Вероятностные модели, условные вероятности и независимость

Антипов Денис Сергеевич

Университет ИТМО, Санкт-Петербург, Россия

УНИВЕРСИТЕТ ИТМО

11 февраля 2021 г.

О лекторе

- PhD in computer science (Университет ИТМО и École Polytechnique)
- Занимаюсь теорией эволюционных вычислений
- Раньше вел матан, курс по теорверу — впервые
- Контакты:
  - telegram: @antipovden
  - email: antipovden@yandex.ru
- Обращаться ко мне на “ты”, по имени (напомню: Денис)
- Не бояться перебивать меня и задавать вопросы

О курсе

- За основу взят курс MIT
- Во многом схож с курсом, который преподавала раньше Ирина Александровна Суслина
- Немного повторяет то, что вы проходили на дискретке
  - Осторожно! Можно случайно не заметить, когда началось что-то совсем для вас новое

Что вы уже знаете

- Множества, операции с ними, законы де Моргана
- Последовательности: пределы, сходимость
- Ряды, порядок суммирования, сумма геометрического ряда
- Счетность и несчетность; почему континуум несчетен
- Мера и ее свойства
- Как интегрировать

План лекции

- Определение вероятностного пространства
- Условная вероятность
- Формула полной вероятности и формула Байеса
- Независимость событий, условная независимость (если успеем)

Часть I. Вероятностное пространство

Вероятностное пространство — это тройка $(\Omega, \Sigma, \mathsf{Pr})$

1. Ω — множество элементарных исходов
2. Σ — σ-алгебра событий
3. Pr — вероятностная мера

Множество элементарных исходов Ω

Мы рассматриваем какой-то эксперимент, у которого возможны разные исходы. Множество всех возможных исходов и есть Ω.

Множество элементарных исходов Ω

Мы рассматриваем какой-то эксперимент, у которого возможны разные исходы. Множество всех возможных исходов и есть Ω.

- Эксперимент не может закончиться сразу двумя исходами одновременно (исходы — взаимоисключающие)

Множество элементарных исходов Ω

Мы рассматриваем какой-то эксперимент, у которого возможны разные исходы. Множество всех возможных исходов и есть Ω.

- Эксперимент не может закончиться сразу двумя исходами одновременно (исходы — взаимоисключающие)
- Эксперимент не может закончиться исходом не из Ω (Ω — полное)

Множество элементарных исходов Ω

Мы рассматриваем какой-то эксперимент, у которого возможны разные исходы. Множество всех возможных исходов и есть Ω.

- Эксперимент не может закончиться сразу двумя исходами одновременно (исходы — взаимоисключающие)
- Эксперимент не может закончиться исходом не из Ω (Ω — полное)
- Множество Ω не должно быть черезчур подробным (Ω — неизбыточное)

Примеры Ω

- Подбрасывание монеты
  - $\Omega = \{\text{орел, решка}\}$

Примеры Ω

- Подбрасывание монеты
  - $\Omega = \{\text{орел, решка}\}$
  - Если вы зануда: $\Omega = \{\text{орел, решка, ребро}\}$

Примеры Ω

- Подбрасывание монеты
  - $\Omega = \{$орел, решка$\}$
  - Если вы зануда: $\Omega = \{$орел, решка, ребро$\}$
- Хоккейный матч (КХЛ, NHL)
  - $\Omega = \{$В, ВО, ВБ, ПБ, ПО, П$\}$
  - Если нас интересуют только очки одной команды, то $\Omega = \{0, 1, 2\}$

Примеры Ω

- Подбрасывание монеты
  - $\Omega = \{\text{орел, решка}\}$
  - Если вы зануда: $\Omega = \{\text{орел, решка, ребро}\}$
- Хоккейный матч (КХЛ, NHL)
  - $\Omega = \{\text{В, ВО, ВБ, ПБ, ПО, П}\}$
  - Если нас интересуют только очки одной команды, то $\Omega = \{0, 1, 2\}$
- Время ожидания автобуса на остановке
  - $\Omega = [0, +\infty)$

σ-алгебра событий Σ

Событие — любое подмножество Ω ($\Leftrightarrow$ множество исходов)

σ-алгебра событий Σ

Событие — любое подмножество Ω ($\Leftrightarrow$ множество исходов)

На Ω должна быть задана σ-алгебра событий Σ, то есть множество подмножеств, такое, что

σ-алгебра событий Σ

Событие — любое подмножество Ω ($\Leftrightarrow$ множество исходов)

На Ω должна быть задана σ-алгебра событий Σ, то есть множество подмножеств, такое, что

1. $\emptyset \in \Sigma$

σ-алгебра событий Σ

Событие — любое подмножество Ω ($\Leftrightarrow$ множество исходов)

На Ω должна быть задана σ-алгебра событий Σ, то есть множество подмножеств, такое, что

1. $\emptyset \in \Sigma$
2. $A \in \Sigma \Rightarrow \bar{A} \in \Sigma$

σ-алгебра событий Σ

Событие — любое подмножество Ω ($\Leftrightarrow$ множество исходов)

На Ω должна быть задана σ-алгебра событий Σ, то есть множество подмножеств, такое, что

1. $\emptyset \in \Sigma$
2. $A \in \Sigma \Rightarrow \bar{A} \in \Sigma$
3. $\{A_1, \dots, A_n, \dots\} \in \Sigma \Rightarrow \bigcap\limits_{i \in \mathbb{N}} A_i \in \Sigma$ и $\bigcup\limits_{i \in \mathbb{N}} A_i \in \Sigma$

Примеры событий

- Домашняя команда выиграла $\{$В, ВО, ВБ$\}$
- Монета упала орлом $\{$орел$\}$
- Автобус приехал в течение 5 минут $[0, 5]$ (если считаем время в минутах)

Вероятность Pr

Вероятностная мера Pr — это функция, заданная на Σ, со следующими свойствами

Вероятность Pr

Вероятностная мера Pr — это функция, заданная на Σ, со следующими свойствами

1. Неотрицательность: $\Pr(A) \geq 0$ для любого события A

Вероятность Pr

Вероятностная мера Pr — это функция, заданная на Σ, со следующими свойствами

1. Неотрицательность: $\Pr(A) \geq 0$ для любого события A
2. Нормализация: $\Pr(\Omega) = 1$

Вероятность Pr

Вероятностная мера Pr — это функция, заданная на Σ, со следующими свойствами

1. Неотрицательность: $\Pr(A) \geq 0$ для любого события A
2. Нормализация: $\Pr(\Omega) = 1$
3. Счетная аддитивнотсь: $A_1, \ldots, A_n, \ldots$ — последовательность попарно непересекающихся событий, тогда

$$\Pr\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \Pr(A_i).$$

NB: В литературе встречаются обозначения $P, p, \mathbb{P}$, их можно использовать

Свойства Pr, следующие из аксиом

- $\Pr(A) \leq 1$
- $\Pr(\emptyset) = 0$
- $\Pr(A) + \Pr(\bar{A}) = 1$
- $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$ — Union bound

Объект теории вероятности

- Мы будем работать с вероятнсотным пространством $(\Omega, \Sigma, \mathrm{Pr})$:
  - Описывать множество элементарных исходов и определять события
  - Задавать вероятностную меру на этих событиях
  - Делать какие-то выводы о построенной модели
- Будем делать это уже на ближайшей практике ☺

Как воспринимать нашу деятельность

- Теорвер — просто раздел математики
  - На основе определенных аксиом выводим какие-то более сложные утверждения и называем их теоремами

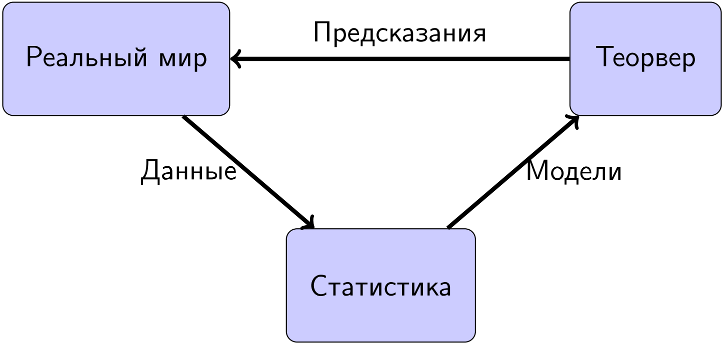
Как воспринимать нашу деятельность

- Теорвер — просто раздел математики
  - На основе определенных аксиом выводим какие-то более сложные утверждения и называем их теоремами
  - Теорема Вероятность события A есть p.

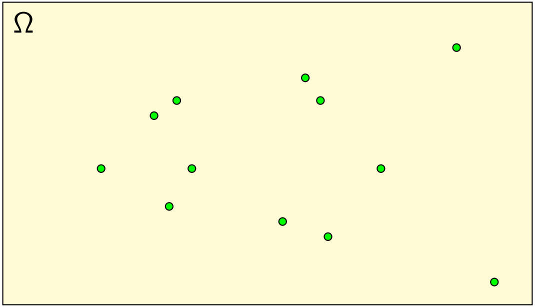
Как воспринимать нашу деятельность

- Теорвер — просто раздел математики
  - На основе определенных аксиом выводим какие-то более сложные утверждения и называем их теоремами
  - Теорема Вероятность события A есть p.
- Как интерпретировать вероятность?
  - Вероятность = частота события (Пример: если много раз кидать монетку, то примерно половина результатов будет “орел”)
  - Вероятность = наша вера в событие (Пример: с вероятностью 0.125 выиграет в этом году кубок Гагарина)

Взаимодействие с реальным миром

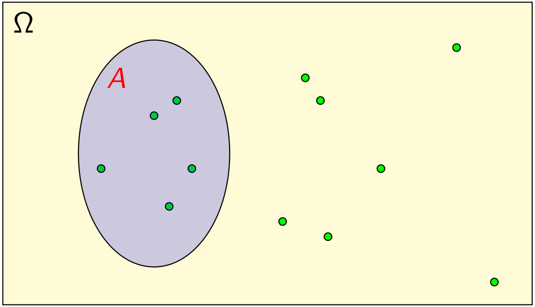


Часть II. Условная вероятность



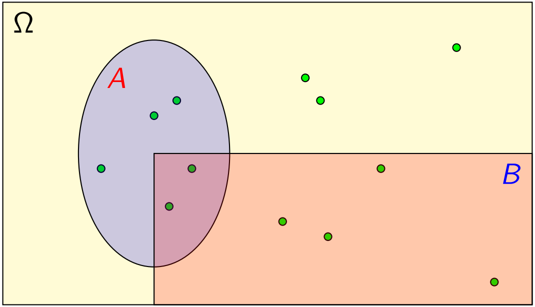
- Пусть все исходы равновероятны (вероятность каждого $\frac{1}{12}$)

Часть II. Условная вероятность



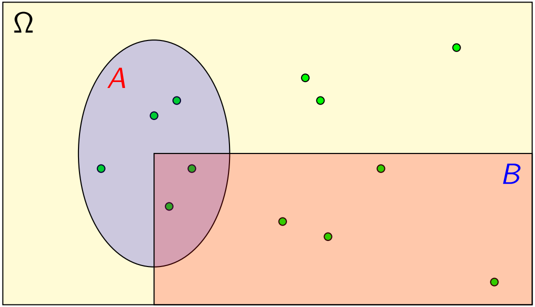
- Пусть все исходы равновероятны (вероятность каждого $\frac{1}{12}$)
- Вероятность события A есть $\Pr(A) = 5 \cdot \frac{1}{12} = \frac{5}{12}$

Часть II. Условная вероятность



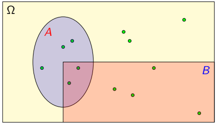
- Пусть все исходы равновероятны (вероятность каждого $\frac{1}{12}$)
- Вероятность события A есть $\Pr(A) = 5 \cdot \frac{1}{12} = \frac{5}{12}$
- Нам стало известно, что произошло событие B. Какой стала вероятность события A?

Часть II. Условная вероятность



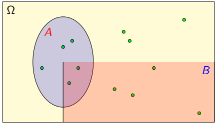
- Пусть все исходы равновероятны (вероятность каждого $\frac{1}{12}$)
- Вероятность события A есть $\Pr(A) = 5 \cdot \frac{1}{12} = \frac{5}{12}$
- Нам стало известно, что произошло событие B. Какой стала вероятность события A?
- В событие B входят 6 равновероятных исходов, из которых только 2 входят в событие A. То есть теперь вероятность события A есть $2 \cdot \frac{1}{6} = \frac{1}{3}$.

Поступление новой информации



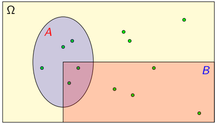
- При появлении новой информации мы хотим изменить свою модель, чтобы она соответствовала новым условиям

Поступление новой информации



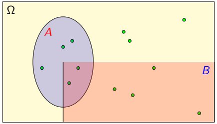
- При появлении новой информации мы хотим изменить свою модель, чтобы она соответствовала новым условиям
- Пусть мы знаем, что произошло событие B. Тогда мы хотим:
  - $A : A \cap B = \emptyset \Rightarrow \Pr(A \mid B) = 0$

Поступление новой информации



- При появлении новой информации мы хотим изменить свою модель, чтобы она соответствовала новым условиям
- Пусть мы знаем, что произошло событие B. Тогда мы хотим:
  - $A : A \cap B = \emptyset \Rightarrow \Pr(A \mid B) = 0$
  - $A \subset B \Rightarrow \Pr(A \mid B) = \frac{\Pr(A)}{\Pr(B)}$ (сохраняем нормализацию)

Поступление новой информации



- При появлении новой информации мы хотим изменить свою модель, чтобы она соответствовала новым условиям
- Пусть мы знаем, что произошло событие B. Тогда мы хотим:
  - $A : A \cap B = \emptyset \Rightarrow \Pr(A \mid B) = 0$
  - $A \subset B \Rightarrow \Pr(A \mid B) = \frac{\Pr(A)}{\Pr(B)}$ (сохраняем нормализацию)
- Любое событие можно представить как $A = (A \cap B) \cup (A \cap \bar{B})$, тогда по аддитивности вероятности

$$\Pr(A \mid B) = \Pr(A \cap B \mid B) + \Pr(A \cap \bar{B} \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} + 0.$$

Определение условной вероятности

$$\boxed{\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}}$$

$\Pr(A \mid B)$ — Вероятность события A при условии события B

Определение условной вероятности

$$\boxed{\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}}$$

$\Pr(A \mid B)$ — Вероятность события A при условии события B

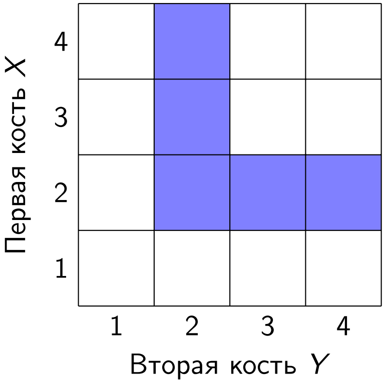
NB: $\Pr(\cdot \mid B)$ — это новая вероятностная мера, заданная на той же самой σ-алгебре, то есть для нее выполняются все аксиомы меры и следствия из них

Пример условной вероятности

Бросок двух тетраидальных костей

Пример условной вероятности

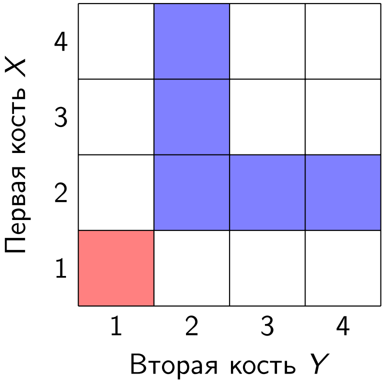
Бросок двух тетраидальных костей



Событие B: $\min\{X, Y\} = 2$

Пример условной вероятности

Бросок двух тетраидальных костей

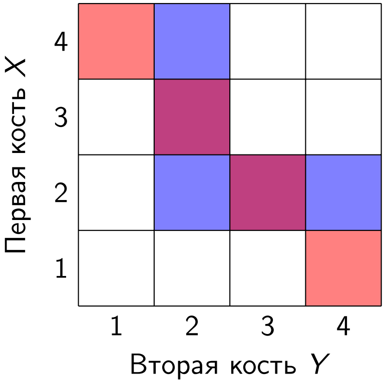


Событие B: $\min\{X, Y\} = 2$

- $\Pr(\max\{X, Y\} = 1 \mid B) = 0$

Пример условной вероятности

Бросок двух тетраидальных костей

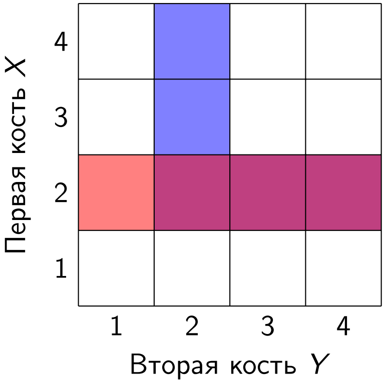


Событие B: $\min\{X, Y\} = 2$

- $\Pr(\max\{X, Y\} = 1 \mid B) = 0$
- $\Pr(X + Y = 5 \mid B) = \frac{2}{5}$

Пример условной вероятности

Бросок двух тетраидальных костей



Событие B: $\min\{X, Y\} = 2$

- $\Pr(\max\{X, Y\} = 1 \mid B) = 0$
- $\Pr(X + Y = 5 \mid B) = \frac{2}{5}$
- $\Pr(X = 2 \mid B) = \frac{3}{5}$

Модели с условной вероятностью

- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 

Модели с условной вероятностью

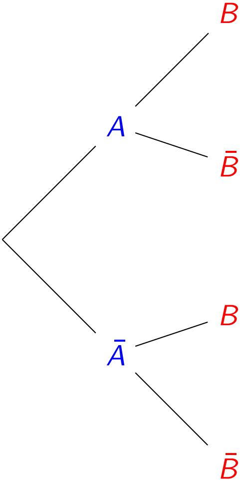
- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 
  - Урон титана 40-60 (выбирается равновероятно)
  - Здоровье дендройда 55
  - Пусть навыки атаки и защиты одинаковы
  - Удача срабатывает с вероятностью 0.125 и удваивает урон

Модели с условной вероятностью

- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 
  - Урон титана 40-60 (выбирается равновероятно)
  - Здоровье дендройда 55
  - Пусть навыки атаки и защиты одинаковы
  - Удача срабатывает с вероятностью 0.125 и удваивает урон
- Событие A: сработала удача
- Событие B: дендройд пал

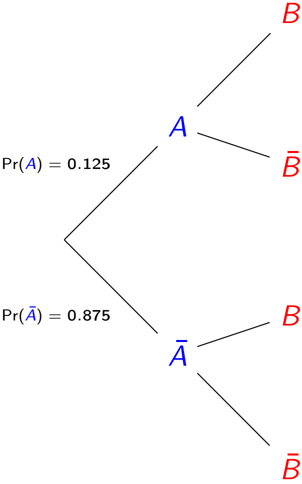
Модели с условной вероятностью

- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 
  - Урон титана 40-60 (выбирается равновероятно)
  - Здоровье дендройда 55
  - Пусть навыки атаки и защиты одинаковы
  - Удача срабатывает с вероятностью 0.125 и удваивает урон
- Событие A: сработала удача
- Событие B: дендройд пал



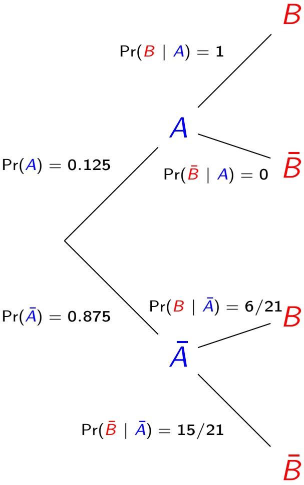
Модели с условной вероятностью

- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 
  - Урон титана 40-60 (выбирается равновероятно)
  - Здоровье дендройда 55
  - Пусть навыки атаки и защиты одинаковы
  - Удача срабатывает с вероятностью 0.125 и удваивает урон
- Событие A: сработала удача
- Событие B: дендройд пал



Модели с условной вероятностью

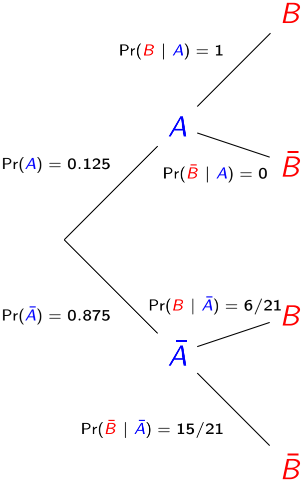
- В HoMM3 у вашего героя экспертный навык удачи 
- Ваш титан  атакует дендройда 
  - Урон титана 40-60 (выбирается равновероятно)
  - Здоровье дендройда 55
  - Пусть навыки атаки и защиты одинаковы
  - Удача срабатывает с вероятностью 0.125 и удваивает урон
- Событие A: сработала удача
- Событие B: дендройд пал



Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

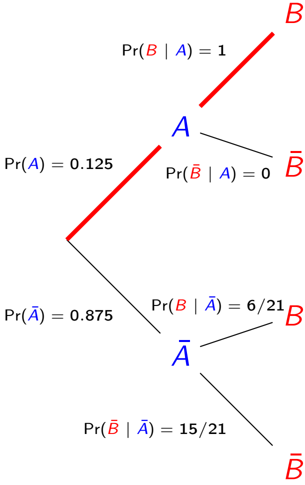
- $\Pr(A \cap B) =$



Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

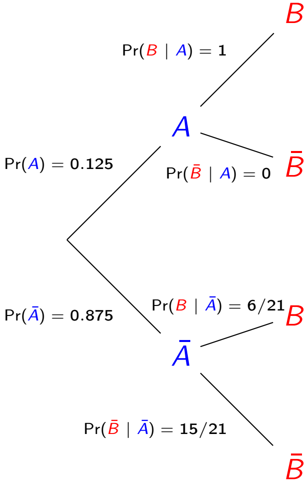
- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = 0.125$



Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = 0.125$
- $\Pr(B) =$

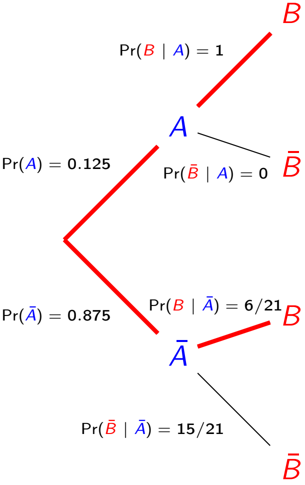


Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = 0.125$
- $\Pr(B) = \quad \Pr(B \cap A) + \Pr(B \cap \bar{A}) =$

  $\Pr(A)\Pr(B \mid A) + \Pr(\bar{A})\Pr(B \mid \bar{A}) =$
  $0.125 \cdot 1 + 0.875 \cdot \frac{6}{21} = 0.375$

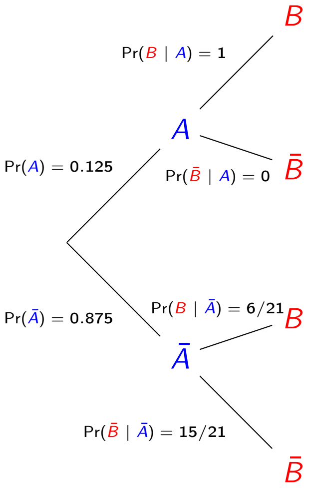


Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = 0.125$
- $\Pr(B) = \quad \Pr(B \cap A) + \Pr(B \cap \bar{A}) =$

  $\Pr(A)\Pr(B \mid A) + \Pr(\bar{A})\Pr(B \mid \bar{A}) =$
  $0.125 \cdot 1 + 0.875 \cdot \frac{6}{21} = 0.375$
- $\Pr(A \mid B) =$



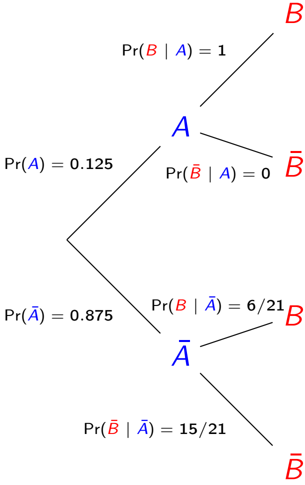
Модели с условной вероятностью

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)}$$

- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = 0.125$
- $\Pr(B) = \quad \Pr(B \cap A) + \Pr(B \cap \bar{A}) =$

