ $\mathbb{P}(\text{будет 1 шестерка в l бросках}) = l\left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{l-1}$. l_{optim} ищем максимизацией $\ln$:

$$\ln(l) + \ln\left(\frac{1}{6}\right) + (l-1)\ln\left(\frac{5}{6}\right) \to \max$$

$$l \sim 5.48...$$

> **Статья 2000 г. Summ odds to 1 and stop .** Независимые индикаторы: $I_1, ..., I_n$
>
> $\mathbb{P}\left(I_j = 1\right) = P_j$
>
> Видим их по очереди, могу 1 раз сказать «стоп». Надо максимизировать $\mathbb{P}(\text{остановиться на последней единице})$

Напомним, что такое шансы. Для $\mathbb{P} = \frac{1}{2}$ шансы $1 : 1$. Для $\mathbb{P} = \frac{1}{3}$ шансы $1 : 2$. Т.е. $\mathrm{odds}_j = \frac{p_j}{1-p_j} = r_j$

Решение: Стратегия: проигнорируем k индикаторов и выберем первую единицу после.

$S_k \coloneqq I_n + ... + I_{k+1}$ - сколько единиц в конце. Задача приобретает вид

$$\max_k \mathbb{P}(S_k = 1)$$

Используем производящие вероятностей функции. Пусть $g_k(t)$ - производящая для I_k, $G_k(t)$ – для S_k, т.е. $g_k(t) = \mathbb{E}\left(t_k^I\right) = p_k t^1 + (1 - p_k)t^0 = 1 - p_k + p_k t$,

$$G_k(t) = g_n \cdot g_{n-1} \cdot ... \cdot g_{k+1} = \prod_{j=k+1}^{n} \left(1 - p_j + p_j t\right)$$

$$\text{Также, } G_k(t) = \mathbb{P}(S_k = 0) \cdot 1 + \mathbb{P}(S_k = 1) \cdot t + \mathbb{P}(S_k = 2) \cdot t^2 + ...$$

$$G_k'(0) = \mathbb{P}(S_k = 1)$$

$$\ln(G_k(t))' = \sum \frac{1}{1 - p_j + p_j t} p_j \Big|_{t=0} = \sum \frac{p_j}{1 - p_j + p_j t} = r_n + ... + r_{k+1}$$

$$\ln(G_k(t))' = \frac{G_k'(0)}{G_k(0)} = r_n + ... + r_{k+1}$$

$$G_k(0) = \mathbb{P}(S_k = 0) = (1 - p_n) \cdot ... \cdot (1 - p_{k+1})$$

$$\rceil V_k = (r_n + ... + r_{k+1})(1 - p_n) \cdot ... \cdot (1 - p_{k+1})$$

Когда верно $V_k < V_{k+1}$?

$$(r_n + ... + r_{k+1})(1 - p_n) \cdot ... \cdot (1 - p_{k+1}) < (r_n + ... + r_{k+1} + r_{k+2})(1 - p_n) \cdot ... \cdot (1 - p_{k+1} \cdot (1 - p_{k+2}))$$

$$(r_n + ... + r_{k+1})(1 - p_{k+1}) + ... + p_{k+1} < (r_n + ... + r_{k+1} + r_{k+2})$$

$$p_{k+1} < p_{k+1}(r_n + ... + r_{k+2})$$

Что верно, когда $(r_n + ... + r_{k+2}) \geq 1$ Выведена стратегия выжидания. Сложим с конца шансы до достижения единицы. Полученное k будет оптимальным.

Лекция 19. Моделирование информации. Сигма-алгебра.

10 февраля 2025 г.

Количество информации $H(X)$ мы умеем моделировать - есть энтропия.

Пример: $X_1, X_2 \sim \mathrm{Bern}(0.5)$ независимы и одинаково распределены

$H(X_1) = H(X_2)$

Создадим объект «информация» с точки зрения тервера.

Обозначение: $\sigma(X_1)$ - информация у того, кто знает X_1

Это НЕ стандартное отклонение! Стандартное отклонение это σ_{X_1}

Для σ нет отношения порядка, это какое-то множество. $\sigma_{X_1, X_2} \supseteq \sigma_{X_1}$

«Знать СВ X» - иметь возможность считать функции от X и отвечать на вопросы например $\{X_1 \leq 7\}$

Пусть $X_1 = 1$ если в первом броске выпала решка. $Y_1 = 1$ если в первом броске выпал орёл. Справедливо $\sigma_{X_1} = \sigma_{Y_1}$ (зная про одну СВ, мы понимаем всё и про другую здесь). То есть $\sigma(X)$ отражает еще и умение обрабатывать информацию.

Имея эти предпосылки (аксиомы), дадим определение.

Определение: Сигма-алгебра $\sigma(X)$, порожденная СВ X - это наименьшее множество событий, обладающее свойствами:

1) События вида $\{X \leq t\}$ лежат в нем, т.е. $\forall t \in \mathbb{R} \hookrightarrow \{X \leq t\} \in \sigma(X)$

2) Если $A_1, A_2, ... \in \sigma(X)$, то $\forall B$, которые можно сконструировать из них с помощью счетного количества операций $\{\cup, \cap, \div\}, B \in \sigma(X)$

Интуиция: $\sigma(X)$ есть список событий, состояние которых гарантированно известно наблюдателю.