  $\Pr(A)\Pr(B \mid A) + \Pr(\bar{A})\Pr(B \mid \bar{A}) =$
  $0.125 \cdot 1 + 0.875 \cdot \frac{6}{21} = 0.375$

- $\Pr(A \mid B) = \quad \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{0.125}{0.375} = \frac{1}{3}$



Правило (теорема) умножения вероятностей

- Мы уже видели, что $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$
- Можно обобщить:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1)\prod_{i=2}^{n}\Pr(A_i \mid A_1 \cap \cdots \cap A_{i-1})$$

Правило (теорема) умножения вероятностей

- Мы уже видели, что $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$
- Можно обобщить:

$$\boxed{\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1)\prod_{i=2}^{n}\Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})}$$

- Докажем по индукции

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr\left(A_n \cap \left(\bigcap_{i=1}^{n-1} A_i\right)\right)$$

Правило (теорема) умножения вероятностей

- Мы уже видели, что $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$
- Можно обобщить:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1)\prod_{i=2}^{n}\Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})$$

- Докажем по индукции

$$\begin{aligned}\Pr\left(\bigcap_{i=1}^{n} A_i\right) &= \Pr\left(A_n \cap \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} A_i\right)\Pr\left(A_n \mid \left(\bigcap_{i=1}^{n-1} A_i\right)\right)\end{aligned}$$

Правило (теорема) умножения вероятностей

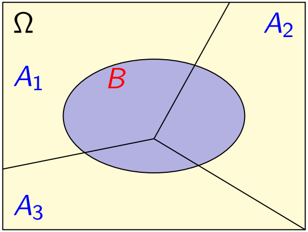
- Мы уже видели, что $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A)$
- Можно обобщить:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1)\prod_{i=2}^{n}\Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})$$

- Докажем по индукции

$$\begin{aligned}
\Pr\left(\bigcap_{i=1}^{n} A_i\right) &= \Pr\left(A_n \cap \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \\
&= \Pr\left(\bigcap_{i=1}^{n-1} A_i\right)\Pr\left(A_n \mid \left(\bigcap_{i=1}^{n-1} A_i\right)\right) \\
&= \Pr(A_1)\prod_{i=2}^{n}\Pr(A_i \mid A_1 \cap \dots \cap A_{i-1})
\end{aligned}$$

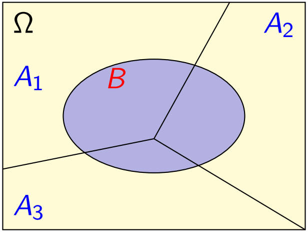
Правило (теорема) полной вероятности



Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

Правило (теорема) полной вероятности

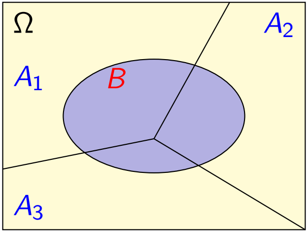


Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid A_i)$$

Правило (теорема) полной вероятности



Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

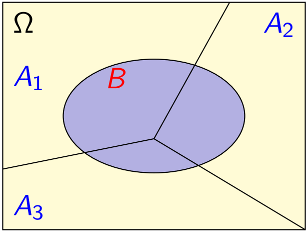
- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid A_i)$$

Доказательство:

- $B = \bigcup_i (B \cap A_i)$ — объединение непересекающихся множеств
- $\Pr(B \cap A_i) = \Pr(A_i) \Pr(B \mid A_i)$ — из правила умножения

Правило (теорема) полной вероятности



Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

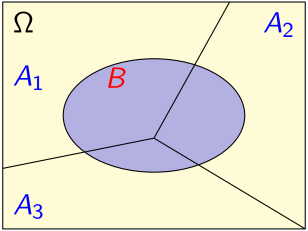
$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid A_i)$$

Доказательство:

- $B = \bigcup_i (B \cap A_i)$ — объединение непересекающихся множеств
- $\Pr(B \cap A_i) = \Pr(A_i) \Pr(B \mid A_i)$ — из правила умножения

NB: Это верно как для конечного, так и для счетного разбиения Ω

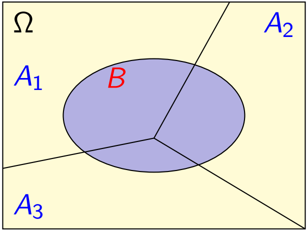
Формула Байеса



Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

Формула Байеса



Есть разбиение Ω на события $A_1, A_2, A_3, \ldots$

- Знаем $\Pr(A_i)$ для всех i
- Знаем $\Pr(B \mid A_i)$ для всех i

$$\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_j \Pr(A_j)\Pr(B \mid A_j)}$$

Часть III. Независимость

- Бросаем два раза нечестную монету:
  - $\Pr(P) = p \neq 0.5$
  - $\Pr(O) = (1 - p)$
- Получаем один из четырех результатов:

$$\{PP, PO, OP, OO\}$$

- Вероятности этих исходов:
  - $\Pr(PP) = \Pr(P*)\Pr(*P \mid P*) = p \cdot p = p^2$
  - $\Pr(PO) = \Pr(OP) = p(1 - p)$
  - $\Pr(OO) = (1 - p)^2$

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

$$\Pr(*P) = p$$

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

$$\Pr(*P) = p$$

- Какова вероятность, что второй бросок упадет решкой, если первый упал орлом?

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

$$\Pr(*P) = p$$

- Какова вероятность, что второй бросок упадет решкой, если первый упал орлом?

$$\Pr(*P \mid O*) = p,$$

то есть условие не влияет на вероятность

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

$$\Pr(*P) = p$$

- Какова вероятность, что второй бросок упадет решкой, если первый упал орлом?

$$\Pr(*P \mid O*) = p,$$

то есть условие не влияет на вероятность

- Какова вероятность, что второй бросок упадет решкой, если есть ровно одна решка?

Интуитивное понимание независимости

- Какова вероятность, что второй бросок упадет решкой?

$$\Pr(*P) = p$$

- Какова вероятность, что второй бросок упадет решкой, если первый упал орлом?

$$\Pr(*P \mid O*) = p,$$

то есть условие не влияет на вероятность

- Какова вероятность, что второй бросок упадет решкой, если есть ровно одна решка?

$$\Pr(*P \mid PO \cup OP) = \frac{\Pr(OP)}{\Pr(PO \cup OP)} = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2},$$

условие влияет на вероятность

Определение независимости

- Интуитивно: события A и B независимы, если $\Pr(A \mid B) = \Pr(A)$
  - Событие B не несет никакой информации о событии A

Определение независимости

- Интуитивно: события A и B независимы, если $\Pr(A \mid B) = \Pr(A)$
  - Событие B не несет никакой информации о событии A
  - Но это не работает, если $\Pr(B) = 0$

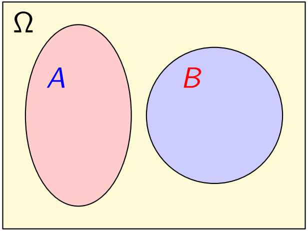
Определение независимости

- Интуитивно: события A и B независимы, если $\Pr(A \mid B) = \Pr(A)$
  - Событие B не несет никакой информации о событии A
  - Но это не работает, если $\Pr(B) = 0$
- Лучше так: события A и B независимы, если

$$\boxed{\Pr(A \cap B) = \Pr(A)\Pr(B)}$$

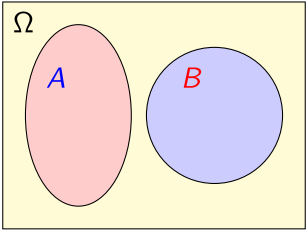
  - Симметрично относительно событий A и B
  - Из него следует и $\Pr(A \mid B) = \Pr(A)$, и $\Pr(B \mid A) = \Pr(B)$ (если условные вероятности определены)
  - Корректно и при $\Pr(A) = 0$, и при $\Pr(B) = 0$

Типичная ошибка в понимании независимости



События A и B — незавсимы?

Типичная ошибка в понимании независимости



События A и B — незавсимы?

Они максимально зависимы: если произошло A, то точно не произошло B, и наоборот

Независимость дополнений

- Если A и B независимы, то независимы и A и $\bar{B}$

Независимость дополнений

- Если A и B независимы, то независимы и A и $\bar{B}$
  - Интуиция: если “событие B произошло” не дает никакой инфы про A, то и “событие B не произошло” не должно ее давать

Независимость дополнений

- Если A и B независимы, то независимы и A и $\bar{B}$
  - Интуиция: если “событие B произошло” не дает никакой инфы про A, то и “событие B не произошло” не должно ее давать
  - Формально:

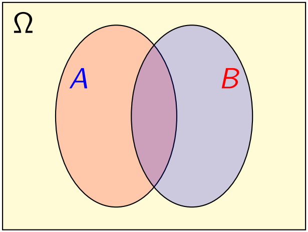
$$
\begin{aligned}
P(A) &= \Pr(A \cap B) + \Pr(A \cap \bar{B}) \\
&= \Pr(A)\Pr(B) + \Pr(A \cap \bar{B}) \\
\Rightarrow \Pr(A \cap \bar{B}) &= P(A) - \Pr(A)\Pr(B) = \Pr(A)(1 - \Pr(B)) \\
&= \Pr(A)\Pr(\bar{B})
\end{aligned}
$$

Условная независимость

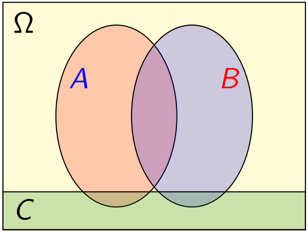
События A и B независимы при условии C, если

$$\Pr(A \cap B \mid C) = \Pr(A \mid C)\Pr(B \mid C)$$

Независимость и условная независимость не особо связаны друг с другом



Пусть A и B — независимы



$A \cap C$ и $B \cap C$ даже не пересекаются

Независимость множества событий

Пусть есть конечный набор событий $\{A_1, \ldots, A_n\}$

Независимость множества событий

Пусть есть конечный набор событий $\{A_1, \ldots, A_n\}$
Они независимы, если никакой набор событий не влияет на вероятность любого другого события.

$$\Pr(A_1 \cap \bar{A}_4) = \Pr(A_1 \cap \bar{A}_4 \mid A_2 \cap (A_3 \cup \bar{A}_5))$$

разные индексы слева и справа от "|"

Независимость множества событий

Пусть есть конечный набор событий $\{A_1, \ldots, A_n\}$
Они независимы, если никакой набор событий не влияет на вероятность любого другого события.

$$\Pr(A_1 \cap \bar{A}_4) = \Pr(A_1 \cap \bar{A}_4 \mid A_2 \cap (A_3 \cup \bar{A}_5))$$

разные индексы слева и справа от "|"

События $\{A_1, \ldots, A_n\}$ независимы (по совокупности), если для любого набора индексов $I \subset [1..n]$ верно

$$\Pr\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \Pr(A_i)$$

Пример для трех событий

Есть события $\{A_1, A_2, A_3\}$. Они независимы, если

$$\left.\begin{array}{l} \Pr(A_1 \cap A_2) = \Pr(A_1)\Pr(A_2) \\ \Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3) \\ \Pr(A_2 \cap A_3) = \Pr(A_2)\Pr(A_3) \end{array}\right\}$$

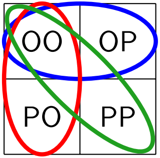
Попарная независимость (слабее, чем независимость по совокупности)

- $\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3)$

Попарная vs По совокупности

Бросаем две монеты

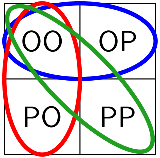
- A — первая монета орлом
- B — вторая монета орлом
- C — обе монеты одинаковы



Попарная vs По совокупности

Бросаем две монеты

- A — первая монета орлом
- B — вторая монета орлом
- C — обе монеты одинаковы



Вероятности событий:

- $\Pr(A) = \frac{1}{2}$
- $\Pr(B) = \frac{1}{2}$
- $\Pr(C) = \frac{1}{2}$

Вероятности комбинаций:

- $\Pr(A \cap B) = \frac{1}{4} = \Pr(A)\Pr(B)$
- $\Pr(A \cap C) = \frac{1}{4} = \Pr(A)\Pr(C)$
- $\Pr(B \cap C) = \frac{1}{4} = \Pr(B)\Pr(C)$
- $\Pr(A \cap B \cap C) = \frac{1}{4} \neq \Pr(A)\Pr(B)\Pr(C)$

Дискретные случайные величины

Антипов Денис Сергеевич

Университет ИТМО, Санкт-Петербург, Россия

УНИВЕРСИТЕТ ИТМО

18 февраля 2021 г.

Случайная величина

Случайная величина (с.в.) — функционал, заданный на Ω.

$$X : \Omega \to \mathbb{R}$$

Случайная величина

Случайная величина (с.в.) — функционал, заданный на Ω.

$$X : \Omega \to \mathbb{R}$$

Случайная величина

Случайная величина (с.в.) — функционал, заданный на Ω.

$$X : \Omega \to \mathbb{R}$$

- NB: этот функционал должен быть измерим по используемой мере $\Pr$

Случайная величина

Случайная величина (с.в.) — функционал, заданный на Ω.

$$X : \Omega \to \mathbb{R}$$

- NB: этот функционал должен быть измерим по используемой мере $\Pr$

Договоримся об обозначениях

- Большая буква (X, Y, Z, etc) — случайная величина (как отображение)
- Маленькая буква (x, y, z, etc) — значение случайной величины (число)

Примеры случайных величин

На конечной Ω

- Бросок кости: X — число на верхней грани
  - {⚀ $\to 1$, ⚁ $\to 2$, ⚂ $\to 3$, ⚃ $\to 4$, ⚄ $\to 5$, ⚅ $\to 6$}
- Хоккейный матч: X — очки домашней команды
  - {В $\to 2$, ВО $\to 2$, ВБ $\to 2$, ПБ $\to 1$, ПО $\to 1$, П $\to 0$}

Примеры случайных величин

На конечной Ω

- Бросок кости: X — число на верхней грани
  - {⚀ $\to 1$, ⚁ $\to 2$, ⚂ $\to 3$, ⚃ $\to 4$, ⚄ $\to 5$, ⚅ $\to 6$}
- Хоккейный матч: X — очки домашней команды
  - $\{\text{В} \to 2, \text{ВО} \to 2, \text{ВБ} \to 2, \text{ПБ} \to 1, \text{ПО} \to 1, \text{П} \to 0\}$

На счетной Ω

- Бросаем монету до первого орла: X — число бросков

Примеры случайных величин

На конечной Ω

- Бросок кости: X — число на верхней грани
  - {⚀ $\to 1$, ⚁ $\to 2$, ⚂ $\to 3$, ⚃ $\to 4$, ⚄ $\to 5$, ⚅ $\to 6$}
- Хоккейный матч: X — очки домашней команды
  - {В $\to 2$, ВО $\to 2$, ВБ $\to 2$, ПБ $\to 1$, ПО $\to 1$, П $\to 0$}

На счетной Ω

- Бросаем монету до первого орла: X — число бросков

На несчетной Ω

- Бросаем дротик в мишень для дартса: X — очки согласно правилам игры
- Крутим волчок в ЧГК: X — угол, на котором он остановится
  - Y — население города, из которого пришел вопрос
  - Z — вероятность того, что на выбранный вопрос знатоки ответят

Дискретные случайные величины

Дискретными называются те с.в., которые определены на не более, чем счетной Ω

Дискретные случайные величины

Дискретными называются те с.в., которые определены на не более, чем счетной Ω

NB: Множеством значений дискретной с.в. может быть любое подмножество $\mathbb{R}$ (не более, чем счетное, разумеется)

Дискретные случайные величины

Дискретными называются те с.в., которые определены на не более, чем счетной Ω

NB: Множеством значений дискретной с.в. может быть любое подмножество $\mathbb{R}$ (не более, чем счетное, разумеется)

NB-2: Никаких проблем с измеримостью дискретной с.в. как функции обычно нет, поэтому этот вопрос мы будем опускать

Функция вероятности

Можно рассматривать $(X = x)$ как событие, так как оно задает какое-то множество исходов из Ω. Тогда у этого события должна быть вероятность.

Функция вероятности

Можно рассматривать $(X = x)$ как событие, так как оно задает какое-то множество исходов из Ω. Тогда у этого события должна быть вероятность.

$$p_X(x) = \Pr(X = x) = \Pr(\{\omega \in \Omega \mid X(\omega) = x\})$$

p_X называется функцией вероятности с.в. X.

Функция вероятности

Можно рассматривать $(X = x)$ как событие, так как оно задает какое-то множество исходов из Ω. Тогда у этого события должна быть вероятность.

$$p_X(x) = \Pr(X = x) = \Pr(\{\omega \in \Omega \mid X(\omega) = x\})$$

p_X называется функцией вероятности с.в. X.
Ее свойства:

- $p_X(x) \geq 0$
- $\sum_x p_X(x) = 1$

Примеры случайных величин

Если нам дано какое-то распределение $\mathcal{D}$ и с.в. X ему следует, мы пишем $X \sim \mathcal{D}$

Примеры распределений (законов):

- Бернулли
- Равномерное
- Биномиальное
- Геометрическое
- Гипер-геометрическое
- Степенное

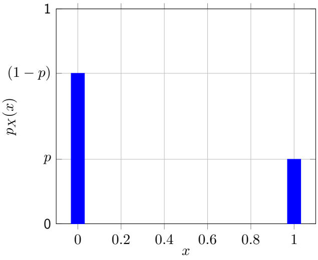
Распределение Бернулли

$$X \sim \mathrm{Bern}(p) \Leftrightarrow \begin{cases} p_X(0) = (1-p), \\ p_X(1) = p, \\ p_X(x) = 0 \text{ для всех остальных } x. \end{cases}$$

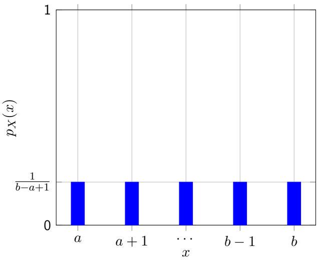
Распределение Бернулли

$$X \sim \mathrm{Bern}(p) \Leftrightarrow \begin{cases} p_X(0) = (1-p), \\ p_X(1) = p, \\ p_X(x) = 0 \text{ для всех остальных } x. \end{cases}$$

- p — параметр распределения, вероятность успеха
- С.в. X может быть индикатором события A:
  - $\omega \in A \Rightarrow X(\omega) = 1$
  - $\omega \notin A \Rightarrow X(\omega) = 0$



Равномерное дискретное распределение



$$X \sim U(a,b) = \begin{cases} p_X(x) = \frac{1}{b-a+1}, & \text{если } x \in [a..b], \\ p_X(x) = 0, & \text{иначе.} \end{cases}$$

- a, b — целочисленные параметры ($a \le b$)
- $\Omega = [a..b]$
- $X(\omega) = \omega$

Биномиальное распределение

Проведем n независимых опытов по схеме Бернулли с вероятностью успеха p. X — число успехов.

$$X \sim \mathrm{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^{n} X_i,$$

где все $X_i \sim \mathrm{Bern}(p)$.

Биномиальное распределение

Проведем n независимых опытов по схеме Бернулли с вероятностью успеха p. X — число успехов.

$$X \sim \mathrm{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^{n} X_i,$$

где все $X_i \sim \mathrm{Bern}(p)$.

$$p_X(x) = ?$$

Биномиальное распределение

Проведем n независимых опытов по схеме Бернулли с вероятностью успеха p. X — число успехов.

$$X \sim \mathrm{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^{n} X_i,$$

где все $X_i \sim \mathrm{Bern}(p)$.

$$p_X(x) =?$$

- Вероятность конкретного исхода, в котором ровно x успехов:

$$\Pr(\omega) = p^x (1-p)^{n-x}$$

- Всего таких исходов $\binom{n}{x}$

Биномиальное распределение

Проведем n независимых опытов по схеме Бернулли с вероятностью успеха p. X — число успехов.

$$X \sim \mathrm{Bin}(n,p) \Leftrightarrow X = \sum_{i=1}^{n} X_i,$$

где все $X_i \sim \mathrm{Bern}(p)$.

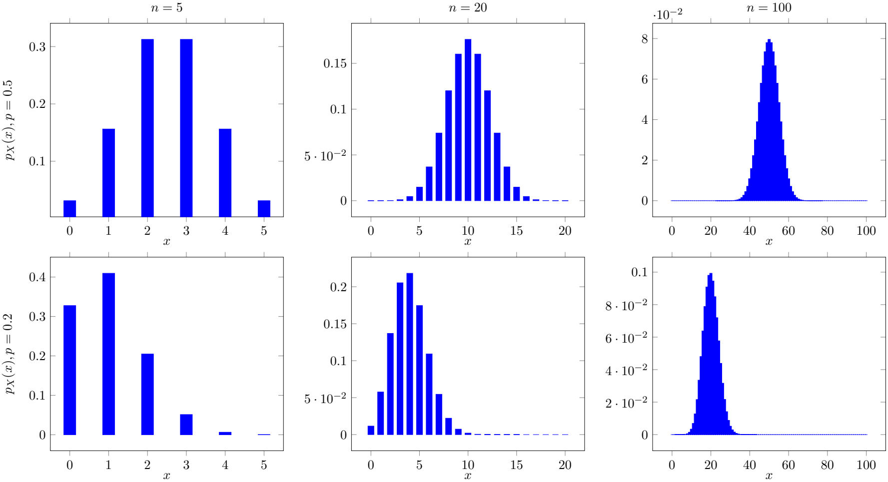
$$\boxed{p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}}$$

- Вероятность конкретного исхода, в котором ровно x успехов:

$$\Pr(\omega) = p^x (1-p)^{n-x}$$

- Всего таких исходов $\binom{n}{x}$

Функции вероятностей биномиального распределения



NB: везде разные масштабы оси OY

Геометрическое распределение

Проводим независимые эксперименты по схеме Бернулли с вероятностью успеха p до первого успешного исхода. X — число экспериментов, которые мы проводим.

$$X \sim Geom(p) \Leftrightarrow X = \min\{i \in \mathbb{N} \mid X_i = 1\},$$

где $X_i \sim \mathrm{Bern}(p)$.

Геометрическое распределение

Проводим независимые эксперименты по схеме Бернулли с вероятностью успеха p до первого успешного исхода. X — число экспериментов, которые мы проводим.

$$X \sim Geom(p) \Leftrightarrow X = \min\{i \in \mathbb{N} \mid X_i = 1\},$$

где $X_i \sim \mathrm{Bern}(p)$.

$$p_X(x) = ?$$

Геометрическое распределение

Проводим независимые эксперименты по схеме Бернулли с вероятностью успеха p до первого успешного исхода. X — число экспериментов, которые мы проводим.

$$X \sim Geom(p) \Leftrightarrow X = \min\{i \in \mathbb{N} \mid X_i = 1\},$$

где $X_i \sim \mathrm{Bern}(p)$.

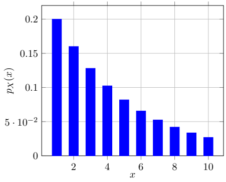
$$p_X(x) = \quad p(1-p)^{x-1}$$

- Вероятность, что первые $x-1$ неуспешны:

$$\Pr(X_1 = X_2 = \dots X_{x-1} = 0) = (1-p)^{x-1}$$

- Вероятность, что x-ый исход успешен:

$$\Pr(X_x = 1) = p$$



Уточнение

Есть исход, при котором $X = +\infty$, но его вероятность равна нулю.

$$\Pr(X = \infty) \leq \Pr(X_1 = \dots = X_n = 0) = (1 - p)^n$$

для любого n. Если допустить, что $\Pr(X = \infty) = q > 0$, то возьмем $n \geq \log_{1-p} q$, получим, что вероятность подсобытия больше, чем вероятность события (Противоречие со свойствами вероятностной меры).

Математическое ожидание с.в.

Если много раз повторить один и тот же эксперимент, то каков будет средний результат?

- Бросаем честную кость d6. Что в среднем?
- $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$

Математическое ожидание с.в.

Если много раз повторить один и тот же эксперимент, то каков будет средний результат?

- Бросаем честную кость d6. Что в среднем?
- $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$

$$E(X) = \sum_x x \cdot p_X(x)$$

Математическое ожидание с.в.

Если много раз повторить один и тот же эксперимент, то каков будет средний результат?

- Бросаем честную кость d6. Что в среднем?
- $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$

$$E(X) = \sum_x x \cdot p_X(x)$$

NB: Матожидание определено только когда этот ряд абсолютно сходится

Математическое ожидание с.в.

Если много раз повторить один и тот же эксперимент, то каков будет средний результат?

- Бросаем честную кость d6. Что в среднем?
- $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$

$$E(X) = \sum_x x \cdot p_X(x)$$

NB: Матожидание определено только когда этот ряд абсолютно сходится

NB-2: Матожидание — функционал на множестве с.в.

Матожидание некоторых с.в.

$X \sim \mathrm{Bern}(p)$

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$X \sim U(a, b)$

$$E(X) = \sum_{i=a}^{b} \frac{i}{b - a + 1} = \frac{a + b}{2}$$

Элементарные свойства матожидания

- Неотрицательность: $X \geq 0 \Rightarrow E(X) \geq 0$
- Ограниченность: $X \in [a, b] \Rightarrow E(X) \in [a, b]$
- Матожидание константы: $E(c) = c$

Функции от с.в. и их матожидание

Пусть есть с.в. X и функция $g : \mathbb{R} \to \mathbb{R}$. Тогда $Y = g(X)$ — тоже с.в.
Как посчитать $E(Y)$?

Функции от с.в. и их матожидание

Пусть есть с.в. X и функция $g : \mathbb{R} \to \mathbb{R}$. Тогда $Y = g(X)$ — тоже с.в.
Как посчитать $E(Y)$?

- по определению: $E(Y) = \sum_y y p_Y(y)$

Функции от с.в. и их матожидание

Пусть есть с.в. X и функция $g : \mathbb{R} \to \mathbb{R}$. Тогда $Y = g(X)$ — тоже с.в.
Как посчитать $E(Y)$?

- по определению: $E(Y) = \sum_y y p_Y(y)$
- $$\boxed{E(Y) = \sum_x g(x) p_X(x)}$$

Функции от с.в. и их матожидание

Пусть есть с.в. X и функция $g : \mathbb{R} \to \mathbb{R}$. Тогда $Y = g(X)$ — тоже с.в.

Как посчитать $E(Y)$?

- по определению: $E(Y) = \sum_y y p_Y(y)$
- $\boxed{E(Y) = \sum_x g(x) p_X(x)}$

$$\sum_x g(x) p_X(x) = \sum_y \sum_{x: g(x) = y} g(x) p_X(x)$$

$$= \sum_y y \sum_{x: g(x) = y} p_X(x) = \sum_y y p_Y(y) = E(Y)$$

Линейность матожидания

$$E(aX + b) = aE(X) + b$$

Линейность матожидания

$$E(aX + b) = aE(X) + b$$

Позволяет легко пересчитывать матожидание с.в. не по определению.

- Пусть X принимает четные значения от 0 до 100 с равной вероятностью.
- $X = 2Y$, где $Y \sim U(0, 50)$
- $E[X] = 2E[Y] = 50$

Линейность матожидания

$$E(aX + b) = aE(X) + b$$

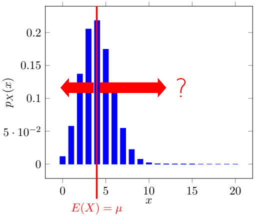
Позволяет легко пересчитывать матожидание с.в. не по определению.

- Пусть X принимает четные значения от 0 до 100 с равной вероятностью.
- $X = 2Y$, где $Y \sim U(0, 50)$
- $E[X] = 2E[Y] = 50$

NB: Для всех линейных функций g выполнено $E(g(X)) = g(E(X))$, но в общем случае это неверно

Дисперсия

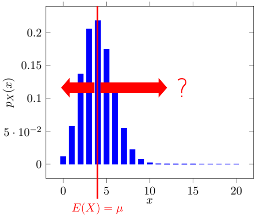
Иногда хотим оценить то, насколько с.в. отклоняется от своего среднего значения μ.



Дисперсия

Иногда хотим оценить то, насколько с.в. отклоняется от своего среднего значения μ.

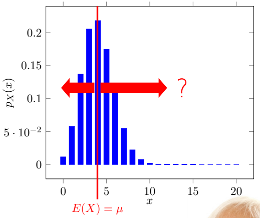
- Первая мысль: посчитать $E(X - \mu)$



Дисперсия

Иногда хотим оценить то, насколько с.в. отклоняется от своего среднего значения μ.

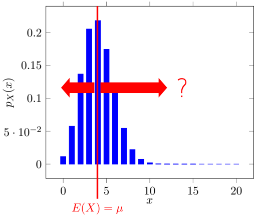
- Первая мысль: посчитать $E(X-\mu)$
- $E(X-\mu) = E(X) - \mu = 0$



Дисперсия

Иногда хотим оценить то, насколько с.в. отклоняется от своего среднего значения μ.

- Первая мысль: посчитать $E(X-\mu)$
- $E(X-\mu) = E(X) - \mu = 0$



$$\boxed{\mathrm{Var}(X) = E((X-\mu)^2)}$$

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X - \mu)^2)}$$

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X - \mu)^2)}$$

- Дисперсия считается как мотожидание функции от с.в.

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X-\mu)^2)}$$

- Дисперсия считается как мотожидание функции от с.в.
- Всегда неотрицательна, равна нулю только у детерминированной с.в.

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X-\mu)^2)}$$

- Дисперсия считается как мотожидание функции от с.в.
- Всегда неотрицательна, равна нулю только у детерминированной с.в.
- Почему не посчитать $|X-\mu|$? Просто так удобнее. Чем — узнаем позже

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X - \mu)^2)}$$

- Дисперсия считается как мотожидание функции от с.в.
- Всегда неотрицательна, равна нулю только у детерминированной с.в.
- Почему не посчитать $|X - \mu|$? Просто так удобнее. Чем — узнаем позже
- Часто полезно знать среднеквадратичное отклонение

$$\boxed{\sigma(X) = \sqrt{\mathrm{Var}(X)}}$$

Несколько замечаний

$$\boxed{\mathrm{Var}(X) = E((X-\mu)^2)}$$

- Дисперсия считается как мотожидание функции от с.в.
- Всегда неотрицательна, равна нулю только у детерминированной с.в.
- Почему не посчитать $|X-\mu|$? Просто так удобнее. Чем — узнаем позже
- Часто полезно знать среднеквадратичное отклонение

$$\boxed{\sigma(X) = \sqrt{\mathrm{Var}(X)}}$$

- В дальнейшем часто будем обозначать $\mu := E(X)$, если понятно, про какую с.в. речь

Свойства дисперсии

$$\boxed{\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)}$$

- Умножая с.в. на a — увеличиваем ее вариацию в a^2 раз
- Прибавляя к с.в. b — не меняем вариацию

Свойства дисперсии

$$\boxed{\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)}$$

- Умножая с.в. на a — увеличиваем ее вариацию в a^2 раз
- Прибавляя к с.в. b — не меняем вариацию

$$\begin{aligned}
\mathrm{Var}(aX+b) &= E\left((aX+b-E(aX+b))^2\right)\\
&= E\left((aX+b-aE(X)-b)^2\right)\\
&= E\left((aX-aE(X))^2\right)\\
&= E\left(a^2(X-E(X))^2\right)\\
&= a^2E\left((X-E(X))^2\right)\\
&= a^2\,\mathrm{Var}(X)
\end{aligned}$$

Свойства дисперсии

$$\boxed{\mathrm{Var}(X) = E(X^2) - (E(X))^2}$$

Свойства дисперсии

$$\boxed{\operatorname{Var}(X) = E(X^2) - (E(X))^2}$$

$$\begin{aligned}
\operatorname{Var}(X) &= E\left((X-\mu)^2\right) \\
&= E\left(X^2 - 2X\mu + \mu^2\right) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}$$

Дисперсия распределения Бернулли

- $X \sim \mathrm{Bern}(p)$

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (E(X))^2 = E(X) - (E(X))^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

Дисперсия распределения Бернулли

- $X \sim \mathrm{Bern}(p)$

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - (E(X))^2 = E(X) - (E(X))^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

- Если кидаем честную монету, мы всегда отклоняемся от μ на $\frac{1}{2}$

$$\sigma(x) = \sqrt{p(1-p)} = \left[p = \frac{1}{2}\right] = \frac{1}{2}$$

Дисперсия равномерного распределения

- $X \sim U(a, b)$ — можем посчитать вариацию $Y = X - a \sim U(0, n)$, где $n = b - a$.

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(Y + a) = \mathrm{Var}(Y) = E(Y^2) - (E(Y))^2 \\
&= \sum_{i=0}^{n} \frac{i^2}{n+1} - \left(\frac{n}{2}\right)^2 \\
&= \frac{1}{n+1} \cdot \frac{n(n+1)(2n+1)}{6} - \frac{n^2}{4} \\
&= \frac{4n^2 + 2n - 3n^2}{12} = \frac{n(n+2)}{12} \\
&= \frac{(b-a)(b-a+2)}{12}
\end{aligned}
$$

Условная с.в. (на событии)

Ничего особо нового:

- $p_X(x \mid A) = \Pr(X = x \mid A)$

Условная с.в. (на событии)

Ничего особо нового:

- $p_X(x \mid A) = \Pr(X = x \mid A)$
- Легко проверить свойства функции вероятностей:
  - $p_X(x \mid A) \geq 0$
  - $\sum_x p_X(x \mid A) = 1$

Условная с.в. (на событии)

Ничего особо нового:

- $p_X(x \mid A) = \Pr(X = x \mid A)$
- Легко проверить свойства функции вероятностей:
  - $p_X(x \mid A) \geq 0$
  - $\sum_x p_X(x \mid A) = 1$
- $E(X \mid A) = \sum_x x \cdot p_X(x \mid A)$

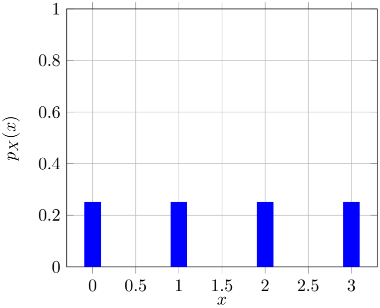
Условная с.в. (на событии)

Ничего особо нового:

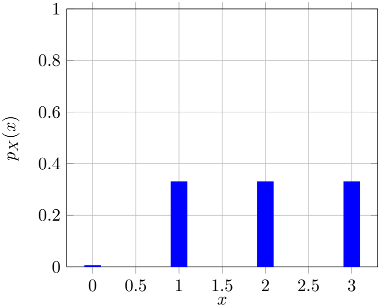
- $p_X(x \mid A) = \Pr(X = x \mid A)$
- Легко проверить свойства функции вероятностей:
  - $p_X(x \mid A) \geq 0$
  - $\sum_x p_X(x \mid A) = 1$
- $E(X \mid A) = \sum_x x \cdot p_X(x \mid A)$
- $E(g(X) \mid A) = \sum_x g(x) \cdot p_X(x \mid A)$

Пример условной с.в.

Безусловное распределение $X \sim U(0, 3)$:



Условное распределение $X \mid X \geq 1$:



Теорема о полном матожидании

Слышали о полной вероятности?

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid (A_i))$$

Теорема о полном матожидании

Слышали о полной вероятности?

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B \mid (A_i))$$

Но событием B может быть событие $X = x$

$$p_X(x) = \sum_i \Pr(A_i) p_X(x \mid A_i)$$

Теорема о полном матожидании

Слышали о полной вероятности?

$$\Pr(B) = \sum_i \Pr(A_i)\Pr(B \mid (A_i))$$

Но событием B может быть событие $X = x$

$$p_X(x) = \sum_i \Pr(A_i) p_X(x \mid A_i)$$

Домножим левую часть на x и просуммируем по всем x

$$\begin{aligned}\sum_x x \cdot p_X(x) &= \sum_x \sum_i x \Pr(A_i) p_X(x \mid A_i) \\ &= \sum_i \Pr(A_i) \sum_x x \cdot p_X(x \mid A_i) \\ &= \sum_i \Pr(A_i) E(X \mid A_i)\end{aligned}$$

Теорема о полном матожидании

$$E(X) = \sum_i \Pr(A_i) E(X \mid A_i)$$

Беспамятство геометрического распределения

У геометрического распределения нет памяти. Пусть $X \sim \mathrm{Geom}(p)$.

$$p_{(X-1)}(x \mid X > 1) = \frac{\Pr(X = x + 1 \cap X > 1)}{\Pr(X > 1)} = \frac{\Pr(X = x + 1)}{1 - p}$$
$$= \frac{p(1-p)^x}{1-p} = p(1-p)^{x-1} = p_X(x)$$

Беспамятство геометрического распределения

У геометрического распределения нет памяти. Пусть $X \sim \mathrm{Geom}(p)$.

$$p_{(X-1)}(x \mid X > 1) = \frac{\Pr(X = x + 1 \cap X > 1)}{\Pr(X > 1)} = \frac{\Pr(X = x + 1)}{1 - p}$$
$$= \frac{p(1-p)^x}{1-p} = p(1-p)^{x-1} = p_X(x)$$

Можно также показать, что $p_{X-n}(x \mid X > n) = p_X(x)$

Беспамятство геометрического распределения

У геометрического распределения нет памяти. Пусть $X \sim \mathrm{Geom}(p)$.

$$p_{(X-1)}(x \mid X>1) = \frac{\Pr(X=x+1 \cap X>1)}{\Pr(X>1)} = \frac{\Pr(X=x+1)}{1-p}$$
$$= \frac{p(1-p)^x}{1-p} = p(1-p)^{x-1} = p_X(x)$$

Можно также показать, что $p_{X-n}(x \mid X>n) = p_X(x)$

При условии, что первые n исходов — неудачные, число оставшихся бросков следует тому же распределению $\mathrm{Geom}(p)$

Матожидание геометрического распределения

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

$$
\begin{aligned}
E(X) &= 1 + E(X-1) \\
&= 1 + \Pr(X=1)E(X-1 \mid X=1) \\
&+ \Pr(X>1)E(X-1 \mid X>1) \\
&= 1 + 0 + (1-p)E(X) \\
pE(X) &= 1 \\
E(X) &= \frac{1}{p}
\end{aligned}
$$

Матожидание геометрического распределения

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

$$\begin{aligned}
E(X) &= 1 + E(X-1) \\
&= 1 + \Pr(X=1)E(X-1 \mid X=1) \\
&+ \Pr(X>1)E(X-1 \mid X>1) \\
&= 1 + 0 + (1-p)E(X) \\
pE(X) &= 1 \\
E(X) &= \frac{1}{p}
\end{aligned}$$

Но можно то же самое вычислить в лоб, посчитав сумму ряда:

$$E(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1}$$

Несколько с.в. (случайный вектор)

Совместная функция вероятностей

$$p_{X,Y}(x,y) = \Pr(X = x \cap Y = y)$$

Маргинальные функции вероятностей

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

Пример случайного вектора

В клетках — $p_{X,Y}(x,y)$

Y				
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		
	1	2	3	4

X

Пример случайного вектора

В клетках — $p_{X,Y}(x,y)$

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

$$\xrightarrow{\Sigma} p_Y(2) = \frac{5}{20}$$

Пример случайного вектора

В клетках — $p_{X,Y}(x, y)$

$Y \backslash X$	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

$\xrightarrow{\Sigma} p_Y(2) = \frac{5}{20}$

$\Sigma \downarrow$ $p_X(2) = \frac{8}{20}$

Пример случайного вектора

В клетках — $p_{X,Y}(x, y)$

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

$\xrightarrow{\Sigma} p_Y(2) = \frac{5}{20}$

$\xrightarrow{\Sigma} p_X(2) = \frac{8}{20}$

NB: все работает точно также и для большего числа с.в.

Линейность матожидания (опять)

Теперь у нас есть возможность смотреть $Z = g(X, Y)$ ☺

Линейность матожидания (опять)

Теперь у нас есть возможность смотреть $Z = g(X, Y)$ ☺

$$p_Z(z) = \Pr(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$$

Линейность матожидания (опять)

Теперь у нас есть возможность смотреть $Z = g(X, Y)$ ☺

$$p_Z(z) = \Pr(g(X,Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x,y)$$

$$E(Z) = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

Линейность матожидания (опять)

Теперь у нас есть возможность смотреть $Z = g(X, Y)$ ☺

$$p_Z(z) = \Pr(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$$

$$E(Z) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

Пусть $Z = X + Y$ как вычислить ее матожидание?

Линейность матожидания (опять)

Теперь у нас есть возможность смотреть $Z = g(X, Y)$ ☺

$$p_Z(z) = \Pr(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$$

$$E(Z) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

Пусть $Z = X + Y$ как вычислить ее матожидание?Хорошо, что оно линейно

$$E(Z) = E(X) + E(Y)$$

Линейность матожидания: доказательство

$$
\begin{aligned}
E(Z) = E(X+Y) &= \sum_x \sum_y (x+y) p_{X,Y}(x,y) \\
&= \sum_x \sum_y x p_{X,Y}(x,y) + \sum_x \sum_y y p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) + \sum_y y p_Y(y) \\
&= E(X) + E(Y)
\end{aligned}
$$

Окончательная линейность матожидания

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$$

NB: случайная величина может быть равна константе, поэтому мы не рассматриваем прибавление к сумме константы

Матожидание биномиального распределения

Напоминание: $X \sim \mathrm{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

Матожидание биномиального распределения

Напоминание: $X \sim \mathrm{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

$X = \sum_{i=1}^{n} X_i$, где все $X_i \sim \mathrm{Bern}(p)$

Матожидание биномиального распределения

Напоминание: $X \sim \mathrm{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

$X = \sum_{i=1}^{n} X_i$, где все $X_i \sim \mathrm{Bern}(p)$

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Матожидание биномиального распределения

Напоминание: $X \sim \mathrm{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

$X = \sum_{i=1}^{n} X_i$, где все $X_i \sim \mathrm{Bern}(p)$

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Всегда можно посчитать в лоб

$$E(X) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$$

Условная с.в. (на другой с.в.)

Было:

$p_X(x \mid A) = \Pr(X = x \mid A)$

Стало:

$p_{X|Y}(x \mid y) = \Pr(X = x \mid Y = y)$

Условная с.в. (на другой с.в.)

Было:

$p_X(x \mid A) = \Pr(X = x \mid A)$

Стало:

$p_{X|Y}(x \mid y) = \Pr(X = x \mid Y = y)$

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

NB: Условная вероятность определена только для y, у которых $p_Y(y) > 0$

Условная с.в. (на другой с.в.)

Было:

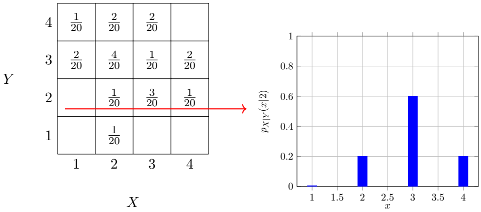
$p_X(x \mid A) = \Pr(X = x \mid A)$

Стало:

$p_{X|Y}(x \mid y) = \Pr(X = x \mid Y = y)$

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

NB: Условная вероятность определена только для y, у которых $p_Y(y) > 0$



Условные случайные векторы

$$p_{X_1,\ldots X_n|Y_1,\ldots,Y_m}(x_1,\ldots x_m \mid y_1,\ldots,y_m)$$
$$= \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n \mid Y_1 = y_1 \cap \ldots Y_m = y_m)$$

Условные случайные векторы

$$p_{X_1,\ldots X_n|Y_1,\ldots,Y_m}(x_1,\ldots x_m \mid y_1,\ldots,y_m)$$
$$= \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n \mid Y_1 = y_1 \cap \ldots Y_m = y_m)$$

Чуть проще:

- $p_{X|Y,Z}(x \mid y, z) = \Pr(X = x \mid Y = y \cap Z = z)$
- $p_{X,Y|Z}(x, y \mid z) = \Pr(X = x \cap Y = y \mid Z = z)$

Условные случайные векторы

$$p_{X_1,\ldots X_n|Y_1,\ldots,Y_m}(x_1,\ldots x_m \mid y_1,\ldots,y_m)$$
$$= \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n \mid Y_1 = y_1 \cap \ldots Y_m = y_m)$$

Чуть проще:

- $p_{X|Y,Z}(x \mid y,z) = \Pr(X = x \mid Y = y \cap Z = z)$
- $p_{X,Y|Z}(x,y \mid z) = \Pr(X = x \cap Y = y \mid Z = z)$

Работает правило умножения:

- Было: $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A \cap B)$
- Стало: $p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x,y)$

Условное маотжидание

$$E(X \mid Y = y) = \sum_x x p_{X|Y}(x \mid y)$$

Условия просто меняют вероятностную меру, поэтому с условными с.в. все работает так же, как и с безусловными

Полные вероятность и матожидание

Формула полной вероятности (оригинал): для разбиения Ω на A_i

$$p_X(x) = \sum_i \Pr(A_i) p_X(x \mid A_i)$$

Полные вероятность и матожидание

Формула полной вероятности (оригинал): для разбиения Ω на A_i

$$p_X(x) = \sum_i \Pr(A_i) p_X(x \mid A_i)$$

Давайте теперь скажем, что $A_i = (Y = y)$, получим

$$\boxed{p_X(x) = \sum_y p_Y(y) p_{X|Y}(x \mid y)}$$

Полные вероятность и матожидание

Формула полной вероятности (оригинал): для разбиения Ω на A_i

$$p_X(x) = \sum_i \Pr(A_i) p_X(x \mid A_i)$$

Давайте теперь скажем, что $A_i = (Y = y)$, получим

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x \mid y)$$

Полное матожидание — по аналогии

$$E(X) = \sum_y p_Y(y) E(X \mid Y = y)$$

NB: работает только когда ряд сходится абсолютно

Независимость и условная независимость с.в.

- Напоминание:A и B независимы $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$

Независимость и условная независимость с.в.

- Напоминание: A и B независимы $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$
- Событие A и с.в. X независимы $\Leftrightarrow$

$$\forall x \quad \Pr(A \cap X = x) = \Pr(A) p_X(x)$$

- X и Y — независимые с.в. $\Leftrightarrow$

$$\forall x, y \quad p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

Независимость и условная независимость с.в.

- Напоминание: A и B независимы $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$
- Событие A и с.в. X независимы $\Leftrightarrow$

$$\boxed{\forall x \quad \Pr(A \cap X = x) = \Pr(A) p_X(x)}$$

- X и Y — независимые с.в. $\Leftrightarrow$

$$\boxed{\forall x, y \quad p_{X,Y}(x, y) = p_X(x) p_Y(y)}$$

NB: независимость по-прежнему означает, что одно событие (или с.в.) не дает никакой информации о другом событии (или с.в)

Условная независимость с.в.

Условие порождает новую меру $\Rightarrow$ зависимость с.в. может поменяться в зависимости от условия

Условная независимость с.в.

Условие порождает новую меру $\Rightarrow$ зависимость с.в. может поменяться в зависимости от условия

- X и Y, очевидно, зависимы:
  - $p_X(1) = \frac{3}{20}$
  - $p_{X|Y}(1,1) = 0$

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

Условная независимость с.в.

Условие порождает новую меру $\Rightarrow$ зависимость с.в. может поменяться в зависимости от условия

$Y \backslash X$	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

- X и Y, очевидно, зависимы:
  - $p_X(1) = \frac{3}{20}$
  - $p_{X|Y}(1, 1) = 0$
- Но при условии $C = (X \leq 2 \cap Y \geq 3)$ становятся независимы:
  - $p_X(1 \mid C) = \frac{1}{3}$, $p_X(2 \mid C) = \frac{2}{3}$
  - $p_Y(3 \mid C) = \frac{2}{3}$, $p_Y(4 \mid C) = \frac{1}{3}$

Условная независимость с.в.