Пример:

	$Y = 1$	$Y = 3$
$X = 1$	$0.1\ (a)$	$0.2\ (b)$
$X = 2$	$0.3\ (c)$	$0.4\ (d)$

То есть $\Omega = \{a, b, c, d\}$

$$\{X \leq 1.5\} = \{a, b\}; \{X \leq -7\} = \varnothing, \{X \leq 2025\} = \{a, b, c, d\}$$

$$\{a, b\}^C = \{c, d\}$$

Итого $\sigma(X) = \{\{a, b\}, \varnothing, \{a, b, c, d\}, \{c, d\}\}$. Заметим, что, скажем, a, b не отпочкуются друг от друга, т.к. для $X < 1$ не попадаем в оба из них.

Условие 2 из определения можно упростить, сформулировав в 3-х свойствах:

$$1) A \in \sigma(X) \Rightarrow A^C \in \sigma(X)$$

$$2) A_1, A_2, ... \in \sigma(X) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \sigma(X)$$

$$3) A_1, A_2, ... \in \sigma(X) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \sigma(X)$$

Условие 2 ЛИБО условие 3 в силу законов Де Моргана избыточно (традиционно 3). Действительно, $A_1 \cap A_2 \cap ... = \left(A_1^C \cup A_2^C \cup ...\right)^C$

Утверждение: $\sigma(X) \subseteq \sigma(X, Y)$

Доказательство: Неформально: если мы знаем про оба события, то можно больше раздробить на подмножества Ω■

Утверждение: $\sigma(X, Y, X+Y) = \sigma(X, Y)$

Доказательство: Надо сконструировать с помощью X, Y множество $\{X + Y \leq t\}$. Очевидный вариант $\bigcup_{a \in \mathbb{R}} (\{X \leq a\} \cap \{Y \leq t - a\})$ неверен - мы можем использовать не более чем счетное количество операций, а тут континуум. Но вспомним, что к любому иррациональному числу мы можем подойти (приблизиться) за счетное количество шагов. Взяв, скажем, $\sqrt{3}$, имеем $\{X + Y \leq t\} = \bigcap_{a, b \in \mathbb{Q}, a \geq \sqrt{3}, t-b \geq t-\sqrt{3}} (\{X \leq a\} \cap \{X \leq t - b\})$■

Определение: $\mathcal{F}$ - σ-алгебра событий на множестве исходов Ω, если

1) $\varnothing \in \mathcal{F}$

2) $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$

3) $A_i \in \mathcal{F} \Rightarrow \bigcup A_i \in \mathcal{F}, 1 \leq i \leq \infty$

Почему важно условие 3? Из первого примера:

$$\sigma(X) = \{\varnothing, \Omega, \{a, b\}, \{c, d\}\}$$

$$\sigma(Y) = \{\varnothing, \Omega, \{a, c\}, \{b, d\}\}$$

$$\sigma(X) \cup \sigma(Y) \text{ не сигма-алгебра}$$

Теорема: Пересечение <u>любого</u> числа σ-алгебр есть σ-алгебра

Доказательство:

$$]\mathcal{F} = \bigcap_{\alpha} \mathcal{F}_\alpha$$

α вообще произвольный, может быть мы даже по континууму итерируемся.

$\forall \mathcal{F}_\alpha$ - σ-алгебра $\Rightarrow \varnothing \in \mathcal{F}_\alpha \Rightarrow \varnothing \in \mathcal{F}_\alpha$

$$a \in \mathcal{F}_\alpha \Rightarrow \forall \alpha \ A \in \mathcal{F}_\alpha \Rightarrow \forall A^C \in \mathcal{F}_\alpha \Rightarrow A^C \in \mathcal{F}_\alpha$$

Третье очевидно. ■

Тогда первоначальное определение можно записать так:

$$\sigma(X) = \bigcap_{\alpha} \mathcal{F}_{\alpha}$$

Где $\{F_\alpha\}_\alpha$ - множество σ-алгебр, содержащих все события $\{X \le t\} \in \mathcal{F}_\alpha \forall t \in \mathbb{R}$

Определение: СВ X измерима относительно σ-алгебры $\mathcal{F}$, если $\sigma(X) \subseteq \mathcal{F}$

Иначе говоря, если σ-алгебра $\mathcal{F}$ содержит достаточно информации, чтобы посчитать X.

Пример: $X + Y$ измеримо относительно $\sigma(X, Y)$

Определение: Борелевская σ-алгебра B - <u>минимальная</u> σ-алгебра, такая что $(-\infty; t] \in B, \forall t \in \mathbb{R}$

> Верно ли $\{2025\} \in B$?