Условие порождает новую меру $\Rightarrow$ зависимость с.в. может поменяться в зависимости от условия

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

- X и Y, очевидно, зависимы:
  - $p_X(1) = \frac{3}{20}$
  - $p_{X|Y}(1, 1) = 0$
- Но при условии $C = (X \le 2 \cap Y \ge 3)$ становятся независимы:
  - $p_X(1 \mid C) = \frac{1}{3}$, $p_X(2 \mid C) = \frac{2}{3}$
  - $p_Y(3 \mid C) = \frac{2}{3}$, $p_Y(4 \mid C) = \frac{1}{3}$
- Поэтому,
  - $p_{X,Y}(1, 4 \mid C) = \frac{1}{9} = p_X(1 \mid C)p_Y(4 \mid C)$
  - $p_{X,Y}(1, 3 \mid C) = \frac{2}{9} = p_X(1 \mid C)p_Y(3 \mid C)$
  - $p_{X,Y}(2, 4 \mid C) = \frac{2}{9} = p_X(2 \mid C)p_Y(4 \mid C)$
  - $p_{X,Y}(2, 3 \mid C) = \frac{4}{9} = p_X(2 \mid C)p_Y(3 \mid C)$

Матожидание независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$\boxed{E(XY) = E(X)E(Y)}$$

Матожидание независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$E(XY) = E(X)E(Y)$$

$$E(XY) = \sum_x \sum_y xyp_{X,Y}(x,y) = \sum_x \sum_y xyp_X(x)p_Y(y)$$
$$= \sum_x xp_X(x) \sum_y yp_Y(y) = E(X)E(Y)$$

Дисперсия независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)}$$

Дисперсия независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)}$$

Предположим, что $E(X) = E(Y) = 0$.
Тогда $E(XY) = E(X)E(Y) = 0$

Дисперсия независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$\boxed{\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)}$$

Предположим, что $E(X)=E(Y)=0$.
Тогда $E(XY)=E(X)E(Y)=0$

$$\begin{aligned}\operatorname{Var}(X+Y)&=E\left((X+Y)^2\right)=E\left(X^2+2XY+Y^2\right)\\&=E(X^2)+2E(XY)+E(Y^2)=E(X^2)+E(Y^2)\\&=\operatorname{Var}(X)+\operatorname{Var}(Y)\end{aligned}$$

Дисперсия биномиалки

Напоминание: $X \sim \mathrm{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^{n} X_i$, где все $X_i \sim \mathrm{Bern}(p)$ — независимы

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$= \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Лекция 3. Остаток про дискретные с.в. и введение в непрерывные

24 февраля 2021 г.

1 Случайный вектор

Несколько с.в. (случайный вектор)

Совместная функция вероятностей

$$p_{X,Y}(x,y) = \Pr(X = x \cap Y = y)$$

Маргинальные функции вероятностей

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$
$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

Пример случайного вектора В клетках — $p_{X,Y}(x,y)$

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

$\xrightarrow{\Sigma} p_Y(2) = \frac{5}{20}$

$\xrightarrow{\Sigma} p_X(2) = \frac{8}{20}$

NB: все работает точно также и для большего числа с.в.

2 Линейность матожидания

Линейность матожидания (опять) Теперь у нас есть возможность смотреть $Z = g(X, Y)$

$$p_Z(z) = \Pr(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$$

$$E(Z) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

Пусть $Z = X + Y$ как вычислить ее матожидание? Хорошо, что оно *линейно*

$$E(Z) = E(X) + E(Y)$$

Линейность матожидания: доказательство

$$\begin{aligned} E(Z) = E(X + Y) &= \sum_x \sum_y (x + y) p_{X,Y}(x, y) \\ &= \sum_x \sum_y x p_{X,Y}(x, y) + \sum_x \sum_y y p_{X,Y}(x, y) \\ &= \sum_x x \sum_y p_{X,Y}(x, y) + \sum_y y \sum_x p_{X,Y}(x, y) \\ &= \sum_x x p_X(x) + \sum_y y p_Y(y) \\ &= E(X) + E(Y) \end{aligned}$$

Окончательная линейность матожидания

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$$

NB: случайная величина может быть равна *константе*, поэтому мы не рассматриваем прибавление к сумме константы

Матожидание биномиального распределения *Напоминание:* $X \sim \mathrm{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

$X = \sum_{i=1}^{n} X_i$, где все $X_i \sim \mathrm{Bern}(p)$

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Всегда можно посчитать *в лоб*

$$E(X) = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$$

3 С.в., условная на другой с.в.

Условная с.в. (на другой с.в.)

Было:

$p_X(x \mid A) = \Pr(X = x \mid A)$

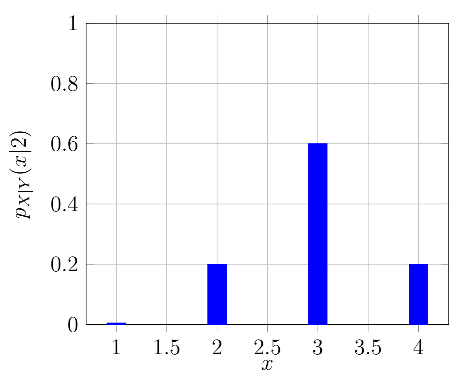
Стало:

$p_{X|Y}(x \mid y) = \Pr(X = x \mid Y = y)$

$$\boxed{p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}}$$

NB: Условная вероятность определена только для y, у которых $p_Y(y) > 0$

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		



Условные случайные векторы

$$\boxed{\begin{aligned} &p_{X_1,\ldots X_n|Y_1,\ldots,Y_m}(x_1, \ldots x_m \mid y_1, \ldots, y_m) \\ &= \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n \mid Y_1 = y_1 \cap \ldots Y_m = y_m) \end{aligned}}$$

Чуть проще:

- $p_{X|Y,Z}(x \mid y,z) = \Pr(X = x \mid Y = y \cap Z = z)$
- $p_{X,Y|Z}(x,y \mid z) = \Pr(X = x \cap Y = y \mid Z = z)$

Работает правило умножения:

- Было: $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A \cap B)$
- Стало: $p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x,y)$

Условное маотжидание

$$E(X \mid Y = y) = \sum_x x p_{X|Y}(x \mid y)$$

Условия просто *меняют вероятностную меру*, поэтому с условными с.в. все работает так же, как и с безусловными

Полные вероятность и матожидание Формула полной вероятности (оригинал): для разбиения Ω на A_i

$$p_X(x) = \sum_i \Pr(A_i)p_X(x \mid A_i)$$

Давайте теперь скажем, что $A_i = (Y = y)$, получим

$$p_X(x) = \sum_y p_Y(y)p_{X|Y}(x \mid y)$$

Полное матожидание — по аналогии

$$E(X) = \sum_y p_Y(y)E(X \mid Y = y)$$

NB: работает только когда ряд сходится *абсолютно*

4 Независимость

Независимость и условная независимость с.в.

- *Напоминание*:A и B независимы $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$
- Событие A и с.в. X независимы $\Leftrightarrow$

$$\boxed{\forall x \quad \Pr(A \cap X = x) = \Pr(A) p_X(x)}$$

- X и Y — *независимые* с.в. $\Leftrightarrow$

$$\boxed{\forall x, y \;\; p_{X,Y}(x, y) = p_X(x) p_Y(y)}$$

NB: независимость по-прежнему означает, что одно событие (или с.в.) *не дает никакой информации* о другом событии (или с.в)

Условная независимость с.в. Условие порождает новую меру $\Rightarrow$ зависимость с.в. может поменяться в зависимости от условия

Y \ X	1	2	3	4
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
1		$\frac{1}{20}$		

- X и Y, очевидно, зависимы:
  - $p_X(1) = \frac{3}{20}$
  - $p_{X|Y}(1, 1) = 0$
- Но при условии $C = (X \leq 2 \cap Y \geq 3)$ становятся независимы:
  - $p_X(1 \mid C) = \frac{1}{3}$, $p_X(2 \mid C) = \frac{2}{3}$
  - $p_Y(3 \mid C) = \frac{2}{3}$, $p_Y(4 \mid C) = \frac{1}{3}$

- Поэтому,
  - $p_{X,Y}(1,4 \mid C) = \frac{1}{9} = p_X(1 \mid C)p_Y(4 \mid C)$
  - $p_{X,Y}(1,3 \mid C) = \frac{2}{9} = p_X(1 \mid C)p_Y(3 \mid C)$
  - $p_{X,Y}(2,4 \mid C) = \frac{2}{9} = p_X(2 \mid C)p_Y(4 \mid C)$
  - $p_{X,Y}(2,3 \mid C) = \frac{4}{9} = p_X(2 \mid C)p_Y(3 \mid C)$

Матожидание независимых с.в.
Пусть X и Y — *независимые* с.в., тогда

$$\boxed{E(XY) = E(X)E(Y)}$$

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyp_{X,Y}(x,y) = \sum_x \sum_y xyp_X(x)p_Y(y) \\ &= \sum_x xp_X(x) \sum_y yp_Y(y) = E(X)E(Y) \end{aligned}$$

Дисперсия независимых с.в.
Пусть X и Y — *независимые* с.в., тогда

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)}$$

Предположим, что $E(X) = E(Y) = 0$.
Тогда $E(XY) = E(X)E(Y) = 0$

$$\begin{aligned} \mathrm{Var}(X+Y) &= E\left((X+Y)^2\right) = E\left(X^2 + 2XY + Y^2\right) \\ &= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) \end{aligned}$$

Дисперсия биномиалки *Напоминание:* $X \sim \mathrm{Bin}(n,p) \Leftrightarrow X = \sum_{i=1}^n X_i$, где все $X_i \sim \mathrm{Bern}(p)$ — *независимы*

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) \\ &= \sum_{i=1}^n p(1-p) = np(1-p) \end{aligned}$$

5 Более сложные распределения

Гипер-геометрическое распределение

Эксперимент: у нас в корзине есть N шаров, D из которых белые, $N-D$ — черные. Мы достаем случайные n шаров (не возвращая их в корзину). X — число белых шаров, которые мы достали. Тогда $X \sim HG(D, N, n)$

Матожидание: представляем X_i как сумму индикаторных величин для всех белых шаров, что он был вытащен. Вероятность успеха $p = \frac{n}{N}$

$$E[X] = \sum_{i\text{- белый шар}} E[X_i] = D \cdot \frac{n}{N} = \frac{nD}{N}$$

Дисперсия:

$$E[X^2] = \sum_{i\text{- белый шар}} E[X_i^2] + \sum_{i,j\text{- белые шары}} E[X_i X_j] = E[X] + D(D-1) \cdot \frac{n}{N} \cdot \frac{n-1}{N-1}$$
$$= \frac{nD(N-n)(N-D)}{N^2(N-1)}$$

При больших N очень близко к биномиальному распределению $\mathrm{Bin}(n, \frac{D}{N})$.

Степенное распределение

$X \sim \mathrm{pow}(u, \beta)$ значит, что мы выбираем число из $[1..u]$, причем $\Pr(X = i) \sim i^{-\beta}$. Более точно:

$$\Pr[X = i] = \begin{cases} C_{u,\beta} i^{-\beta}, \text{ если } i \in [1..u], 0, \text{иначе.} \end{cases}$$

$C_{u,\beta}$ — коэффициент нормализации

$$C_{u,\beta} = \left(\sum_{i=1}^{u} i^{-\beta}\right)^{-1} = H_{u,\beta}^{-1}$$

$H_{u,\beta}^{-1}$ — обобщенное гармоническое число

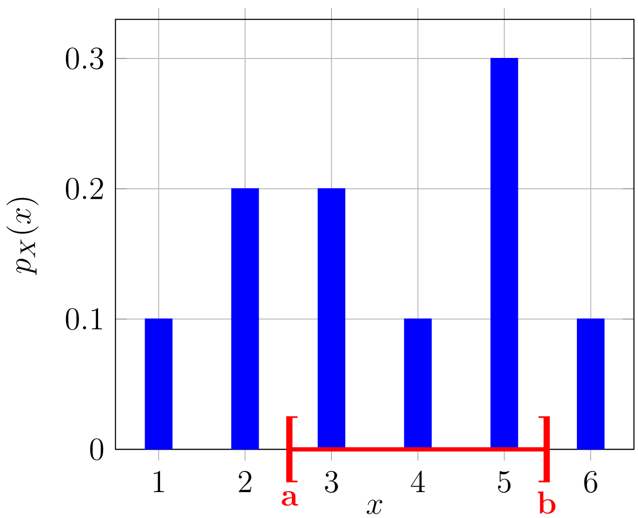
Замечание: если $\beta > 1$, то u может быть $+\infty$

Рассказать про приближение интегралами

6 Непрерывные случайные величины

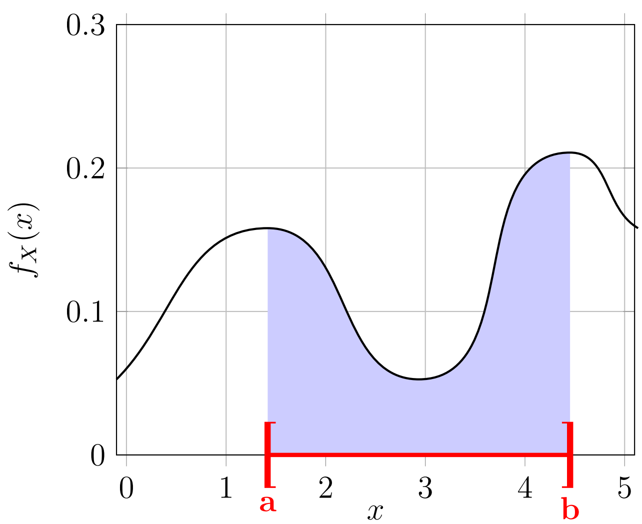
Определение непрерывных с.в.

В случае с дискретными величинами:



$$\Pr[a \leq X \leq b] = \sum_{x: a \leq x \leq b} p_X(x)$$

Но если X может принимать люые значения из этого интервала? Тогда нам нужна функция, которая показывает, сколько вероятностной массы лежит на каждом элементарном отрезке.



$$\Pr[a \leq X \leq b] = \int_a^b f_X(x)dx \tag{1}$$

Определение: Случайная величина называется непрерывной, если для нее существует такая функция $f_X(x)$, что для любых $a, b \in \mathbb{R}$ (где $a \leq b$) верно (1).

$f_X(x)$ — плотность вероятности с.в. X:

$$\Pr(a \leq X \leq a + \varepsilon) \approx f_X(a) \cdot \varepsilon$$

Плотность вероятности — аналог функции вероятностей для непрерывных с.в.:

- $p_X(x) \geq 0$
- $\sum_x p_X(x) = 1$

То же самое

- $f_X(x) \geq 0$
- $\int_{-\infty}^{+\infty} p_X(x) = 1$

NB:

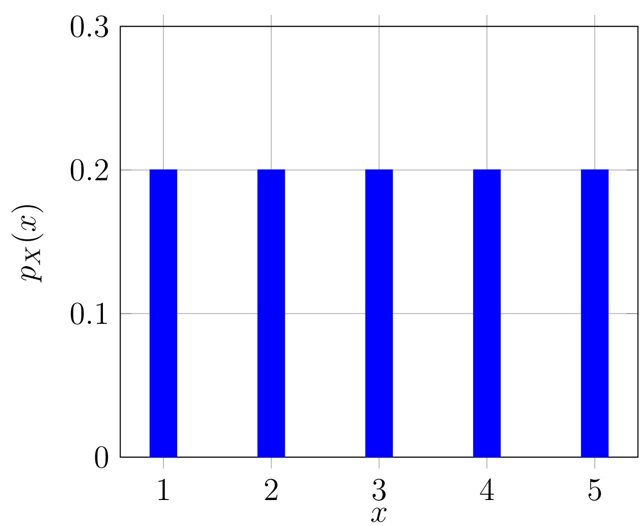
$$\Pr(X = a) = \Pr(a \leq X \leq a) = \int_a^a f_X(x)dx = 0$$

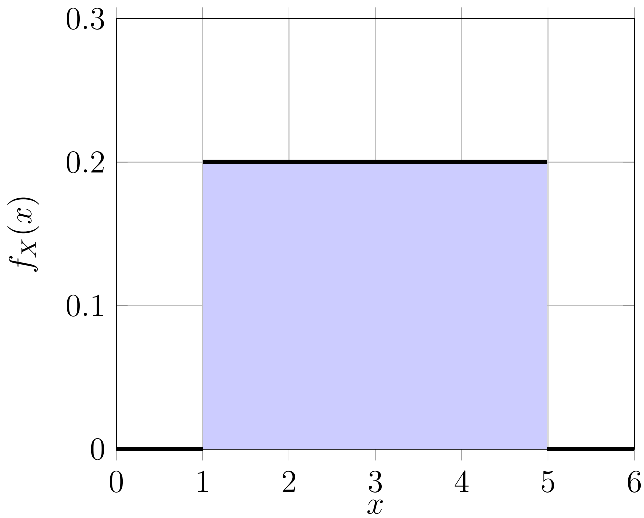
Поэтому:

$$\Pr(a \leq X \leq b) = \Pr(X = a) + \Pr(X = b) + \Pr(a < X < b) = \Pr(a < X < b)$$

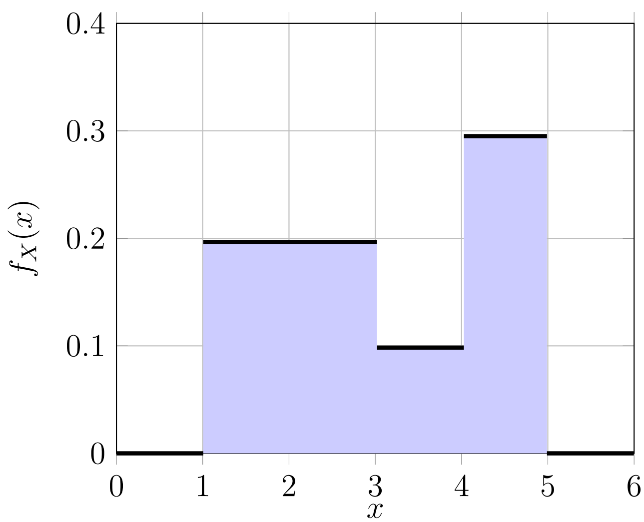
NB: Мы ушли от понятия событий, но у нас по-прежнему есть какая-то Ω, на которой и задана с.в. X. Просто нам сейчас проще быть чисто в терминах с.в.

Пример: равномерное распределение





Обобщение: частично равномерное распределение



7 Матожидание

Для дискретных величин:

$$E(X) = \sum_x x p_X(x)$$

Для непрерывных: заменяем сумму на интеграл, а функцию вероятностей на плотность вероятности

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Важно: интеграл должен сходиться абсолютно
Интерпретация: центр масс вероятностной массы
Свойства матожидания

- $X \geq 0 \Rightarrow E[X] > 0$
- $X \in [a, b] \Rightarrow E[X] \in [a, b]$
- Матожидание функции от с.в.:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

- Пример

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

- Линейность: $E(aX + b) = aE(X) + b$

8 Дисперсия

Как и для дискретных:

$$\operatorname{Var}(X) = E((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx$$

среднеквадратичное отклонение:

$$\sigma_X = \sqrt{\operatorname{Var}(X)}$$

Свойства — те же:

- $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$
- $\operatorname{Var}(X) = E(X^2) - (E(X))^2$

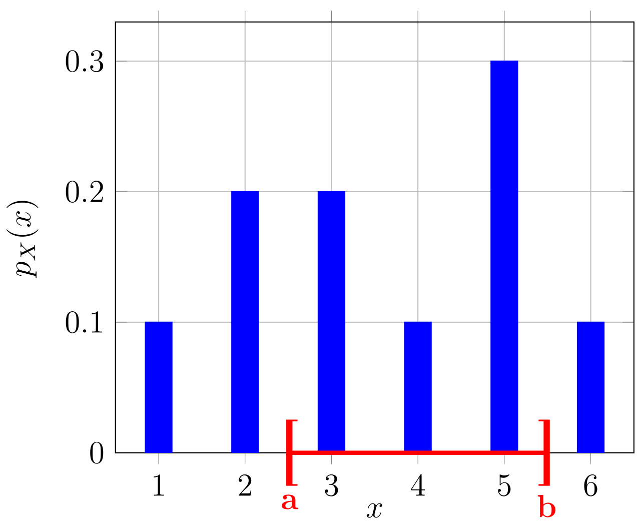
Лекция 3. Непрерывные с.в., часть 1

2 марта 2021 г.

1 Непрерывные случайные величины

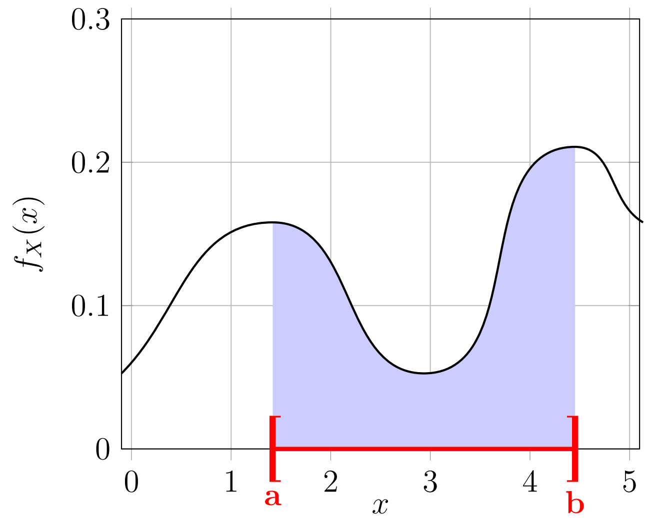
Определение непрерывных с.в.

В случае с дискретными величинами:



$$\Pr[a \le X \le b] = \sum_{x: a \le x \le b} p_X(x)$$

Но если X может принимать любые вещественные значения из этого интервала? Тогда нам нужна функция, которая показывает, сколько вероятностной массы лежит на каждом элементарном отрезке.



$$\Pr[a \leq X \leq b] = \int_a^b f_X(x)dx \tag{1}$$

Определение: Случайная величина называется непрерывной, если для нее существует такая функция $f_X(x)$, что для любых $a, b \in \mathbb{R}$ (где $a \leq b$) верно (1).

$f_X(x)$ — *плотность вероятности* с.в. X:

$$\Pr(a \leq X \leq a + \varepsilon) \approx f_X(a) \cdot \varepsilon$$

Плотность вероятности — аналог функции вероятностей для непрерывных с.в.:

- $p_X(x) \geq 0$
- $\sum_x p_X(x) = 1$

То же самое

- $f_X(x) \geq 0$
- $\int_{-\infty}^{+\infty} p_X(x) = 1$

NB:

$$\Pr(X = a) = \Pr(a \leq X \leq a) = \int_a^a f_X(x)dx = 0$$

Поэтому:

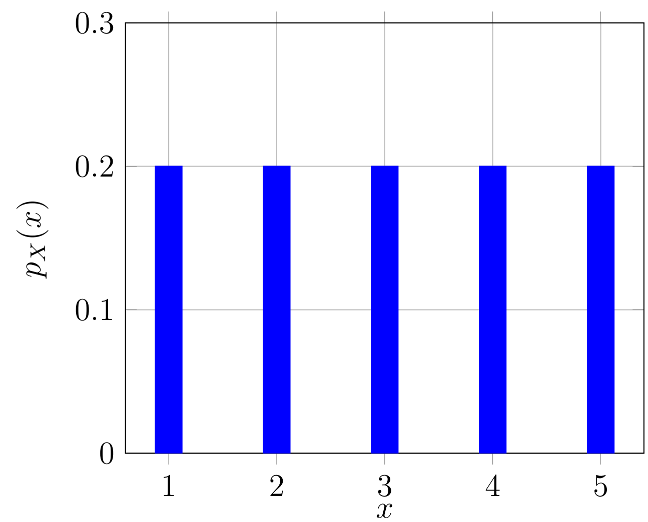
$$\Pr(a \leq X \leq b) = \Pr(X = a) + \Pr(X = b) + \Pr(a < X < b) = \Pr(a < X < b)$$

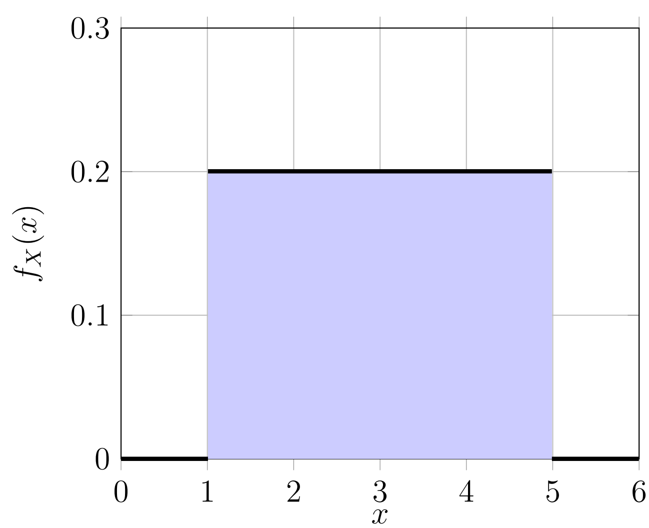
NB: Мы ушли от понятия событий, но у нас по-прежнему есть какая-то Ω, на которой и задана с.в. X. Просто нам сейчас проще быть чисто в терминах с.в.

NB: Переопредилим дискретные с.в. как с.в., для которых есть функция вероятностей, то есть число возможныъх значений которых счетно.

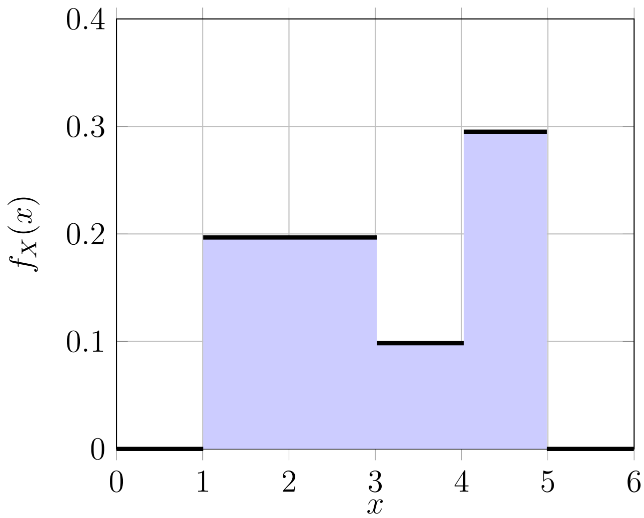
NB: Вы уже могли догадаться, что с.в. могут быть и смешанные, но про это позже.

Пример: равномерное распределение





Обобщение: частично равномерное распределение



2 Матожидание

Для дискретных величин:

$$E(X) = \sum_x x p_X(x)$$

Для непрерывных: заменяем сумму на интеграл, а функцию вероятностей на плотность вероятности

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Важно: интеграл должен сходиться абсолютно

Интерпретация: центр масс вероятностной массы

Свойства матожидания

- $X \geq 0 \Rightarrow E[X] > 0$
- $X \in [a, b] \Rightarrow E[X] \in [a, b]$
- Матожидание функции от с.в.:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

- Пример

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

- Линейность: $E(aX + b) = aE(X) + b$

3 Дисперсия

Как и для дискретных:

$$\mathrm{Var}(X) = E((X-\mu)^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x) dx$$

среднеквадратичное отклонение:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

Свойства — те же:

- $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$
- $\mathrm{Var}(X) = E(X^2) - (E(X))^2$

4 Моменты стандартных распределений

NB: i-й момент распределения — $\int_{-\infty}^{+\infty} x^i f_X(x) dx$ (для дискретных с.в — сумма)

Равномерное распределение

$X \sim U(a,b)$

Матожидание:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

Дисперсия:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3} \\ \mathrm{Var}(X) = E[X^2] - (E[X])^2 &= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Экспоненциальное распределение

Говорим, что X следует экспоненциальному распределению $\mathrm{Exp}(\lambda)$ с параметром λ, если

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

NB: Это аналог геометрического распределения с параметром $p = \lambda$.

Матожидание:

$$E[X] = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = -\int_0^{+\infty} x de^{-\lambda x} = -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx$$
$$= 0 - \frac{1}{\lambda}\Big|_0^{+\infty} = \frac{1}{\lambda}.$$

Дисперсия (два раза интегрируем по частям):

$$E[X] = \int_0^{+\infty} x^2\lambda e^{-\lambda x} dx = \frac{2}{\lambda}. \operatorname{Var}(X) \qquad = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Довольно хорошо сконцентрирована, так как вероятность хвоста экспоненциально падает:

$$\Pr(X \geq a) = \int_a^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda a}.$$

5 Функция распределения

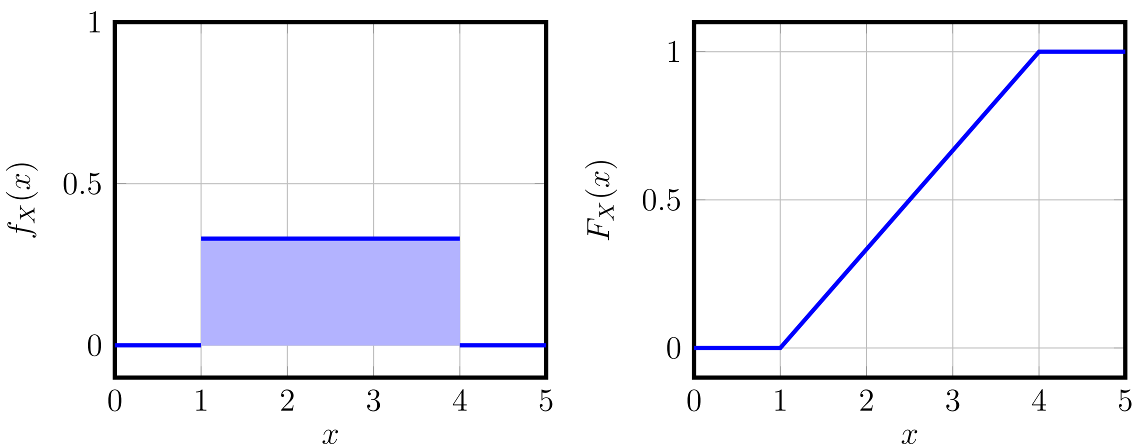
$F_X(x) = \Pr(X \leq x)$ — *функция распределения* с.в. X (как дискретной, так и непрерывной).

Как считать:

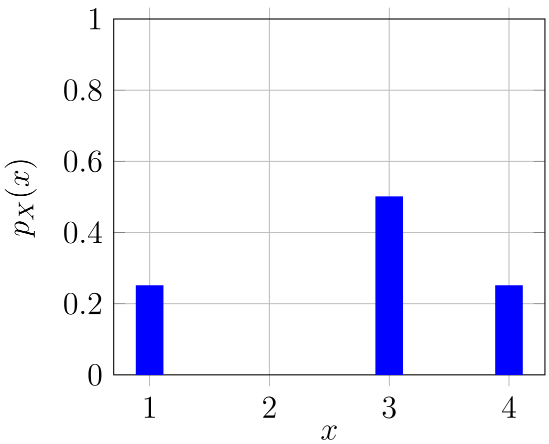
$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt$$

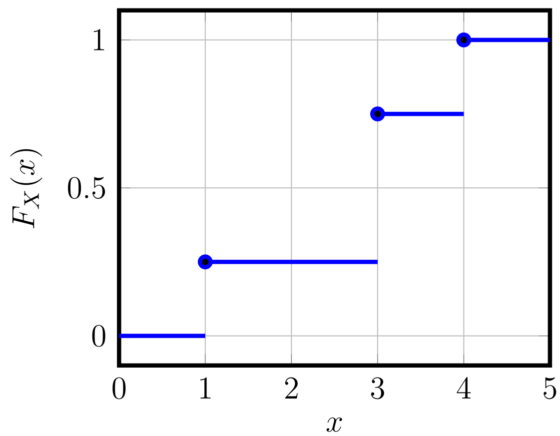
Легко заметить: $F'_X(x) = f_X(x)$

Пример: равномерное распределение.



Функцию распределения можно считать и для дискретной случайной величины:





Свойства функции распределения:

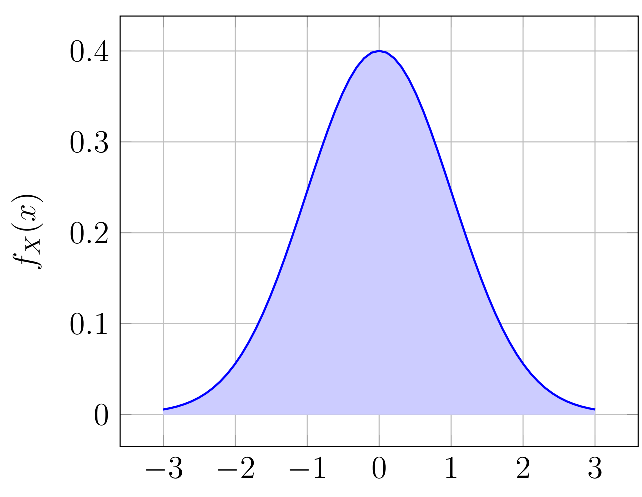
- Неубывающая
- $\lim_{x \to +\infty} F_X(x) = 1$
- $\lim_{x \to -\infty} F_X(x) = 0$

6 Нормальное распределение (распределение Гаусса)

Очень важная штука:

- Важна в центральной предельной теореме
- Часто на практике неизвестные распределения приближаются нормальным

Стандартное нормальное: $X \sim N(0,1) \leftrightarrow f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$



Откуда берется коэффициент нормализации $\frac{1}{\sqrt{2\pi}}$? Из интеграла Гаусса

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Свойства $N(0,1)$:

$$E[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0,$$

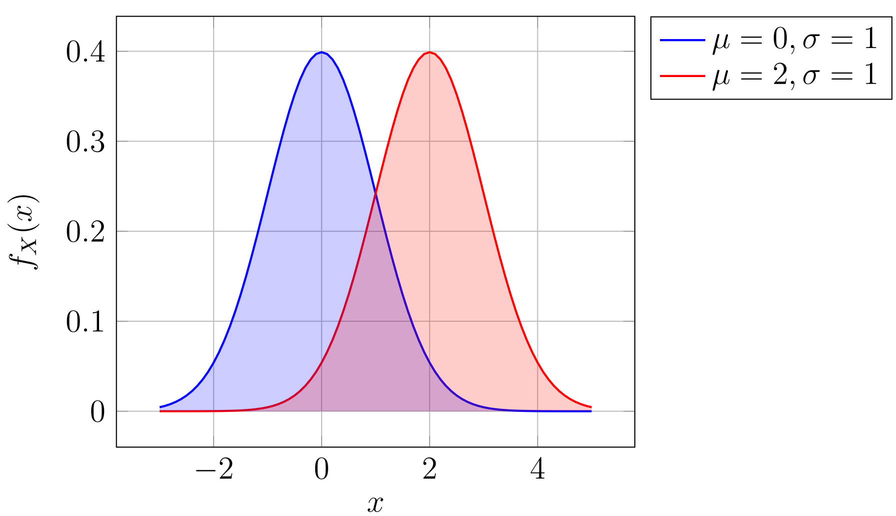
так как это интеграл нечетной функции, которая в бесконечности очень маленькая.

$$\begin{aligned}
\mathrm{Var}(X) = E[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} d\frac{x^2}{2} \\
= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} de^{-x^2/2} \\
= -\frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx \\
= 0 + \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1.
\end{aligned}$$

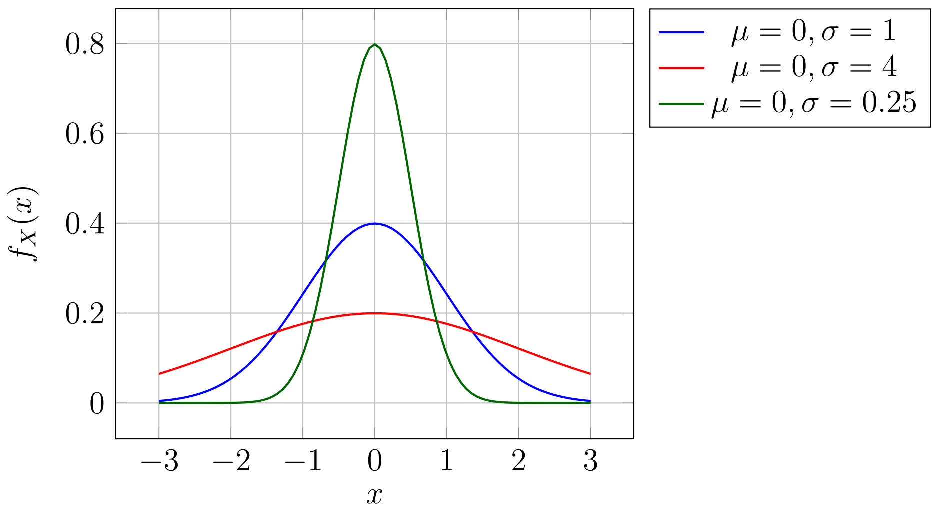
В записи $N(0,1)$ нолик как раз обозначает матожидание, а единица — дисперсию

Обобщенное нормальное распределение: $X \sim N(\mu, \sigma^2) \Leftrightarrow f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- μ — новое матожидание, насколько мы сдвигаем ось симметрии X.
- σ — новая дисперсия, насколько мы растягиваем распределение от оси симметрии.



Чем меньше σ (срежнеквадратичное отклонение), тем больше распределение сжато вокруг оси симметрии



Полезное свойство нормальной случайной величины: если $X \sim N(\mu, \sigma^2)$ и при этом $Y = aX + b$, то Y тоже имеет нормальное распределение. Это мы докажем позже, но пока давайте посмотрим, какому именно распределению.

- $E(Y)$ должно равняться $E(aX + b) = aE(X) + b = a\mu + b$
- $\operatorname{Var}(Y)$ должна равняться $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X) = a^2\sigma^2$

Значит, $Y \sim N(a\mu + b, a^2\sigma^2)$.

Функция распределения $X \sim N(0, 1)$ обозначается буквой Φ:

$$\Phi(x) = F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Так как интеграл неберущийся, ее значения считаются с помощью таблиц.

Как работать с такими таблицами. Пусть мы хотим посчитать вероятность того, что $X \sim N(0, 1)$ не больше, чем 2.39. Находим строчку 2.3, столбец 0.09, смотрим на число в их пересечении, добавляем к нему 0.5. Если хотим посчитать $\Pr(X \leq -2.39)$, то вычитаем это число из 0.5.

Как пользоваться таблицей, если $X \sim N(\mu, \sigma^2)$? И пусть мы хотим найти вероятность $\Pr(X \in [a, b])$. Для этого Давайте рассмотрим $Y = \frac{X-\mu}{\sigma}$. Заметим, что $Y \sim N(0, 1)$. Теперь запишем интересующее нас событие следующим образом

Таблица 1: Таблица для вычисления функции распределения стандартного нормального распределения $N(0,1)$. В i-й строке и j-ом столбце число равно $\int_0^{0.1i+0.01j} f_X(x)dx$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

$$
\begin{aligned}
a \leq X && \leq b \\
a-\mu \leq X-\mu && \leq b-\mu \\
\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} && \leq \frac{b-\mu}{\sigma}
\end{aligned}
$$

То есть искомая вероятность равна $\Pr(Y \in [\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}])$, а это уже вычисляется по таблице.

7 Условная плотность вероятности

Напомним трактовку плотность вероятности. Это то, сколько вероятностной массы приходится на маленький интервал значений с.в.:

$$f_X(x)\delta \approx \Pr(X \in [x, x+\delta])$$

Но вероятностная мера может меняться при условии, что произошло событие A. В этом случае определяем

$$f_{X|A}(x)\delta \approx \Pr(X \in [x, x+\delta] \mid A)$$

Точное определение: плотность вероятности с.в. X при условии A есть такая функция $f_{X|A}(x)$, что для любого измеримого множества B верно, что

$$\Pr(X \in B \mid A) = \int_B f_{X|A}(x)dx.$$

NB: у нас опять просто поменялась вероятностная мера. То есть у условной плотности вероятности будут все те же свойства:

- $f_{X|A}(x) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|A}(x)dx = 1$

Вычисляем аналогично условной функции вероятности. Пусть $x \in A$, тогда

$$
\begin{aligned}
f_{X|A}(x)\delta \approx \Pr(X \in [x, x+\delta] \mid A) &= \frac{\Pr(X \in [x, x+\delta] \cap X \in A)}{\Pr(A)} \\
&= \frac{\Pr(X \in [x, x+\delta])}{\Pr(A)} \approx \frac{f_X(x)\delta}{\Pr(A)}
\end{aligned}
$$

Поэтому строго говоря, она вычисляется так:

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\Pr(A)}, & \text{если } x \in A, \\ 0, & \text{иначе.} \end{cases}$$

То есть масштабируем плотность вероятности по событию A.

8 Условное матожидание

Условное матожидание определяем аналогично:

$$E(X \mid A) = \int_{-\infty}^{+\infty} x f_{X|A}(x) dx$$

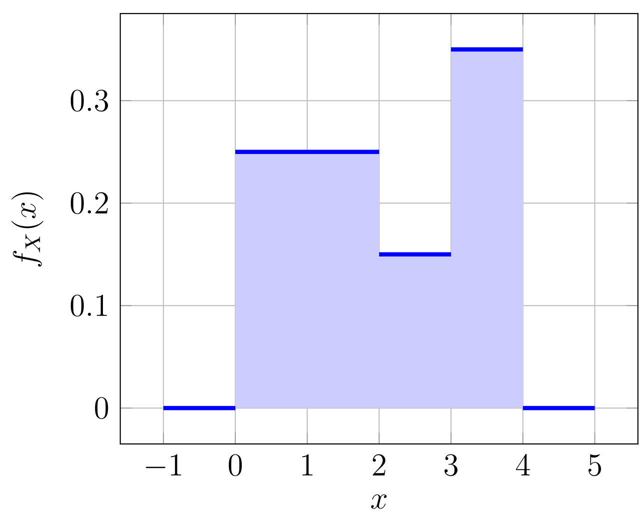
Работает то же самое правило и для функций от с.в. (у нас просто новая плотность вероятности при условии A):

$$E(g(X) \mid A) = \int_{-\infty}^{+\infty} g(x) f_{X|A}(x) dx$$

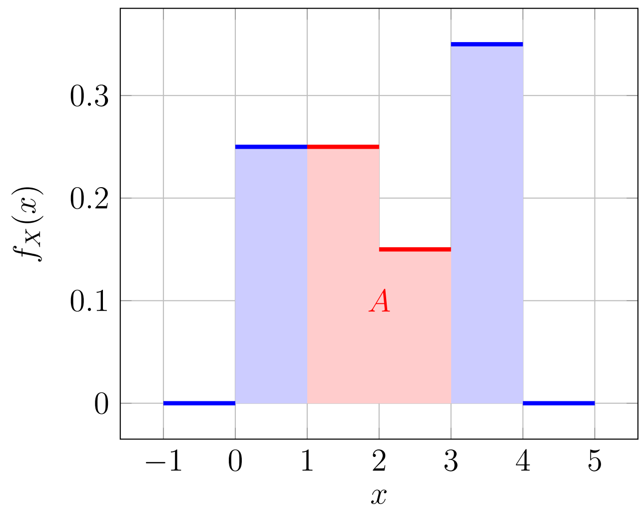
9 Пример условных с.в.

Рассмотрим частично равномерное распределение:

$$f_X(x) = \begin{cases} 0.25, x \in [0, 2) \\ 0.15, x \in [2, 3) \\ 0.35, x \in [3, 4] \\ 0, \text{ иначе.} \end{cases}$$

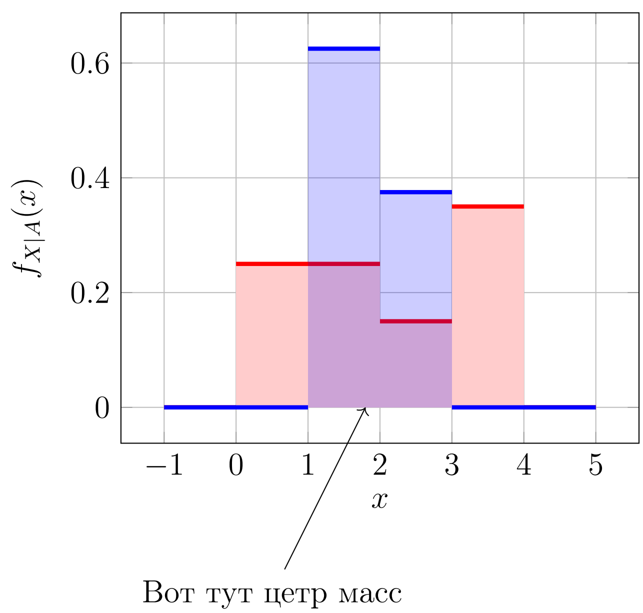


Пусть $A = [1, 3]$. Тогда $\Pr(A) = 0.25 \cdot 1 + 0.15 \cdot 1 = 0.4$



Значит, новая плотность вероятности выглядит так:

$$f_{X|A}(x) = \begin{cases} \frac{0.25}{0.4} = 0.625, x \in [1, 2) \\ \frac{0.15}{0.4} = 0.375, x \in [2, 3) \\ 0, \text{ иначе.} \end{cases}$$



$$E(X \mid A) = \int_1^2 t \cdot 0.625dt + \int_2^3 t \cdot 0.375dt = 1.875$$

10 Беспамятство экспоненциального распределения

Как уже говорилось, экспоненциальное распределение очень похоже на геометрическое. В том числе вот почему. Пусть продолжительность жизни лампочки T следует $\mathrm{Exp}(\lambda)$. Следует ли поменять лампочку после того, как она проработала время t? Посмотрим, сколько она еще проживет, то есть распределение $T-t$ при условии $T \geq t$.

$$\begin{aligned} F_{(T-t)|T\geq t}(x) = \Pr(T-t \geq x \mid T > t) &= \frac{\Pr(T-t \geq x \cap T > t)}{\Pr(T>t)} \\ = \frac{\Pr(T \geq x+t)}{\Pr(T>t)} &= \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x}, \end{aligned}$$

то есть распределение точно то же, как если мы заменим лампочку на новую.

11 Полные вероятность и матожидание

Напомним: пусть есть разбиение Ω на $\{A_i\}$ (не более, чем счетное), тогда

$$\begin{aligned} \Pr(B) &= \Pr(A_1)\Pr(B \mid A_1) + \dots + \Pr(A_n)\Pr(B \mid A_n) + \dots \\ p_X(x) &= \Pr(A_1)p_{X|A_1}(x) + \dots + \Pr(A_n)p_{X|A_n}(x) + \dots \end{aligned}$$

Ничего не меняется и в непрерывном случае. Сначала функция распределения:

$$\begin{aligned} F_X(x) = \Pr(X \leq x) &= \Pr(A_1)\Pr(X \leq x \mid A_1) + \dots + \Pr(A_n)\Pr(X \leq x \mid A_n) + \dots \\ &= \Pr(A_1)F_{X|A_1}(x) + \dots + \Pr(A_n)F_{X|A_n}(x) + \dots \end{aligned}$$

Дифференцируем, получаем:

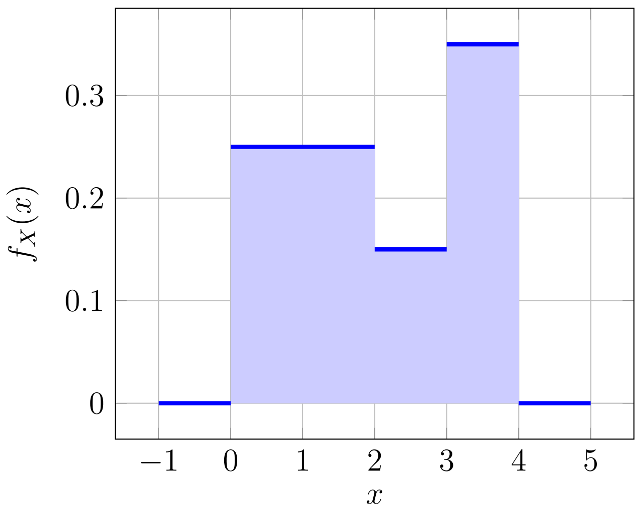
$$f_X(x) = \Pr(A_1)f_{X|A_1}(x) + \dots + \Pr(A_n)f_{X|A_n}(x) + \dots$$

Умножаем на x и интегрируем по всему $\mathbb{R}$:

$$E(X) = \Pr(A_1)E(X \mid A_1) + \dots + \Pr(A_n)E(X \mid A_n) + \dots$$

Пример (который уже был):

$$f_X(x) = \begin{cases} 0.25, x \in [0,2) \\ 0.15, x \in [2,3) \\ 0.35, x \in [3,4] \\ 0, \text{ иначе.} \end{cases}$$



Посчитаем матожидание

$$\begin{aligned} E(X) = \Pr(X \in [0,2))E(X \mid X \in [0,2)) \\ + \Pr(X \in [2,3))E(X \mid X \in [2,3)) \\ + \Pr(X \in [3,4))E(X \mid X \in [3,4)) \end{aligned}$$

Заметим, что на каждом отрезке матожидание – это положение центра масс, то есть середина отрезка. Поэтому

$$E(X) = 0.5 \cdot 1 + 0.15 \cdot 2.5 + 0.35 \cdot 3.5 = 2.1$$

Лекция 5. Непрерывные с.в., часть 2

10 марта 2021 г.