Решение:

$$\bigcup_{k=1}^{\infty} \left(-\infty; 2025 - \frac{1}{k}\right] = (-\infty, 2025) \in B$$

$$A^C = [2025, +\infty) \in B$$

$$(-\infty, 2025] \cap A^C = \{2025\} \in B$$

$\mathbb{N} \in B$ тоже. Действительно, достаточно сделать шаги выше для всех натуральных. Более того, Борелевская алгебра содержит все интервалы, полуинтервалы, отрезки и их счетные объединения.

Определение (альтернативное): Борелевская σ-алгебра $B(\mathbb{R})$ - <u>минимальная</u> σ-алгебра, такая что $(-\infty, t_1] \times (-\infty, t_2] \times ... \times (-\infty, t_n] \in B(\mathbb{R})\ \forall t_1, ..., t_n$

Лекция 20. Условное мат. ожидание относительно σ-алгебр

18 февраля 2025 г.

20.1 Примеры к теме прошлой лекции

Разберем пару задач на алгебры к предстоящей контрольной [22.02.2025]

Пример: Сколько разных σ-алгебр на множестве $\Omega = \{a, b, c\}$

Решение: На конечном множестве каждой σ-алгебре можно сопоставить какое-то разделение множества. Мы можем различать все исходы: $(1)\{a\}, \{b\}, \{c\}$. Можем склеить все исходы вместе: $(2)\{a, b, c\}$, а можем лишь некоторые из них: $(3)\{a, b\}, \{c\}$. Оставшиеся исходы: $(4)\{a\}, \{b, c\}; (5)\{a, c\}, \{b\}$. Получаемые алгебры:
- $\mathcal{F}_2 = \{\{a, b, c\}, \varnothing\} = \{\Omega, \varnothing\}$
- $\mathcal{F}_3 = \{\{c\}, \{a, b\}, \varnothing, \{a, b, c\}\}$
- $\mathcal{F}_1 = 2^\Omega = \{\varnothing, \Omega, \{a, b\}, \{a, c\}, \{b, c\}, \{a\}, \{b\}, \{c\}\}$

Оставшиеся две очень похожи на $\mathcal{F}_3$.

Ответ: 5 различных σ-алгебр .

Пример: X, Y - случайные величины. X принимает 3 значения, Y — тоже 3 значения. Сколько событий [=элементов σ-алгебры] может быть в $\sigma(X)$? В $\sigma(X + Y)$?

Решение: Не повлияет на ответ присваивание уникальным значениям СВ X условных числовых индексов 1, 2, 3. Получили элементарные события $\{X = 1\}, \{X = 2\}, \{X = 3\}$. То есть $|\sigma(X)| = 2^3 = 8$

В случае суммы минимальное количество уникальных значений, которое может принимать $X + Y$, достигается при идеальной зависимости X с Y, т.е. $X = Y$. Получается 3, т.е. $|\sigma(X + Y)| = 8$. Пусть Y независимо от X принимает значения $10, 20, 30$. В сумме получается 9 разных значений, отсюда $|\sigma(X + Y)| = 512$. Для произвольного же случая берем промежуточные степени двойки.

20.2 Независимость алгебр и СВ

Сигма-алгебры мы используем для описании концепта наделенности информацией на общем языке. Как описать на этом языке независимость?

Определение: σ-алгебры $\mathcal{F}_1, \mathcal{F}_2$ независимы, если независима любая пара событий $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$ (то есть $P(A_1 \cap A_2) = P(A_1)P(A_2)$))

Определение: Случайные величины X, Y независимы, если независимы минимальные σ-алгебры, относительно которых они измеримы, т.е. $\sigma(X), \sigma(Y)$

Определение: СВ X независима от σ-алгебры $\mathcal{F}$, если независимы $\mathcal{F}$ и $\sigma(X)$

Пусть есть два множества $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ и функция $f : \Omega_1 \to \Omega_2$ (Чаще всего удобнее говорить, что Ω_1 произвольно, $\Omega_2 = \mathbb{R}, \mathcal{F}_2 = \mathcal{B}$ - борелевская)

Определение: f называется измеримой функцией из Ω_1 в Ω_2, если $\forall A_2 \in \mathcal{F}_2 \hookrightarrow f^{-1}(A_2) = \{\omega_1 \in \Omega_1 \mid f(\omega_1) \in A_2\} \in \mathcal{F}_1$ (f^{-1} это прообраз)

Пример:

$$\Omega_1 = \{a, b, c\}, \mathcal{F}_1 = \{\{a, b\}, \{c\}, \{a, b, c\}, \varnothing\}; \Omega_2 = \mathbb{R}, \mathcal{F}_2 = \{\{6\}, \mathbb{R} \setminus \{6\}, \varnothing, \mathbb{R}\}$$

X, Y - такие функции, что $X(a) = 6, X(b) = 7, X(c) = 6, Y(a) = 6, Y(b) = 6, Y(c) = 11$

Какая из функций измерима?