1 Условная плотность вероятности

Напомним трактовку плотность вероятности. Это то, сколько вероятностной массы приходится на маленький интервал значений с.в.:

$$f_X(x)\delta \approx \Pr(X \in [x, x+\delta])$$

Но вероятностная мера может меняться при условии, что произошло событие A. В этом случае определяем

$$f_{X|A}(x)\delta \approx \Pr(X \in [x, x+\delta] \mid A)$$

Точное определение: плотность вероятности с.в. X при условии A есть такая функция $f_{X|A}(x)$, что для любого измеримого множества B верно, что

$$\Pr(X \in B \mid A) = \int_B f_{X|A}(x)dx.$$

NB: у нас опять просто поменялась вероятностная мера. То есть у условной плотности вероятности будут все те же свойства:

- $f_{X|A}(x) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|A}(x)dx = 1$

Вычисляем аналогично условной функции вероятности. Пусть $x \in A$, тогда

$$f_{X|A}(x)\delta \approx \Pr(X \in [x, x+\delta] \mid A) = \frac{\Pr(X \in [x, x+\delta] \cap X \in A)}{\Pr(A)}$$
$$= \frac{\Pr(X \in [x, x+\delta])}{\Pr(A)} \approx \frac{f_X(x)\delta}{\Pr(A)}$$

Поэтому строго говоря, она вычисляется так:

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\Pr(A)}, & \text{если } x \in A, \\ 0, & \text{иначе.} \end{cases}$$

То есть масштабируем плотность вероятности по событию A.

Иногда придется иметь дело с условной функцией распределения:

$$F_{X|A}(x) = \Pr(X \leq x \mid A)$$

2 Условное матожидание

Условное матожидание определяем аналогично:

$$E(X \mid A) = \int_{-\infty}^{+\infty} x f_{X|A}(x) dx$$

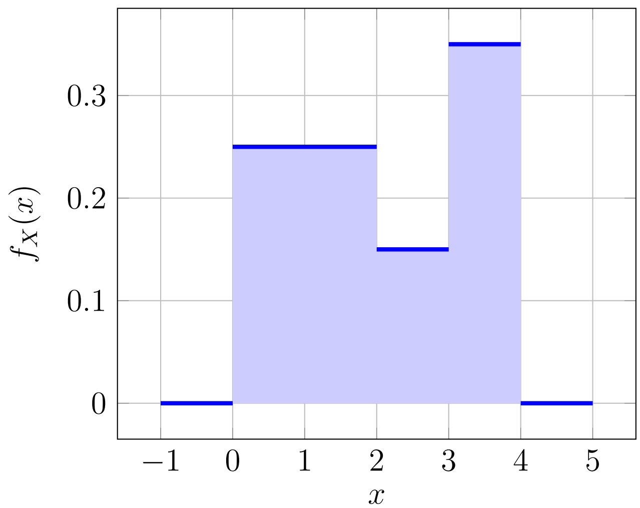
Работает то же самое правило и для функций от с.в. (у нас просто новая плотность вероятности при условии A):

$$E(g(X) \mid A) = \int_{-\infty}^{+\infty} g(x) f_{X|A}(x) dx$$

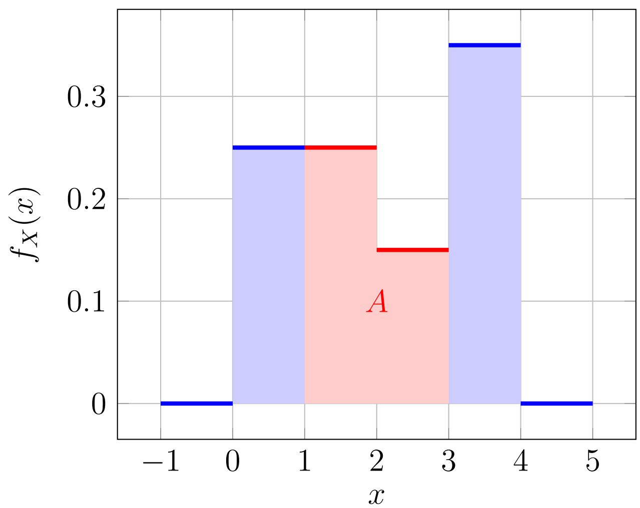
3 Пример условных с.в.

Рассмотрим частично равномерное распределение:

$$f_X(x) = \begin{cases} 0.25, x \in [0, 2) \\ 0.15, x \in [2, 3) \\ 0.35, x \in [3, 4] \\ 0, \text{ иначе.} \end{cases}$$

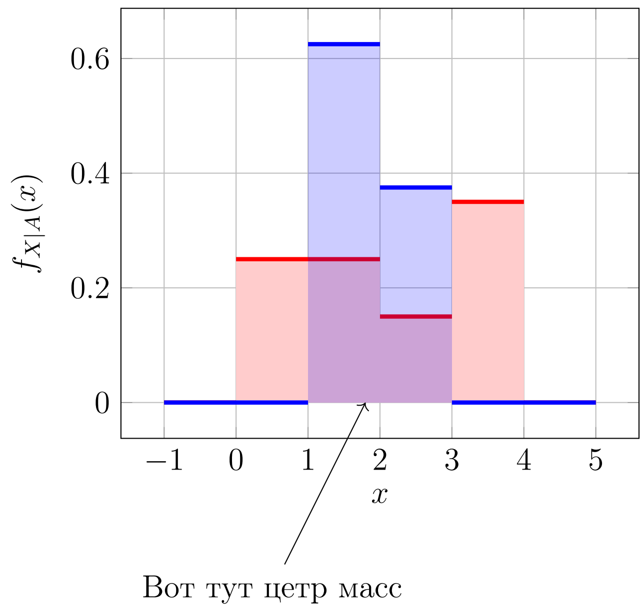


Пусть $A = [1, 3]$. Тогда $\Pr(A) = 0.25 \cdot 1 + 0.15 \cdot 1 = 0.4$



Значит, новая плотность вероятности выглядит так:

$$f_{X|A}(x) = \begin{cases} \frac{0.25}{0.4} = 0.625, x \in [1, 2) \\ \frac{0.15}{0.4} = 0.375, x \in [2, 3) \\ 0, \text{ иначе.} \end{cases}$$



$$E(X \mid A) = \int_1^2 t \cdot 0.625 dt + \int_2^3 t \cdot 0.375 dt = 1.875$$

4 Беспамятство экспоненциального распределения

Как уже говорилось, экспоненциальное распределение очень похоже на геометрическое. В том числе вот почему. Пусть продолжительность жизни лампочки T следует $\mathrm{Exp}(\lambda)$. Следует ли поменять лампочку после того, как она проработала время t? Посмотрим, сколько она еще проживет, то есть распределение $T - t$ при условии $T \geq t$.

$$F_{(T-t)|T \geq t}(x) = \Pr(T - t \geq x \mid T > t) = \frac{\Pr(T - t \geq x \cap T > t)}{\Pr(T > t)}$$
$$= \frac{\Pr(T \geq x + t)}{\Pr(T > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x},$$

то есть распределение точно то же, как если мы заменим лампочку на новую.

5 Полные вероятность и матожидание

Напомним: пусть есть разбиение Ω на $\{A_i\}$ (не более, чем счетное), тогда

$$\Pr(B) = \Pr(A_1) \Pr(B \mid A_1) + \dots + \Pr(A_n) \Pr(B \mid A_n) + \dots$$
$$p_X(x) = \Pr(A_1) p_{X|A_1}(x) + \dots + \Pr(A_n) p_{X|A_n}(x) + \dots$$

Ничего не меняется и в непрерывном случае. Сначала функция распределения:

$$\begin{aligned} F_X(x) = \Pr(X \le x) &= \Pr(A_1)\Pr(X \le x \mid A_1) + \dots + \Pr(A_n)\Pr(X \le x \mid A_n) + \dots \\ &= \Pr(A_1)F_{X|A_1}(x) + \dots + \Pr(A_n)F_{X|A_n}(x) + \dots \end{aligned}$$

Дифференцируем, получаем:

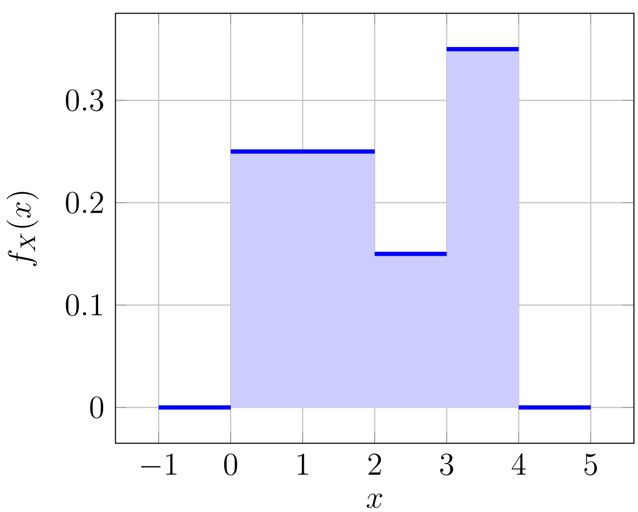
$$f_X(x) = \Pr(A_1)f_{X|A_1}(x) + \dots + \Pr(A_n)f_{X|A_n}(x) + \dots$$

Умножаем на x и интегрируем по всему $\mathbb{R}$:

$$E(X) = \Pr(A_1)E(X \mid A_1) + \dots + \Pr(A_n)E(X \mid A_n) + \dots$$

Пример (который уже был):

$$f_X(x) = \begin{cases} 0.25, x \in [0,2) \\ 0.15, x \in [2,3) \\ 0.35, x \in [3,4] \\ 0, \text{ иначе.} \end{cases}$$



Посчитаем матожидание

$$\begin{aligned} E(X) = &\Pr(X \in [0,2))E(X \mid X \in [0,2)) \\ &+ \Pr(X \in [2,3))E(X \mid X \in [2,3)) \\ &+ \Pr(X \in [3,4))E(X \mid X \in [3,4)) \end{aligned}$$

Заметим, что на каждом отрезке матожидание – это положение центра масс, то есть середина отрезка. Поэтому

$$E(X) = 0.5 \cdot 1 + 0.15 \cdot 2.5 + 0.35 \cdot 3.5 = 2.1$$

6 Смешанные распределения

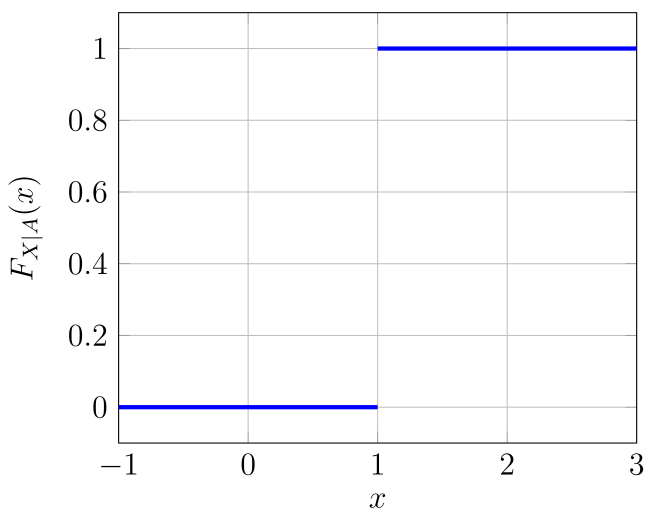
Иногда с.в. могут быть ни дискретными, ни непрерывными, например. Пусть у нас есть слеющий эксперимент. Сначала бросам честную монетку, потом, если выпал орел, то выбираем случайное число из отрезка $[0, 2]$ (равномерно). Случайная величина X при этом равна 1 в случае решки и равна выбранному числу в случае орла.

У данной с.в. нет функции вероятностей, как нет и плотности вероятности. Фукнция распределения все-такие есть. Как ее посчитать? По формуле полной вероятности, которая работает и для функции распределения. Пусть A — событие "выпала решка"

$$F_X(x) = F_{X|A}(x) \Pr(A) + F_{X|\bar{A}}(x) \Pr(\bar{A})$$

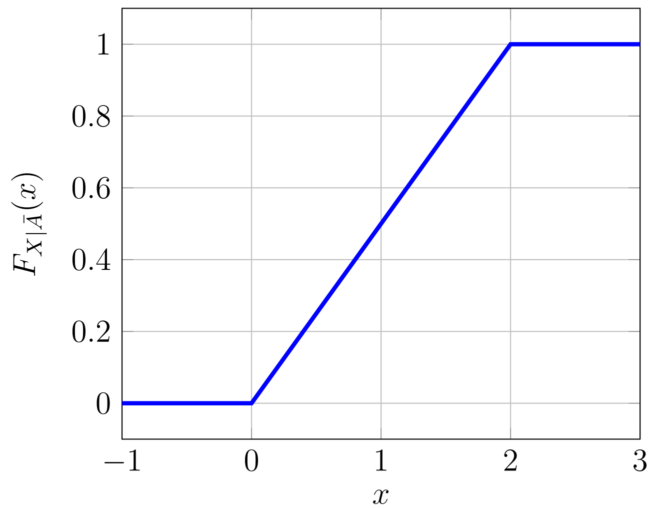
Заметим, что если A, то $X = 1$ с вероятностью 1. То есть,

$$F_{X|A}(x) = \begin{cases} 0, x < 1 \\ 1, x \geq 1 \end{cases}$$



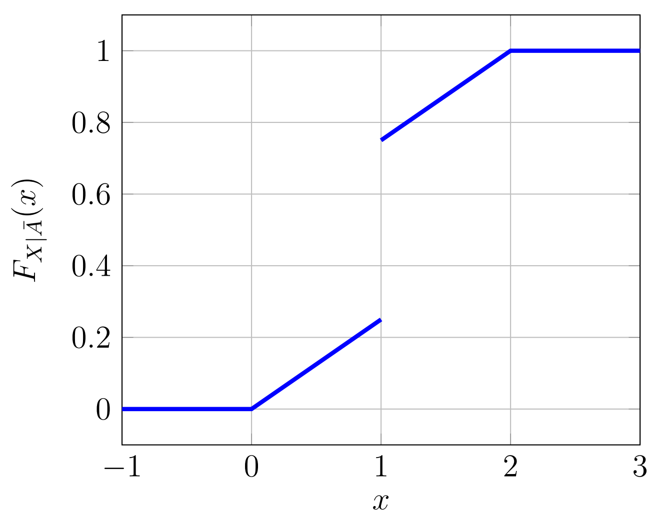
А если $\bar{A}$, то X следует равномерному распределению на отрезке $[0, 2]$.

$$F_{X|\bar{A}}(x) = \begin{cases} 0, x < 0 \\ x/2, x \in [0, 2] \\ 1, x > 2 \end{cases}$$



Итоговая функция распределения выглядит так:

$$F_X(x) = \frac{1}{2}F_{X|A}(x) + \frac{1}{2}F_{X|\bar{A}}(x) = \begin{cases} 0, x < 0 \\ x/4, x \in [0,1) \\ x/4 + 1/2, x \in [1,2] \\ 1, x > 2 \end{cases}$$



7 Векторы из непрерывных с.в.

Пусть у нас есть 2 непрерывных с.в. X и Y. Тогда можно говорить о совместной плотности вероятности:

$$f_{X,Y}(x,y)\delta^2 \approx \Pr(X \in [x, x+\delta] \cap Y \in [y, y+\delta])$$

Более строго: если у нас есть такая функция $f_{X,Y}(x,y)$, такая что для любого измеримого множества $A \subset \mathbb{R}^2$ верно

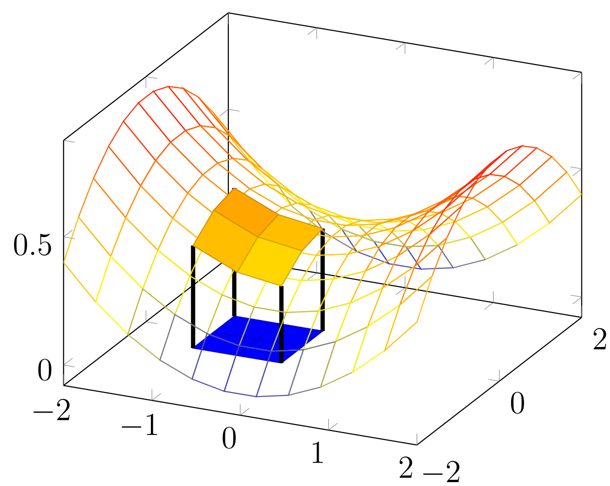
$$\Pr((X,Y) \in A) = \iint\limits_A f_{X,Y}(x,y)dxdy,$$

то с.в. X и Y называются совместно непрерывными, а $f_{X,Y}(x,y)$ называется их совместной плотностью вероятности

Свойства, аналогичные совместной функции вероятности, только суммы заменены интегралами:

- $f_{X,Y}(x,y) \geq 0$
- $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1$

Как визуализировать:



Чтобы посчитать вероятность, что X попадает в какое-то событие, надо просто посчитать объем подграфика на этом событии

NB: Одномерные события имеют вероятность ноль. Например, событие $Y = X$.

8 Маргинальные распределения

Покажем, что

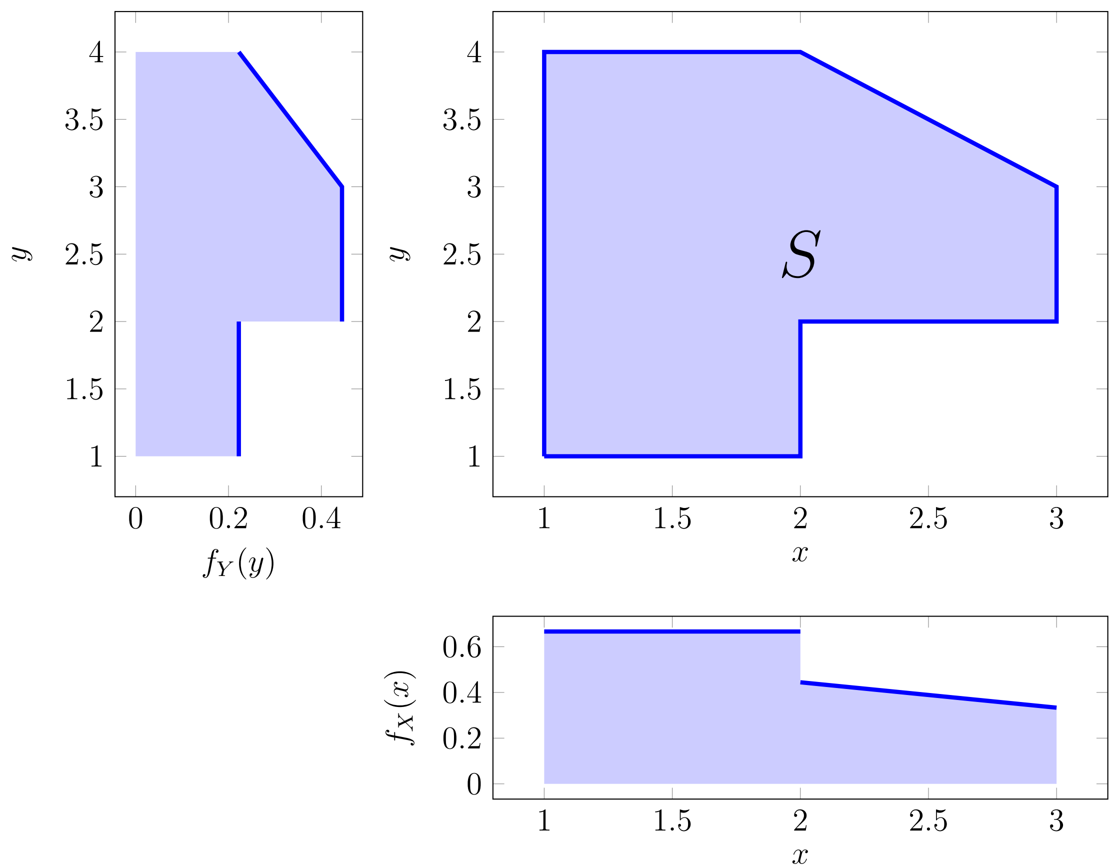
$$\begin{cases} f_X(x) & = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \\ f_Y(y) & = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \end{cases}$$

Для этого рассмотрим функцию распределения X:

$$F_X(x) = \Pr(X \leq X) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f_{X,Y}(t,y)dy \right) dt f_X(x) \quad = F'_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

Пример: равномерное распределение на множестве S площадью 4.5

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{|S|}, & \text{если } (x,y) \in S, \\ 0, & \text{иначе.} \end{cases}$$



9 Более двух с.в. и функции от многих с.в.

Совместное распределение может быть задано на более, чем одной с.в., тогда есть плотность вероятности $f_{X,Y,Z}(x,y,z)$. Все то же самое, что было со многими дискретными с.в., только вместо сумм интегралы:

- $f_{X,Y,Z}(x,y,z) \geq 0$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y,Z}(x,y,z)dxdydz = 1$

И на многих с.в. мы также можем задавать функции, которые будут по сути новыми с.в.: $Z = g(X, Y)$. Их матожидание считается так:

$$E(g(X,Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{X,Y}(x,y)dxdy$$

При этом имеет место быть линейность матожидания:

$$E(\sum_i a_i X_i) = \sum_i a_i E(X_i)$$

10 Совместная функция распределения

Для нескольких с.в. определим функцию распределения:

$$F_{X,Y}(x,y) = \Pr(X \leq x \cap Y \leq Y) = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f_{X,Y}(s,t)dt\right)ds$$

Также можно заметить:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

11 С.в., условные на других с.в.

Было в дискретном случае:

- $p_{X,Y}(x,y)$ — совместная функция вероятности
- $p_{X|A}(x) = \frac{p_X(x)\cdot[x\in A]}{\Pr(A)}$ — условная функция вероятности (где $[\cdot]$ — скобка Айверсона)
- $p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ — функция вероятности X, условная на Y (определена только для таких y, что $p_Y(y) > 0$)

То же самое есть для непрерывного случая:

- $f_{X,Y}(x,y)$ — совместная плотность вероятности
- $f_{X|A}(x) = \frac{f_X(x)\cdot[x\in A]}{\Pr(A)}$ — условная плотность вероятности
- $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ — плотность вероятности X, условная на Y (определена только для таких y, что $p_Y(y) > 0$)

Чуть ближе к формальному определению:

$$\begin{aligned}\Pr\left(X \in [x, x+\delta] \mid Y \in [y, y+\varepsilon]\right) &= \frac{\Pr\left(X \in [x, x+\delta] \cap Y \in [y, y+\varepsilon]\right)}{\Pr(Y \in [y, y+\varepsilon])} \\ &\approx \frac{f_{X,Y}(x,y)\delta\varepsilon}{f_Y(y)\varepsilon} = \frac{f_{X,Y}(x,y)}{f_Y(y)}\delta\end{aligned}$$

И совсем формальное. Если существует такая функция $f_{X|Y}(x \mid y)$, что для всех y и для всех A верно

$$\Pr(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)dx,$$

то эта функция называется условной плотностью вероятности X при условии Y.