Решение: Проверяем все прообразы

$X^{-1}(\{6\}) = \{a, c\} \notin \mathcal{F}_1 \Rightarrow X$ не измерима

$Y^{-1}(\{6\}) = \{a, b\} \in \mathcal{F}_1$

$Y^{-1}(\varnothing) = \varnothing \in \mathcal{F}_1$

$Y^{-1}(\mathbb{R}) = \{a, b, c\} \in \mathcal{F}_1$

$Y^{-1}(\mathbb{R} \setminus \{6\}) = \{c\} \in \mathcal{F}_1 \Rightarrow Y$ измерима.

20.3 Мат ожидание относительно сигма-алгебры

Центральный объект лекции - $\mathbb{E}(Y \mid \mathcal{F})$. Интуиция: наилучший прогноз Y исходя из информации, доступной в $\mathcal{F}$. Это тоже случайная величина. Формализуем требования, чтобы дать строгое определение:
- $\hat{Y} = \mathbb{E}(Y \mid \mathcal{F})$ измерима относительно $\mathcal{F}$.
- Наилучший=невозможно предсказать ошибку прогноза, т.е. $Y - \hat{Y}$ непредсказуема исходя из информации в $\mathcal{F}$
- Можно различить состояние каждого события из $\mathcal{F}$, т.е. знаем значения всех индикаторов I_A событий $A \in \mathcal{F}$. И при этом они независимы с $Y - \hat{Y}$. Это можно записать как:

$$\mathbb{E}\Big(\Big(Y - \hat{Y}\Big) I_A\Big) = \mathbb{E}\Big(Y - \hat{Y}\Big)\mathbb{E}(I_A)$$

Мы целимся в $\mathbb{E}\Big(\Big(Y - \hat{Y}\Big) I_A\Big) = 0$ в среднем, отсюда $\mathbb{E}(Y I_A) = \mathbb{E}\Big(\hat{Y} I_A\Big)$

Определение: Y - СВ, $\mathcal{F}$ - σ-алгебра; $\mathbb{E}(Y \mid \mathcal{F}) = \hat{Y}$ - СВ, удовлетворяющая свойствам:
- $\hat{Y}$ измерима относительно $\mathcal{F}$ ($\sigma\Big(\hat{Y}\Big) \subseteq \mathcal{F}$)
- $\forall A \in \mathcal{F} \hookrightarrow \mathbb{E}(Y I_A) = \mathbb{E}\Big(\hat{Y} I_A\Big)$

Пример: $\Omega = \{a, b, c\}$

ω	a	b	c
$X(\omega)$	2	2	3
$Y(\omega)$	1	4	8
P	0.1	0.2	0.7
$\hat{Y}(\omega)$	?	?	?

$\sigma(X) = ?, \hat{Y} = ?$

Решение: $\mathcal{F} = \sigma(X) = \{\varnothing, \Omega, \{a, b\}, \{c\}\}$

Сначала по здравому смыслу. Если $X = 3$ нам подходят только событие $\{c\} \in \mathcal{F}$, отсюда $\hat{Y} = 8$. $X = 2 \Rightarrow \hat{Y} = 1 \cdot \frac{1}{3} + 4 \cdot \frac{2}{3} = \frac{9}{3} = 3$ ($\frac{1}{3}, \frac{2}{3}$ есть условные вероятности в силу невозможности наступления события $\{c\}$)

А если формально по определению? Имеем переменные:

ω	a	b	c
$\hat{Y}(\omega)$	α	α	γ

- Измеримость $\hat{Y}$ относительно $\mathcal{F}$ есть.
- Требуем равенство мат. ожиданий для индикаторов. Из них получаются уравнения:
- $\mathbb{E}(Y \cdot I_\varnothing) = \mathbb{E}\left(\hat{Y} \cdot I_\varnothing\right)$ $(I_\varnothing = 0 \Rightarrow 0 = 0)$
- $\mathbb{E}(Y \cdot I_\Omega) = \mathbb{E}\left(\hat{Y} \cdot I_\Omega\right)$ $(I_\Omega = 1 \Rightarrow \mathbb{E}(Y) = \mathbb{E}\left(\hat{Y}\right))$
- $\mathbb{E}\left(Y \cdot I_{\{c\}}\right) = \mathbb{E}\left(\hat{Y} \cdot I_{\{c\}}\right) \Leftrightarrow 8 \cdot 0.7 = \gamma \cdot 0.7 \Rightarrow \gamma = 8$
- $\mathbb{E}\left(Y \cdot I_{\{a,b\}}\right) = \mathbb{E}\left(\hat{Y} \cdot I_{\{a,b\}}\right) \Leftrightarrow 0.1 \cdot 1 + 0.2 \cdot 4 = 0.1 \cdot \alpha + 0.2 \cdot \alpha \Rightarrow \alpha = 3$

Так мы сконструировали СВ $\hat{Y}$. Выразим её как функцию от X:

$$\hat{Y} = \begin{cases} 3 \text{ if } X = 2 \\ 8 \text{ if } X = 3 \end{cases} \Rightarrow \hat{Y} = 3 + 5 \cdot (X - 2)$$

Свойства:

1. Еще раз подчеркнем, если ранее для СВ Y и событий A $\mathbb{E}(Y|A), \mathbb{E}(Y)$ - константы, то $\mathbb{E}(Y \mid \mathcal{F})$ - СВ.