Опять при условиях у нас просто появляется новая вероятностная мера. Для всех $y : f_Y(y) > 0$ мы имеем

- $f_{X|Y}(x \mid y) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y)dx = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x|y)dx}{f_Y(y)} = 1$

Из последнего видно, что стоит воспринимать условную плотность как срез совместной плотности по какому-то значению с.в.. Заметьте, что при этом вероятность этого среза равна нулю (одномерное множество), поэтому только терминами условности на события тут не обойтись.

Работает правило умножения:

$$\begin{aligned}f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y \mid x) \\ &= f_Y(y)f_{X|Y}(x \mid y)\end{aligned}$$

А значит, работают полные вероятность и матожидание. Полная вероятность (следует из правила умножения):

$$f_X(x) = \int_{-\infty}^{+\infty} f_Y(y)f_{X|Y}(x,y)dy$$

Также определено условное матожидание

$$E(X \mid Y = y) = \int_{-\infty}^{+\infty} xf_{X|Y}(x,y)dx$$

и работает теорема о полном матожидании (TODO: доказать)

$$E(X) = \int_{-\infty}^{+\infty} f_Y(y)E(X \mid Y = y)dy$$

Также можно посчитать условное матожидание функции с.в.

$$E(g(X) \mid Y = y) = \int_{-\infty}^{+\infty} g(x)f_{X|Y}(x,y)dx$$

12 Независимость с.в.

Было у дискретных:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

Логично определить непрерывные с.в. X и Y независимыми, если

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Это автоматически подразумевает, что для всех y верно, что $f_{X|Y}(x \mid y) = f_X(x)$.
Свойств независимых с.в. те же, что и у дискретных

- $E(XY) = E(X)E(Y)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
- $g(X)$ и $h(Y)$ тоже будут независимы

Пример про зависимые с.в.

Есть палка длиной ℓ. Ломаем ее в случайном месте $X \in [0,\ell]$, остаток тоже ломаем в случайном месте $Y \in [0,X]$. Найти:

- $f_{X,Y}$
- $f_Y(y)$
- $E[Y]$

Лекция 6. Непрерывные с.в., часть 3

17 марта 2021 г.

Закончили на определении плотности вероятности с.в. X при известном значении с.в. Y.

$$\Pr(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)dx,$$

И ее свойствах:

- $f_{X|Y}(x \mid y) \geq 0$
- $\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y)dx = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x|y)dx}{f_Y(y)} = 1$

Еще раз подчеркнем, что стоит воспринимать условную плотность как срез совместной плотности по какому-то значению с.в.. Заметьте, что при этом вероятность этого среза равна нулю (одномерное множество), поэтому только терминами условности на события тут не обойтись.

Работает правило умножения:

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_{Y|X}(y \mid x) \\ &= f_Y(y) f_{X|Y}(x \mid y) \end{aligned}$$

А значит, работают полные вероятность и матожидание. Полная вероятность (следует из правила умножения и определения маргинальной плотности вероятности):

$$f_X(x) = \int_{-\infty}^{+\infty} f_Y(y) f_{X|Y}(x \mid y)dy$$

Также определено условное матожидание

$$E(X \mid Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)dx$$

и работает теорема о полном матожидании:

$$E(X)=\int_{-\infty}^{+\infty} x f_X(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_Y(y) f_{X|Y}(x\mid y)dydx$$
$$=\int_{-\infty}^{+\infty} f_Y(y)\left(\int_{-\infty}^{+\infty} x f_{X|Y}(x\mid y)dx\right)dy$$
$$=\int_{-\infty}^{+\infty} f_Y(y)E(X\mid Y=y)dy$$

Также можно посчитать условное матожидание функции с.в.

$$E(g(X)\mid Y=y)=\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x,y)dx$$

1 Независимость с.в.

Было у дискретных:

$$p_{X,Y}(x,y)=p_X(x)p_Y(y)$$

Логично определить непрерывные с.в. X и Y независимыми, если

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

Это автоматически подразумевает, что для всех y верно, что $f_{X|Y}(x\mid y)=f_X(x)$. Свойства независимых с.в. те же, что и у дискретных

- $E(XY)=E(X)E(Y)$
- $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$
- $g(X)$ и $h(Y)$ тоже будут независимы

Пример независимых с.в.

Пусть есть два трамвая, которые ходят с интервалом 11 и 17 минут, независимо друг от друга. Вы приходите на остановку в случайный момент времени. Сколько ожидаемо вы будете ожидать трамвай, если вам подходит любой из двух?

Определим с.в. Пусть T_1 — время, через которое придет первый трамвай, а T_2 — второй. Заметим, что $T_1\sim U(0,11)$, а $T_2\sim U(0,17)$, то есть

$$F_{T_1}(t)=\begin{cases}0, t<0\\ \frac{t}{11}, t\in[0,11]\\ 1, t>11,\end{cases}$$

$$F_{T_2}(t)=\begin{cases}0, t<0\\ \frac{t}{17}, t\in[0,17]\\ 1, t>17,\end{cases}$$

А время, которое придется провести на остановке есть $Y = \min\{T_1, T_2\}$, функция от двух независимых с.в.

Определим функцию распределения Y.

$$\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) = \Pr(T_1 \leq y \cup T_2 \leq y) \\
&= 1 - \Pr(T_1 > y \cap T_2 > y) \\
&= 1 - \Pr(T_1 > y)\Pr(T_2 > y) = 1 - (1 - F_{T_1}(y))(1 - F_{T_2}(y))
\end{aligned}$$

Воспользуемся полезной формулой с прошлой практики для неотрицательных с.в. (*которую доказали не все*)

$$\begin{aligned}
E(Y) &= \int_0^{+\infty} (1 - F_Y(y))dy = \int_0^{+\infty} (1 - F_{T_1}(y))(1 - F_{T_2}(y))dy \\
&= \int_0^{11} \left(1 - \frac{y}{11}\right)\left(1 - \frac{y}{17}\right) dy \approx 4.31
\end{aligned}$$

Пример про зависимые с.в.

Есть палка длиной ℓ. Ломаем ее в случайном месте $X \in [0, \ell]$, остаток тоже ломаем в случайном месте $Y \in [0, X]$. Какова длина остатка Y? Заметим, что величины зависимы: если $X \geq \ell/2$, то есть шанс, что $Y \geq \ell/2$, а иначе нет.

Совместная функция распределения:

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y \mid x) = \frac{1}{\ell x}$$

Теперь можем найти плотность вероятности Y и его матожидание. Для этого интегрируем совместную ПВ по всем возможным X, то есть от y до ℓ

$$f_Y(y) = \int_y^{\ell} f_{X,Y}(x, y)dx = \int_0^{\ell} \frac{1}{\ell x} dx = \frac{1}{\ell} \ln \frac{\ell}{y}.$$

Два способа посчитать матожидание. Первый:

$$E(Y) = \int_0^{\ell} y f_Y(y)dy = \int_0^{\ell} \frac{y}{\ell} \ln \frac{\ell}{y} dy = \left(\frac{\ln(\ell)}{\ell} \cdot \frac{y^2}{2} - \frac{y^2 \ln(y)}{2\ell} + \frac{y^2}{4\ell}\right)\Bigg|_0^{\ell} = \frac{\ell}{4}$$

Второй:

$$E(Y) = \int_0^{\ell} f_X(x) E(Y \mid X = x)dx = \int_0^{\ell} \frac{1}{\ell} \cdot \frac{x}{2} dx = \frac{x^2}{4\ell}\Bigg|_0^{\ell} = \frac{\ell}{4}$$

Интуитивная логика: первый раз палка ломается в среднем посередине, потом снова в среднем посередине, то есть средняя длина должна быть $\ell/4$. Она работает не всегда, а только для Y, матожидание которых линейно относительно X. Пусть $E(Y \mid X = x) = g(x)$ тогда:

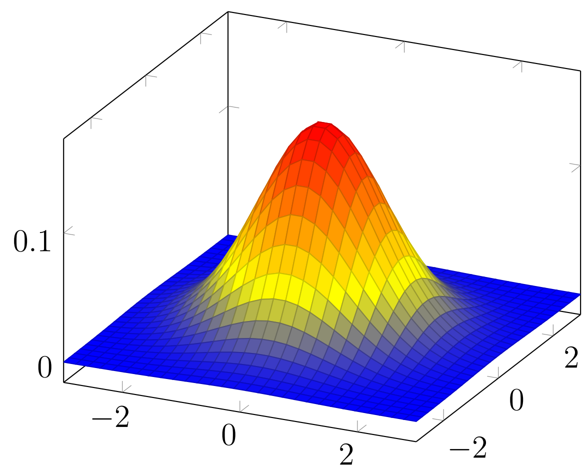
$$E(Y) = \int_{-\infty}^{+\infty} f_X(x) E(Y \mid X = x) dx = \int_{-\infty}^{+\infty} f_X(x) g(x) dx = E(g(X)) \neq g(E(X)).$$

Независимые нормальные распределения

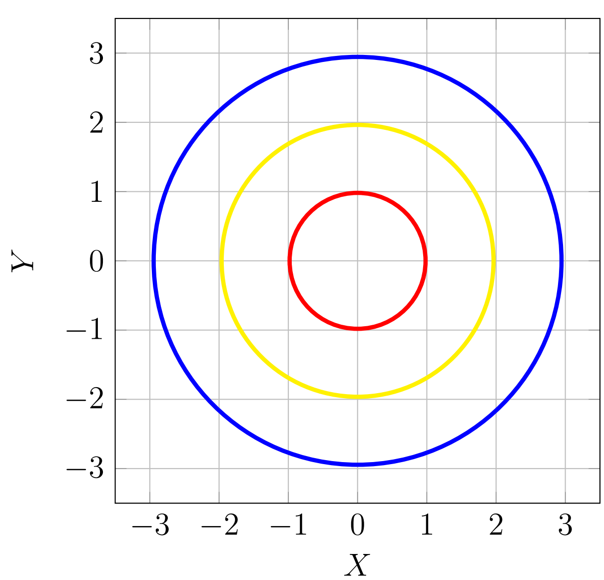
Пусть есть $X \sim N(0,1)$ и $Y \sim N(0,1)$, и они независимы. Тогда

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right).$$

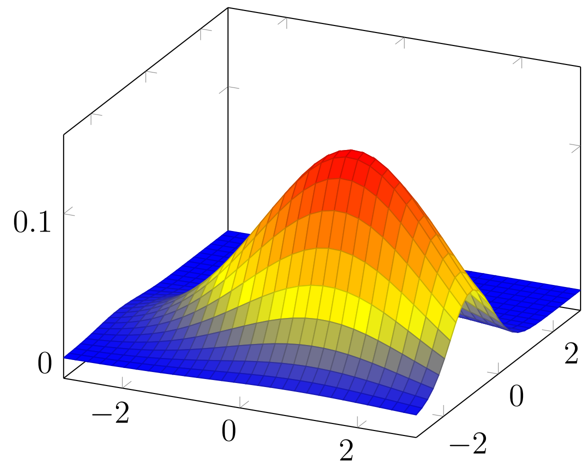
То есть их совместная плотность вероятности пропорциональна $e^{-r^2/2}$, где r — расстояние до точки (0, 0).



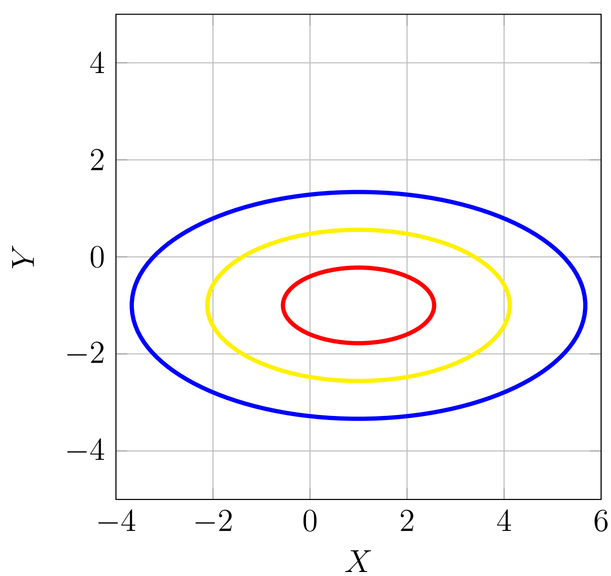
Эквиплотные линии:



Если у нас нестандартные нормальные распределения, то бугорок может сплющиваться или растягиваться вдоль оси, и его центр будет свдигаться



Эквиплотные линии превратятся в эллипсы.



Заметим, что оси эллипса должны быть направлены вдоль осей, иначе с.в. будут зависимые

2 Формула Байеса для случайных величин

Напомним смысл формулы Байеса. С помощью нее мы выражаем вероятность события, которое не можем пронаблюдать (A_i в формуле) через априорные оценки вероятностей другого события(ий) (B в формуле), которое мы можем наблюдать, после его наблюдения.

Априорная оценка вероятности A_i

Априорная оценка условной вероятности наблюдаемого события B

$$\boxed{\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_j \Pr(A_j)\Pr(B \mid A_j)}}$$

Апостериорная оценка вероятности A_i

Полная вероятность B

Легко вывести из правила умножения и формулы полной вероятности для двух дискретных и двух непрерывных случанйых величин. Начнем с дискретных

$$\begin{aligned}
p_{X,Y}(x,y) &= p_Y(y)p_{X|Y}(x \mid y) \\
&= p_X(x)p_{Y|X}(y \mid x) \\
\Rightarrow p_{X|Y}(x \mid y) &= \frac{p_X(x)p_{Y|X}(y \mid x)}{p_Y(y)} \\
&= \frac{p_X(x)p_{Y|X}(y \mid x)}{\sum_x p_X(x)p_{Y|X}(y \mid x)}
\end{aligned}$$

$$\boxed{p_{X|Y}(x \mid y) = \frac{p_X(x)p_{Y|X}(y \mid x)}{\sum_x p_X(x)p_{Y|X}(y \mid x)}}$$

То же самое для непрерывных с.в., но говорим про плотности вероятности, а не про функцию вероятности, и суммы заменяем интегралами

$$\begin{aligned}
f_{X,Y}(x,y) &= f_Y(y)f_{X|Y}(x \mid y) \\
&= f_X(x)f_{Y|X}(y \mid x) \\
\Rightarrow f_{X|Y}(x \mid y) &= \frac{f_X(x)f_{Y|X}(y \mid x)}{f_Y(y)} \\
&= \frac{f_X(x)f_{Y|X}(y \mid x)}{\int_{\mathbb{R}} f_X(x)f_{Y|X}(y \mid x)dx}
\end{aligned}$$

$$\boxed{f_{X|Y}(x \mid y) = \frac{f_X(x)f_{Y|X}(y \mid x)}{\int_{\mathbb{R}} f_X(x)f_{Y|X}(y \mid x)dx}}$$

Но иногда могут быть случаи, когда есть две с.в.: одна дискретная, другая непрерывная. Можем сделать примерно следующее. Пусть X — дискретная, а Y — непрерывная

$$\begin{aligned}
&\Pr(X = x \cap y \leq Y \leq y + \delta) \\
&\qquad = \Pr(X = x)\Pr(y \leq Y \leq Y + \delta \mid X = x) \approx p_X(x) f_{Y|X}(y \mid x)\delta \\
&\qquad = \Pr(y \leq Y \leq y + \delta)\Pr(X = x \mid y \leq Y \leq y + \delta) \approx f_Y(y)\delta p_{X|Y}(x \mid y)
\end{aligned}$$

Откуда следует, что

$$p_X(x) f_{Y|X}(y \mid x) = f_Y(y) p_{X|Y}(x \mid y)$$

Чтобы доказать более строго, надо просто δ устремить к нулю, тогда вместо "$\approx$" будет "$=$". Получаем две формулы.

$$p_{X|Y}(x \mid y) = \frac{p_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

$$f_{Y|X}(y \mid x) = \frac{f_Y(y) p_{X|Y}(x \mid y)}{p_X(x)}$$

И теперь в левую формулу можно подставить формулу полной вероятности, которая работает для $f_Y(y)$:

$$f_Y(y) = \sum_{x'} p_X(x') f_{Y|X}(y \mid x').$$

С правой чуть посложнее, но пока поверьте наслово, что

$$p_X(x) = \int_{\mathbb{R}} f_Y(y') p_{X|Y}(x \mid y') dy.$$

Пример: наблюдаем непрерывную с.в., оцениваем дискретную

$$p_{X|Y}(x \mid y) = \frac{p_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

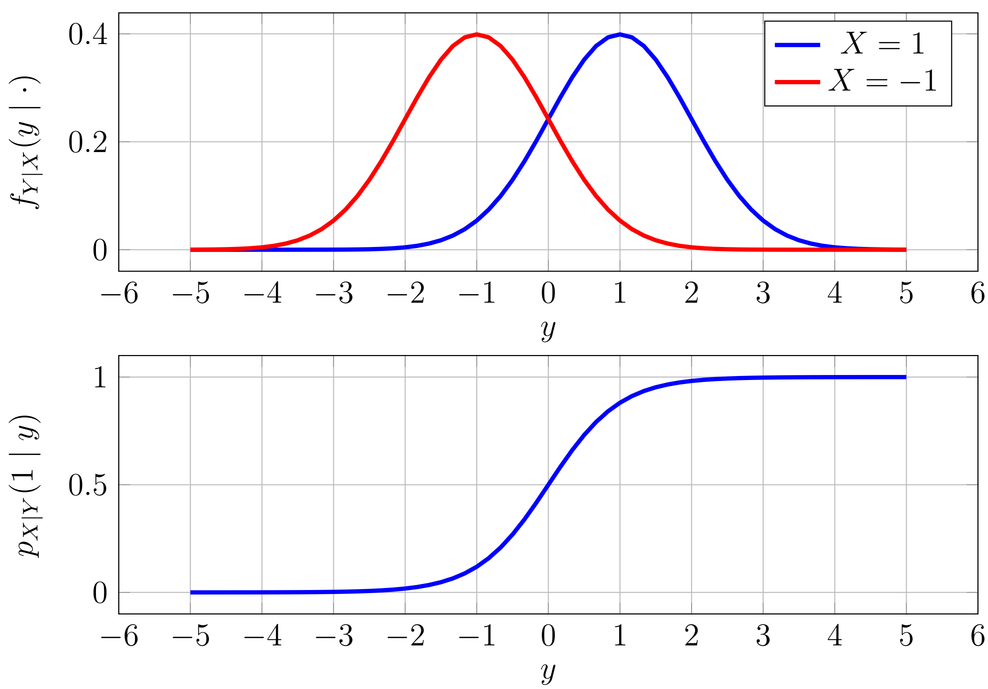
Ситуация: посылаем дискретный сигнал $X \in [-1, 1]$, но к нему добавляется шум $Z \sim N(0, 1)$. В итоге мы можем замерить только $Y = X + Z$. Давайте определим вероятности каждого варианта посланного сигнала, если изначально отправка каждого равновероятна.

- $p_X(-1) = p_X(1) = \frac{1}{2}$

- $Y \sim N(0,1) + X$, то есть если $X = 1$, то $Y \mid X = 1 \sim N(1,1)$. Аналогично $Y \mid X = -1 \sim N(-1,1)$
- Из предыдущего понимаем, что
  - $f_{Y|X}(y,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}}$
  - $f_{Y|X}(y,-1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2}}$
- $f_Y(y) = \frac{1}{2} f_{Y|X}(y,1) + \frac{1}{2} f_{Y|X}(y,-1)$

$$\begin{aligned} p_{X|Y}(1 \mid y) &= \frac{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}}}{\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2}}} \\ &= \frac{1}{1 + e^{-\frac{(y+1)^2 - (y-1)^2}{2}}} = \frac{1}{1 + e^{-2y}} \end{aligned}$$

Иллюстрация: распределение наблюдаемого значения при разных сигналов и вероятность, что посланный сигнал равне единице, в зависимости от наблюдаемого значения



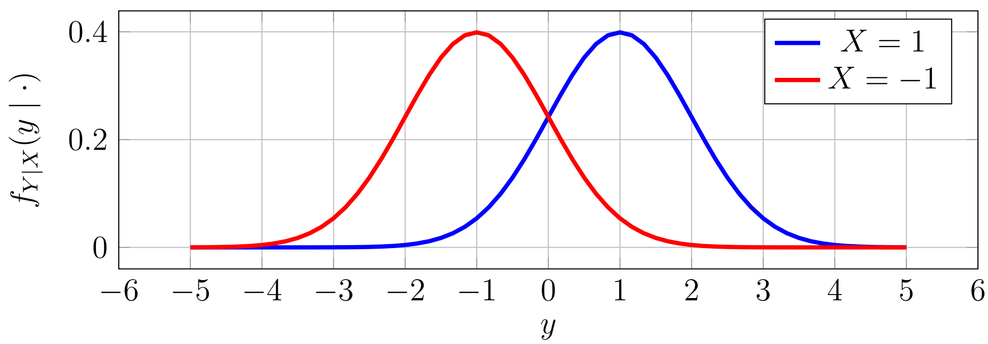
Пример: наблюдаем дискретную с.в., оцениваем непрерывную

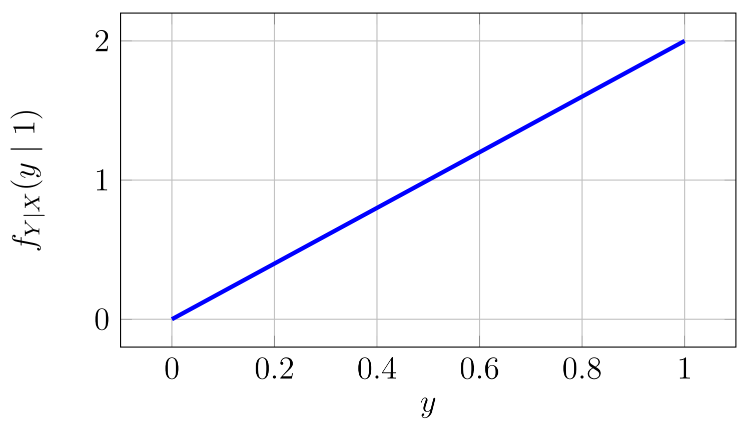
$$\boxed{f_{Y|X}(y \mid x) = \frac{f_Y(y) p_{X|Y}(x \mid y)}{p_X(x)}}$$

Эксперимент: берем нечестную монету, бросаем ее и исходя из результата хотим оценить степень ее нечестности. Наблюдаемая с.в. $X \in \{0, 1\}$ и неизвестная с.в. $Y = \Pr(X = 1) \sim U(0, 1)$.

- $f_Y(y) = 1$ на отрезке $[0, 1]$
- $p_{X|Y}(1 \mid y) = y$
- $p_{X|Y}(0 \mid y) = 1 - y$
- $p_X(1) = \int_0^1 f_Y(y') p_{X|Y}(1 \mid y') dy' = \int_0^1 y' dy = \frac{y^2}{2}\Big|_0^1 = \frac{1}{2}$

$$f_{Y|X}(y \mid 1) = \frac{1 \cdot y}{1/2} = 2y$$





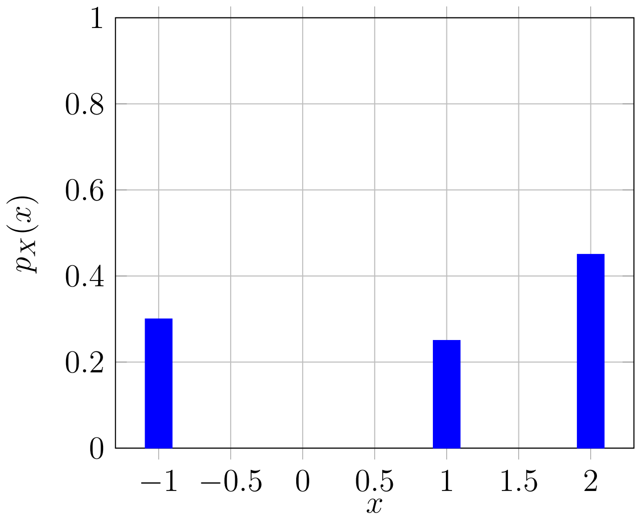
3 Линейные функции от с.в.

Дискретные с.в.