2. Краевые случаи.

2.1. Y измерима относительно $\mathcal{F} \Rightarrow \mathbb{E}(Y \mid \mathcal{F}) = Y$.

2.2. $\mathcal{F} = \{\varnothing, \Omega\}$, то $\mathbb{E}(Y|\mathcal{F}) = \mathbb{E}(Y)$.

Это доказывается простой проверкой условий определения.

Лекция 21. Условные мат ожидания. Продолжение.

24 февраля 2025 г.

21.1 Повторение, интуиция, свойства

На прошлой лекции было определено условное математическое ожидание как случайная величина $\hat{Y} = \mathbb{E}(Y|\mathcal{F})$ такая что $\sigma\left(\hat{Y}\right) \subseteq \mathcal{F} \& \forall A \in \mathcal{F}\ \mathbb{E}\left(\hat{Y} I_A\right) = \mathbb{E}(Y I_A)$.

Условие 2 можно переформулировать: $\mathbb{E}\left(Y - \hat{Y}|A\right) = 0 \forall A \in \mathcal{F}$

И можно еще раз переформулировать: $\mathbb{E}(Y) = \mathbb{E}\left(\hat{Y}\right)\ \forall A \in \mathcal{F} : \mathbb{C}\mathrm{ov}\left(Y - \hat{Y}, I_A\right) = 0$

Несколько вариантов интуиции для осознания:

- $\hat{Y}$ есть наилучший прогноз Y с учетом информации из $\mathcal{F}$
- $\hat{Y}$ есть «огрубление» Y до $\mathcal{F}$. То есть на различных исходах для Y (на Ω), различить которые не позволяет $\mathcal{F}$, функция $\hat{Y}$ выровнена и неточно приближает Y.
- $\hat{Y}$ есть проекция Y на множество $\mathcal{F}$-измеримых СВ. Можно заметить, что тогда вторая переформулировка задает условие ортогональности индикаторов и вектора $Y - \hat{Y}$), который является расстоянием от Y до $\mathbb{E}(Y|\mathcal{F})$. Легко тогда понять, что

$$\mathbb{E}(\mathbb{E}(Y|\mathcal{F})) = \mathbb{E}(Y);\ \ \mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(E(Y|\mathcal{F})) + \mathbb{V}\mathrm{ar}(Y - \mathbb{E}(Y|\mathcal{F}))$$

Это ни что иное, как теорема о трех перпендикулярах и теорема пифагора в нашей геометрии.

Снова о свойствах (проверяются по аксиомам)

Вспомним крайние случаи:

- Y измерима относительно $\mathcal{F} \Rightarrow \mathbb{E}(Y\ |\mathcal{F}) = Y$.
- $\mathcal{F} = \{\varnothing, \Omega\}$, то $\mathbb{E}(Y|\mathcal{F}) = \mathbb{E}(Y)$.

Другие свойства:

- $Y, \mathcal{F}$ независимы $\Rightarrow \mathbb{E}(Y|\mathcal{F}) = \mathbb{E}(Y)$
- W $\mathcal{F}$-измерима $\Rightarrow \mathbb{E}(W \cdot Y \mid \mathcal{F}) = W\mathbb{E}(Y \mid \mathcal{F})$
- Линейность: $\mathbb{E}(\alpha X + Y \mid \mathcal{F}) = \alpha\mathbb{E}(X|\mathcal{F}) + \mathbb{E}(Y|\mathcal{F})$
- $X \leq Y \Rightarrow \mathbb{E}(X \mid \mathcal{F}) \leq \mathbb{E}(Y \mid \mathcal{F})$

Теорема (Непротиворечивость определения): $\mathbb{E}(|Y|) < \infty \Rightarrow \exists \mathbb{E}(Y|\mathcal{F})$

Теорема (Про единственность): $\hat{Y}_1 = \mathbb{E}(Y|\mathcal{F}), \hat{Y}_2 = \mathbb{E}(Y|\mathcal{F}) \Rightarrow \mathbb{P}\left(\hat{Y}_1 = \hat{Y}_2\right) = 1$

Доказательство: От противного: $\mathbb{P}\left(\hat{Y}_1 > \hat{Y}_2\right) > 0 \Rightarrow \exists \delta : \mathbb{P}\left(\hat{Y}_1 - \hat{Y}_2 > \delta\right) > 0$

$A = \left\{\hat{Y}_1 - \hat{Y}_2 > \delta\right\} \in \mathcal{F}$.

Применяем аксиому 2: $0 < \delta\mathbb{P}(A) \leq \mathbb{E}\left(\left(\hat{Y}_1 - \hat{Y}_2\right) I_A\right) = \mathbb{E}\left(\hat{Y}_1 I_A - \hat{Y}_2 I_A\right) = 0$ противоречие ■

21.2 Условная дисперсия относительно сигма-алгебры

Обозначение: Если $\mathcal{F} = \sigma(W, X)$, то $\mathbb{E}(Y|\mathcal{F}) = \mathbb{E}(Y|W, X)$

Обозначение: $\mathbb{V}\mathrm{ar}(Y|\mathcal{F}) \coloneqq \mathbb{E}(Y^2|\mathcal{F}) - (\mathbb{E}(Y|\mathcal{F}))^2$

Обозначение: $\mathbb{C}\mathrm{ov}(Y, W \mid \mathcal{F}) = \mathbb{E}(YW \mid \mathcal{F}) - \mathbb{E}(Y|\mathcal{F})\mathbb{E}(W|\mathcal{F})$

Обозначение: A - событие, тогда $P(A|\mathcal{F}) = \mathbb{E}(I_A \mid \mathcal{F})$

Как нам считать условное мат ожидание? Рецепты подсчета на простых случаях

- $\rceil \Omega = A_1 \sqcup A_2 \sqcup ...(\leq |\mathbb{N}|);\ \mathcal{F} = \sigma(A_1, A_2, ...) \Rightarrow \mathbb{E}(Y \mid \mathcal{F}) = \mathbb{E}(Y|A_i)$ на A_i
- У пары (X, Y) есть совместная функция плотности, то $\mathbb{E}(Y|X) = \int\limits_{\infty}^{\infty} yf(y \mid X)\,\mathrm{d}x$, где $f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)}$

Свойства $\mathbb{V}\mathrm{ar}(Y \mid \mathcal{F})$:

- Если $Y, \mathcal{F}$ независимы, то $\mathbb{V}\mathrm{ar}(Y|\mathcal{F}) = \mathbb{V}\mathrm{ar}(Y)$

Доказательство: $\mathbb{V}\mathrm{ar}(Y|\mathcal{F}) = \mathbb{E}(Y^2|\mathcal{F}) - (\mathbb{E}(Y|\mathcal{F}))^2 = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{V}\mathrm{ar}(Y)$■

- Y $\mathcal{F}$-измерима, тогда

$$\mathbb{V}\mathrm{ar}(X + Y \mid \mathcal{F}) = \mathbb{V}\mathrm{ar}(X \mid \mathcal{F}); \mathbb{V}\mathrm{ar}(YX \mid \mathcal{F}) = Y^2\,\mathbb{V}\mathrm{ar}(X \mid \mathcal{F})$$

Доказательство (первое свойство):

$$\mathbb{V}\mathrm{ar}(X + Y|\mathcal{F}) = \mathbb{E}((X + Y)^2|\mathcal{F}) - (\mathbb{E}(X + Y|\mathcal{F}))^2 =$$

$$= \mathbb{E}(X^2 + 2XY + Y^2|\mathcal{F}) - (Y + \mathbb{E}(X|\mathcal{F}))^2 = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{V}\mathrm{ar}(Y)■$$

Вернемся к «теореме пифагора»: $\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(E(Y|\mathcal{F})) + \mathbb{V}\mathrm{ar}(Y - \mathbb{E}(Y|\mathcal{F}))$. Второе слагаемое можно переписать в виде $\mathbb{E}(\mathbb{V}\mathrm{ar}(Y|\mathcal{F}))$. Доказать можно тупым раскрытием скобок.

Это интересное свойство имеет приложение для задач со случайным количеством слагаемых.

Пример: Пусть $N, Y_1, Y_2, ...$ независимы и $S = Y_1 + Y_2 + ... + Y_N$.

$\mathbb{V}\mathrm{ar}(S) -?$

Решение: $\mathbb{V}\mathrm{ar}(S) = \mathbb{V}\mathrm{ar}(E(S|N)) + \mathbb{E}(\mathbb{V}\mathrm{ar}(S|N)) = (\mathbb{V}\mathrm{ar}(\mathbb{E}(Y_1|N) + ... + \mathbb{E}(Y_N|N))) + (\mathbb{E}(N\,\mathbb{V}\mathrm{ar}(Y_1))) = \mathbb{V}\mathrm{ar}(N\mathbb{E}(Y_1)) + \mathbb{E}(N\,\mathbb{V}\mathrm{ar}(Y_1)) = (\mathbb{E}(Y_1))^2\,\mathbb{V}\mathrm{ar}(N) + \mathbb{E}(N)\,\mathbb{V}\mathrm{ar}(Y_1)$

21.3 Парадокс двух конвертов

Есть два неотличимых конверта, в одном X, в другом $2X$. Выбираем один, смотрим сумму. У нас есть возможность безвозвратно сменить конверт. Стоит ли нам это делать?

Пусть мы открыли, там 300к. С вероятностями $\frac{1}{2}$ в другом может лежать 150к или 600к, то есть мат ожидание 375к. Вопрос - почему же мы сразу не выбрали другой конверт, если там ожидаемо больше деняк?

Контраргумент-1: неверно посчитаны эти вероятности для конечного количества вариантов. Если у организаторов этой лотереи есть минимальная и максимальная возможная сумма в конверте, то если в конверте найдена минимальная сумма, то вероятности будут не по 0.5, а 0 и 1. Пусть теперь организаторы будут подсовывать нам конверты так, чтобы вероятности были по 0.5. Такой пример: $\frac{\mathbb{P}(\text{подложили конверты 100-200})\mathbb{P}(\text{подсунули меньший})}{\mathbb{P}(\text{подложили 50-100})\mathbb{P}(\text{меньший}) + \mathbb{P}(\text{подложили 100-200})\mathbb{P}(\text{больший})} = \frac{1}{2}$ подбирается.

Контраргумент-2: противоречие получается в том, что $\mathbb{E}(Y|X) > X$. Но в нашем определении важно было вспомнить про $\mathbb{E}(Y) < \infty$. Тогда чтобы было возможно подтасовать так, как написано абзацем выше, мы вынуждены нарушить это условие, тогда такого объекта в целом не будет существовать.

Лекция 22. Мартингалы. Точки остановки.

24 февраля 2025 г.

Определение: Фильтрация - неубывающая последовательность сигма-алгебр: $\forall s \leq t \mathcal{F}_s \subseteq \mathcal{F}_t$

Интуиция: рациональный агент, который ничего не забывает.

Определение: Последовательность СВ $(M_t)_{t=0}^{\infty}$ - мартингал (martingale) относительно фильтрации $(\mathcal{F}_t)$, если $\mathbb{E}(M_{t+1} \mid \mathcal{F}_t) = M_t \forall t$

Интуиция: «справедливая игра», благосостояние

Определение (Лайт): Последовательность СВ (M_t) - мартингал, если M_t есть мартингал относительно $\mathcal{F}_k = \sigma(M_1, M_2, ..., M_t)$

Определение (Неформально): τ - момент остановки относительно фильтрации $(\mathcal{F}_t)$ - момент времени, когда информации в $(\mathcal{F}_t)$ достаточно, чтобы сказать «стоп».

Определение (Формально): $\forall t \hookrightarrow \{\tau \leq t\} \in \mathcal{F}_t$

τ это СВ, принимает изначения $\{0, 1, 2, ...\} \cup \{+\infty\}$ $(+\infty = \text{никогда})$

Пример: (X_t) независимы и одинаково распределены, равновероятно принимают значения ± 1; $\mathcal{F} = \sigma(X_1, ..., X_t), \tau = \min\{t |\ X_t = 1, X_{t+1} = 1, X_{t+2} = 1\}$

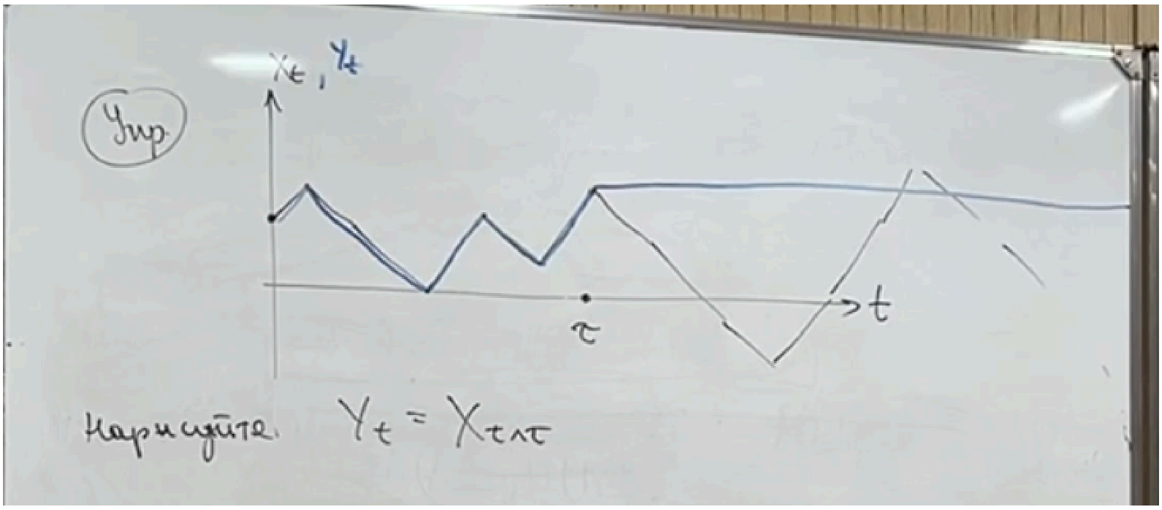
(X_t) - мартингал относительно $(\mathcal{F}_t)$? $S_t = X_1 + ... + X_t, (S_t)$ это мартингал относительно $(\mathcal{F}_k)$? τ это момент остановки относительно $\mathcal{F}_t$?

Решение: По определению в силу независимости $\mathbb{E}(X_{t+1} \mid \mathcal{F}_t) = \mathbb{E}(X_{t+1}) = 0 \neq X_t \Rightarrow$ не мартингал

$\mathbb{E}(S_{t+1} \mid \mathcal{F}_t) = \mathbb{E}(S_t + X_{t+1} \mid \mathcal{F}_t) = \mathbb{E}(S_t \mid \mathcal{F}_t) = S_t \Rightarrow$ мартингал

$\{\tau \leq 3\} \in \mathcal{F}_3$? $\{\tau \leq 3\} = \{X_1 = 1, X_2 = 1, X_3 = 1\} \in \mathcal{F}_3$? это верно. $\{\tau \leq 4\} = \{\tau = 3\} \cup \{\tau = 4\}$. Первое событие лежит в $\mathcal{F}_3$, второе расписывается так же и лежит в $\mathcal{F}_4$, короче говоря, τ это момент остановки.

Тривиальные обозначения для минимума и максимума: $\min(a, b) =: a \vee b; \max(a, b) =: a \wedge b$. Можно проверить и убедиться, что эти логические нотации вполне сочетаются со смыслом булевой логики.

> Упражнение на изображении:
>
> 
>

Решение: Синяя черта действительно ответ по логике: сначала мы следуем за значением X_t, а затем после τ благосостояние остается на фиксированном уровне.

Определение: Если τ - момент остановки относительно фильтрации $(\mathcal{F}_t)$, то σсигма-алгеброй событий, известных на момент времени τ - $\mathcal{F}_\tau$

$A \in \mathcal{F}_\tau$, если $A \cap \{\tau \le t\} \in F_t \forall t$

Интуиция: знаю всё, что было до момента τ включительно.

> (X_t) независимы и равновероятно равны ± 1. $\tau = \min\limits_{t \ge 3}\{t \mid X_t = X_{t-1} = X_{t-2} = 1\}$
>
> Пример $A \in \mathcal{F}_\tau$ не лежащего ни в одной $\mathcal{F}_t$?
>
> Пример $B \in \mathcal{F}_{100}$ не лежащего в $\mathcal{F}_\tau$?

Решение: $A = \{x_1 + x_2 + ... + x_\tau = 5\}$

$B = \{X_1 + ... + X_{100} = 3\} \in \mathcal{F}_{100}; \notin \mathcal{F}_\tau$

Так, $\mathcal{F}_\tau$ не является элементом цепочки включений $\mathcal{F}_i$

Теорема (Дуб): (обозначения как в определениях). $Y_t = M_{t \wedge \tau}$ и выполнено хотя бы одно условие: А: τ ограничен: $\exists M \in \mathbb{R}$ т.ч. $\tau \le M$

B: $\mathbb{P}(\tau = +\infty) = 0$ и Y_t ограничены: $\exists M \in \mathbb{R} \hookrightarrow |Y_t| \le M$

C: $\mathbb{E}(\tau) < +\infty$ и $\mathbb{E}(|\delta Y_{t+1}| \cdot |\mathcal{F}_t|)$ ограничено

То $\mathbb{E}(M_\tau) = \mathbb{E}(M_0)$

Это формализация простого соображения: если не ждать слишком долго и не ставить много денег, то наше мат ожидание благосостояния на конец игры будет равно стартовому.

Пример: Макака рандомно нажимает кнопки на клавиатуре (33 буквы всего). τ - момент времени, когда она впервые напишет АБРАКАДАБРА. $\mathbb{E}(\tau)$ $-?$

Решение: Очевидно сразу, что $\tau \geq 11 \Rightarrow \mathbb{E}(\tau) \geq 11$.

Построим «казино» под условия теоремы Дуба. Каждый момент t будем пускать одного посетителя с 1 рублем. Посетитель последовательно ставит все деньги на А, на Б, Р, А,... Если буква угадана, возвращается x33 - эта балансировка воплощает условие С.

Благосостояние i-го игрока это $M_{i,n}$. Тогда игрок 1: $M_{1,0} = 1, M_{1,1} = 33$ или $0, M_{1,2} = 33^2$ или 0, ...

$$\mathbb{E}\big(M_{1,t+1} \mid \mathcal{F}_t\big) = M_{1,t} = \frac{1}{33} \cdot 33 \cdot M_{1,t} + \frac{32}{33} \cdot 0s$$

Определим S_t как благосостояние всех игроков внутри казино. $S_t = M_{1,t} + M_{2,t} + ... + M_{t,t}$. (S_t) это не мартингал, а вот $Y_t = S_t - t$ - да.

$$\mathbb{E}(Y_\tau) = \mathbb{E}(Y_0) = 0 \Leftrightarrow \mathbb{E}(S_\tau - \tau) = 0 \Leftrightarrow \mathbb{E}(\tau) = \mathbb{E}(S_\tau)$$

Тогда для примера, где мартышка выстукивает ПРИВЕТАБРАКАДАБРА, первые 6 игроков, входящих в казино, сразу же обнулятся. Игрок 7 получит 33^{11}, так как он попадет в нужное время и будет угадывать каждый раз. Повезет также игроку 14, начавшему со второй «АБРА» - он возьмет 33^4 и, наконец, последнему, который уйдет с 33. То есть $\mathbb{E}(S_\tau) = 33^{11} + 33^4 + 33 = \mathbb{E}(\tau)$.