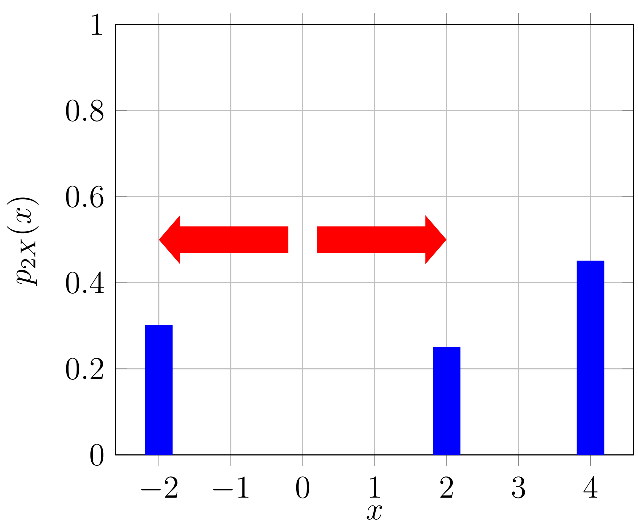
Если $Y = g(X)$ и мы знаем функцию вероятности $p_X(x)$, то не соствит труда посчитать

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x)$$

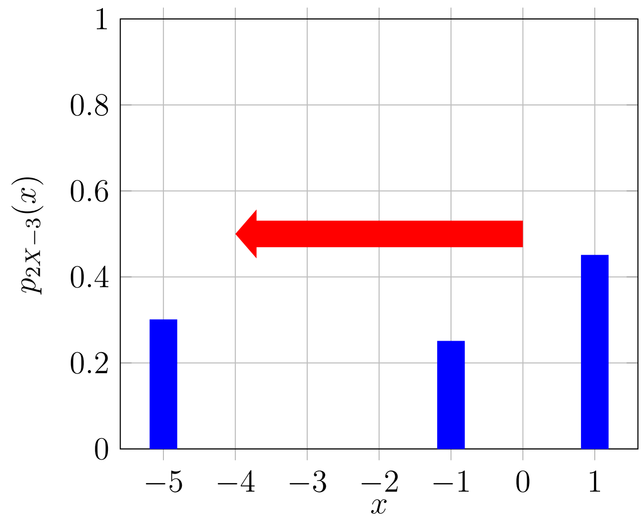
рассмотрим простой случай: $g(x) = ax + b$ — линейная функция. Что происходит с функцией распределения?



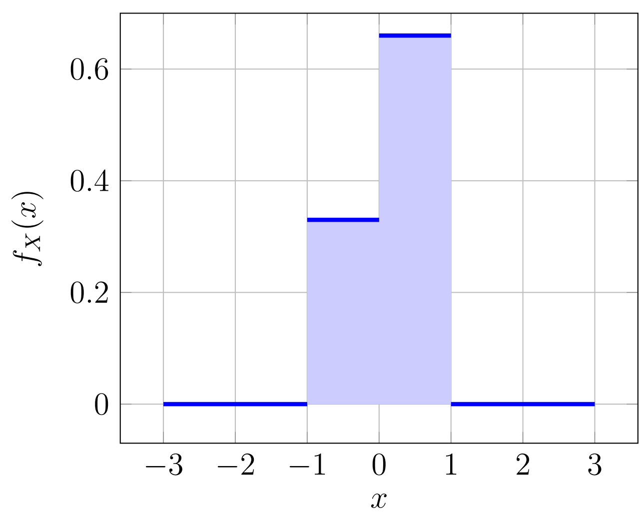
Расмотрим сначала, например, $g(x) = 2x$. Легко понять, что функция вероятностей сохранила свою форму, просто столбики отъехали от оси OY в два раза.



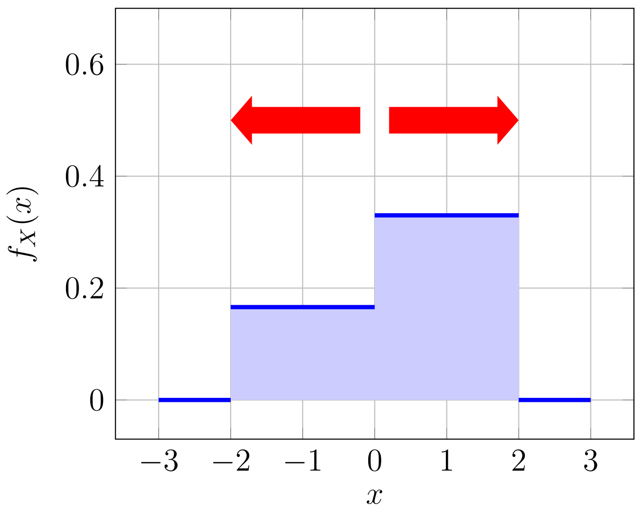
Если мы еще прибавим константу $b = -3$, то просто сдвинем все столбики влево на 3.



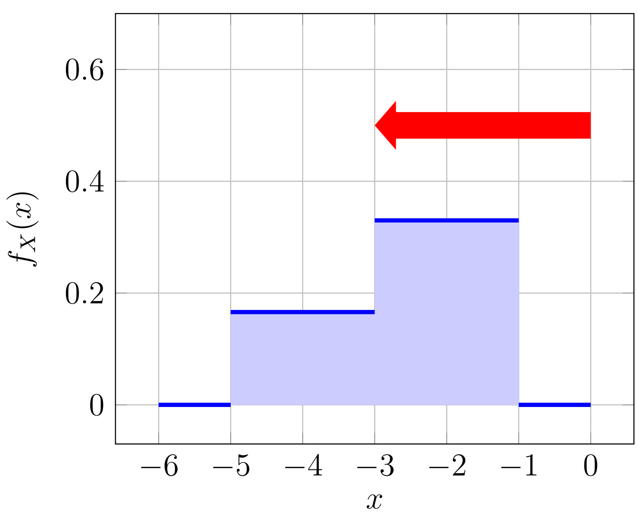
Очень похожая история с непрерывными. Рассмотрим пример.



Для того, чтобы получить функцию плотности вероятности для с.в. $Y = 2x$, надо растянуть ее от оси OY в два раза. Заметьте, что при растягивании сама плотность упадет также в два раза.



Ну и сдвиг при добавлении константы аналогичный. Рассмотрим, например, $g(x) = 2x - 3$



Как это в общем случае делать с.в.? Рассмотрим $Y = aX + b$ (где $a \neq 0$, иначе это скучный случай, когда $Y = b$ с вероятностью 1). Пусть сначала X дискретная, и нам известна ее функция вероятностей. Тогда

$$p_Y(y) = \Pr(Y = y) = \Pr(aX + b = y) = \Pr\left(X = \frac{y-b}{a}\right) = p_X\left(\frac{y-b}{a}\right)$$

С непрерывными такое не работает, так как вероятность, что непрерывная с.в. равна конкретному числу, есть ноль. Но мы можем работать с функциями распределения! Допустим мы знаем $f_X(x)$ и $F_X(x)$. Рассмотрим случай $a > 0$

$$F_Y(y) = \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right),$$
$$f_Y(y) = F'_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

И отдельно $a < 0$

$$F_Y(y) = \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),$$
$$f_Y(y) = F'_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Объединяя два случая, получаем:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Докажем теперь, что линейное преобразование нормального распределения оставляет его нормальным. Пусть $X \sim N(\mu, \sigma^2)$ и $Y = aX + b$. Значит,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{(|a|\sigma)\sqrt{2\pi}} \exp\left(\frac{(y - (b + a\mu))^2}{2(a\sigma)^2}\right), \end{aligned}$$

что есть функция плотности вероятности для $N(a\mu + b, (a\sigma)^2)$.

Лекция 7. Функции от с.в.

22 марта 2021 г.

1 Линейные преобразования

Как это в общем случае посчитать новое распределение линейной функции от с.в.? Рассмотрим $Y = aX + b$ (где $a \neq 0$, иначе это скучный случай, когда $Y = b$ с вероятностью 1). Пусть сначала X дискретная, и нам известна ее функция вероятностей. Тогда

$$p_Y(y) = \Pr(Y = y) = \Pr(aX + b = y) = \Pr\left(X = \frac{y-b}{a}\right) = p_X\left(\frac{y-b}{a}\right)$$

С непрерывными такое не работает, так как вероятность, что непрерывная с.в. равна конкретному числу, есть ноль. Но мы можем работать с функциями распределения! Допустим мы знаем $f_X(x)$ и $F_X(x)$. Рассмотрим случай $a > 0$

$$\begin{aligned}
F_Y(y) &= \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right),\\
f_Y(y) &= F'_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.
\end{aligned}$$

И отдельно $a < 0$

$$\begin{aligned}
F_Y(y) &= \Pr(Y \le y) = \Pr(aX + b \le y) = \Pr\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),\\
f_Y(y) &= F'_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.
\end{aligned}$$

Объединяя два случая, получаем:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Докажем теперь, что линейное преобразование нормального распределения оставляет его нормальным. Пусть $X \sim N(\mu, \sigma^2)$ и $Y = aX + b$. Значит,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{(|a|\sigma)\sqrt{2\pi}} \exp\left(\frac{(y-(b+a\mu))^2}{2(a\sigma)^2}\right), \end{aligned}$$

что есть функция плотности вероятности для $N(a\mu + b, (a\sigma)^2)$.

2 Нелинейные преобразования

Алгоритм поиска нового распределения $Y = g(X)$:

1. Найти функцию распределения $F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y)$
2. Продифференцировать: $f_Y(y) = F'_Y(y)$.

Рассмотрим случай, когда $g(x)$ монотонно возрастает и дифференцируема.

$$F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y) = \Pr(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$
$$f_Y(y) = F'_Y(y) = F'_X(g^{-1}(y)) = f_X(g^{-1}(y))(g^{-1}(y))' = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

Аналогично можно рассмотреть случай, когда $g(x)$ монотонно убывает и дифференцируема

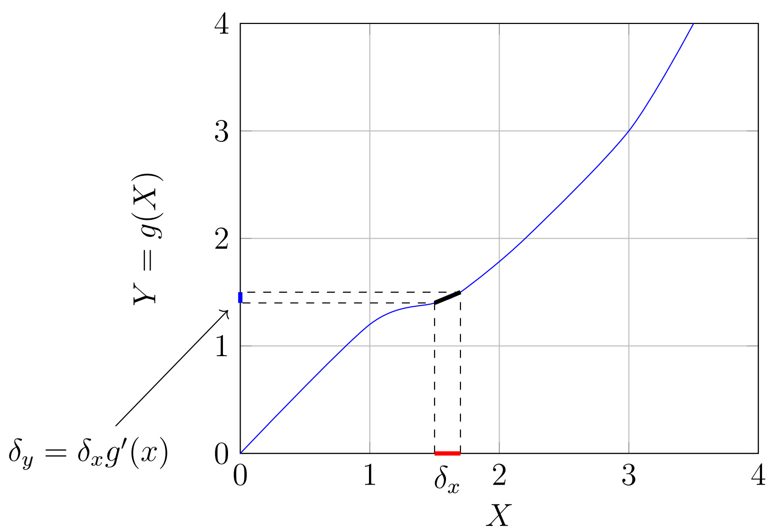
$$F_Y(y) = \Pr(Y \le y) = \Pr(g(X) \le y) = \Pr(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$
$$f_Y(y) = F'_Y(y) = -F'_X(g^{-1}(y)) = -f_X(g^{-1}(y))(g^{-1}(y))' = -\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

То есть в общем случае (когда $g(x)$ строго монотонна):

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

Интуитивное объяснение (которое использовалось на практике). Пусть $y = g(x)$

$$f_Y(y)\delta_y \approx \Pr(y \le Y \le y + \delta_y) = \Pr(x \le X \le x + \delta_y) = f_X(x)\delta_x = f_X(x)\frac{\delta_y}{g'(x)}$$



Для немонотонных $g(x)$ алгоритм остается прежний, но нет гарантий, что все пройдет легко. рассмотрим $g(x) = x^2$ и $Y = g(X)$ (при известной плотности X).

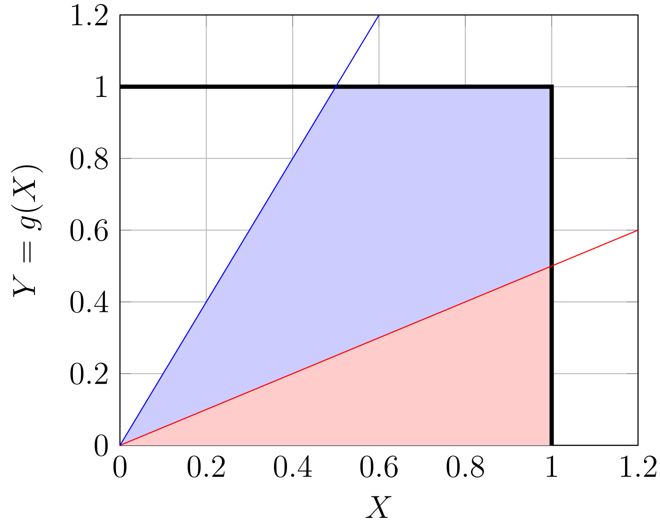
$$\begin{aligned}
F_Y(y) &= \Pr(Y \le y) = \Pr(X^2 \le y) = 1 - \Pr(X^2 > y) \\
&= 1 - \Pr(X > \sqrt{y}) - \Pr(X < -\sqrt{y}) \\
&= \Pr(X \le \sqrt{y}) = \Pr(X < -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-sqrty) \\
f_Y(y) &= F_Y'(y) = F_X'(\sqrt{y}) - F_X'(-\sqrt{y}) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\
&= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}
\end{aligned}$$

3 Функции нескольких с.в.

Если есть $Z = g(X, Y)$, то алгоритм нахождения ее функции распределения или плотности такой же. Рассмотрим на примере. X, Y – независимые, обе следуют $U(0, 1)$. Рассмотрим с.в. $Z = \frac{Y}{X}$.

Найдем функцию распределения $F_Z(z)$

$$F_Z(z) = \Pr(\frac{Y}{X} \le z).$$



Если z меньше единицы, то это просто интеграл по красному треугольнику, то есть его площадь (так как совместная плотность $f_{X,Y}(x, y) = 1$ на всем квадрате).

$$F_Z(z) = \frac{z}{2}, \text{ если } z \in [0, 1].$$

Если $z > 1$, то это площадь всего квадрата минус площадь незакрашенного треугольника

$$F_Z(z) = 1 - \frac{1}{2z}.$$

Объединяя все вместе, получаем:

$$F_Z(z) = \begin{cases} 0, & \text{если } z < 0, \\ \frac{z}{2}, & \text{если } z \in [0, 1), \\ 1 - \frac{1}{2z}, & \text{если } z \geq 1. \end{cases}$$

3.1 Сумма независимых с.в.

В данном подразделе мы имеем две независимых с.в. X и Y, знаем их распределение и хотим узнать распределение с.в. $Z = X + Y$.

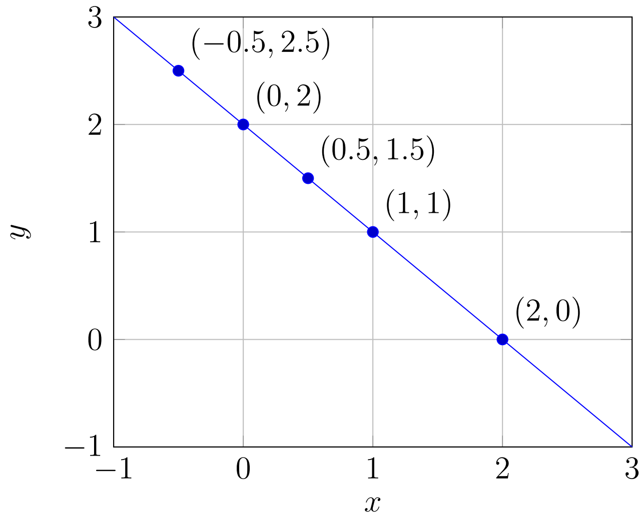
Дискретный случай

Чтобы посчитать вероятность того, что $Z = z$, необходимо, чтобы одновременно выполнялось:

- $X = x$ (где x — какое-то число из множества значений X)

- $Y = z - x$.

То есть мы берем все пары (x, y) с прямой $x+y=z$ и суммируем вероятности этих пар.



Заметим, что из-за независимости X и Y события $X = x$ и $Y = z - x$ независимы, поэтому

$$\Pr(X = x \cap Y = z - x) = p_X(x) p_Y(z - x).$$

Таким образом, формула выглядит так:

$$\boxed{p_Z(z) = \sum_x p_X(x) p_Y(z - x)}$$

Непрерывный случай

Аналогичная формула для непрерывных с.в.:

$$\boxed{f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx}$$

Чтобы ее доказать, обусловимся сначала на событии $X = x$. Тогда

$$f_{Z|X}(z \mid x) = f_{Y+x|X}(z \mid x) = f_{Y+x}(z) = f_Y(z - x).$$

Далее, по формуле полной вероятности имеем

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_{Z|X}(z \mid x) dx = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx/$$

Интересное следствие

Сумма независимых нормальных распределений есть нормальное распределение. Пусть $X \sim N(\mu_X, \sigma_X^2)$ и $Y \sim N(\mu_Y, \sigma_Y^2)$ — независимы. Посчитаем $f_Z(z)$, где $Z = X + Y$.

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}} \exp\left(-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right)
\end{aligned}
$$

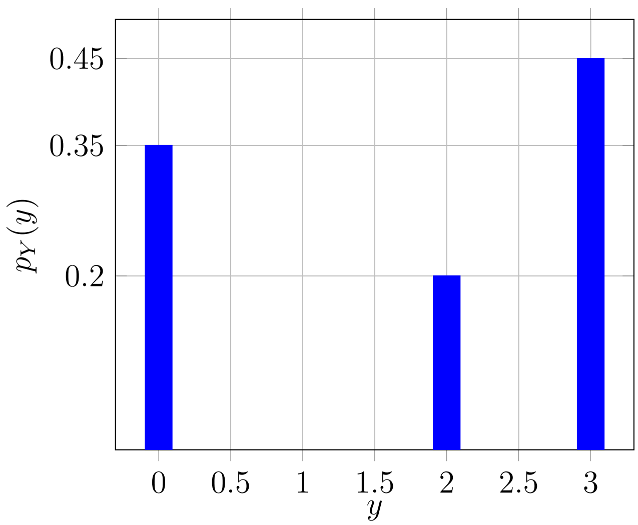
Тут было много несложной, но громоздкой математики, которую мы опустили. Таким образом, $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Единственное удивительное здесь – то, что мы по-прежнему имеем нормальное распределение. Его параметры при этом очевидны из линейности матожиданий и из свойств дисперсии независимых с.в.

По индукции легко доказать, что для любого конечного множества независимых нормальных с.в. их сумма будет также нормальной с.в.

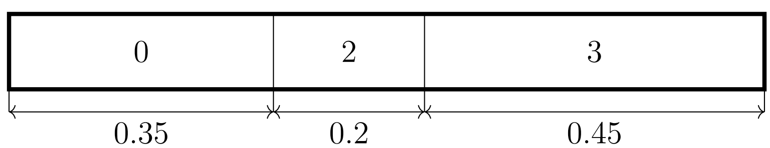
4 Эмитация одного распределения другим

Практически любое современное вычислительное устройство умеет генерировать (псевдо)случайные числа, но чаще всего только равномерное распределение. Этого на самом деле достаточно, чтобы получить случайную величину, следующую любому другому распределению. В данной части мы предполагаем, что мы умеем получать с.в. $X \sim U(0,1)$ и хотим получить с.в. Y, для которой знаем функцию распределения $F_Y(y)$.

Рассмотрим простой случай. Y – дискретная с.в. с конечным набором значений. Пусть ее функция вероятности выглядит так:

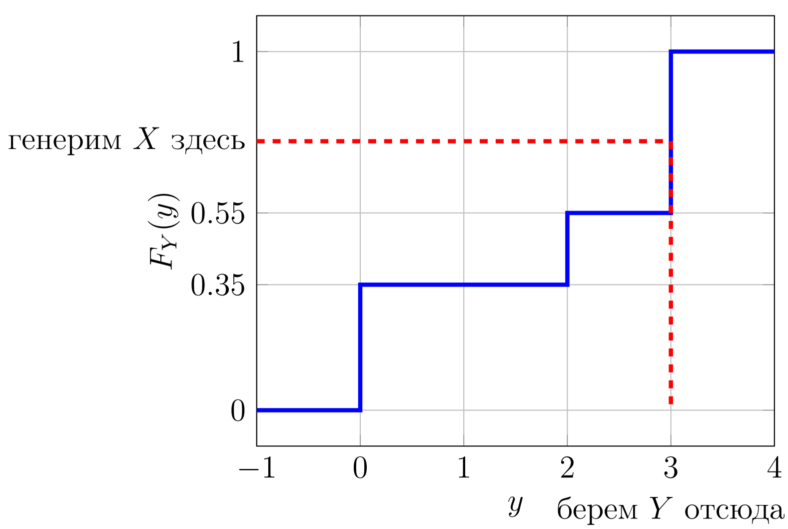


Как мы можем поступить. Возьмем отрезок единичной длины, разобъем его на отрезки длиной, соответствующей вероятности каждого значения с.в., сгенерим X и вернем значение, соответствующее тому значению, в которое попал X .

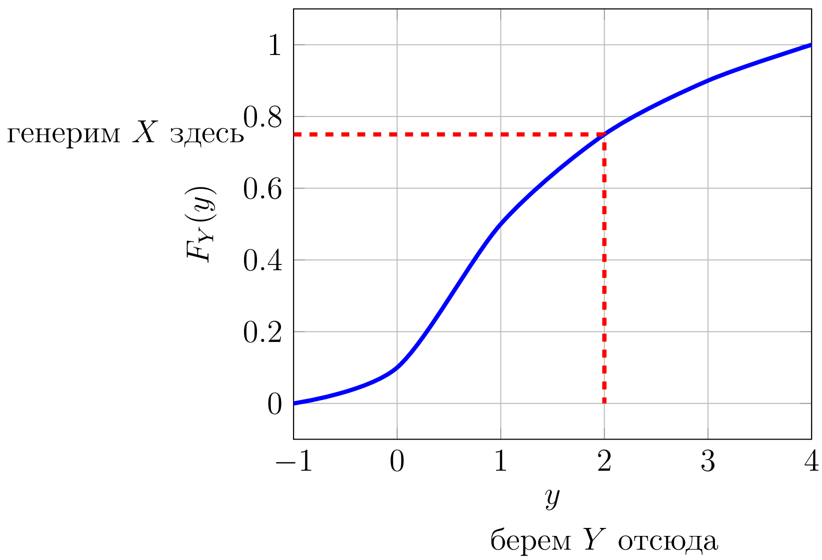


Таким методом можно сэмитировать любую с.в. с конечным множеством значений размера n за время запроса к $X \sim U(0,1)$ плюс время $\Theta(\log(n))$, необходимое для нахождения отрезка, в который мы попали, двоичным поиском.

Интерпретация наших действий на основе функции распределения:



Давайте используем эту интерпретацию для непрерывной Y в случае с монотонной $F_Y(y)$.

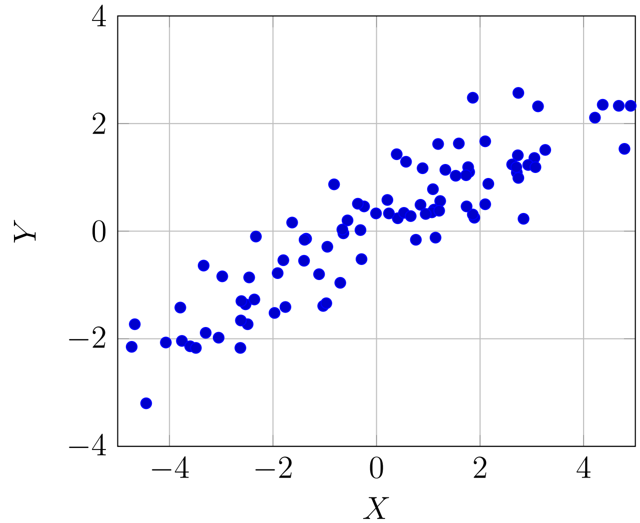


То есть после выбора X мы берем $Y = F_Y^{-1}(X)$. Почему это работает:

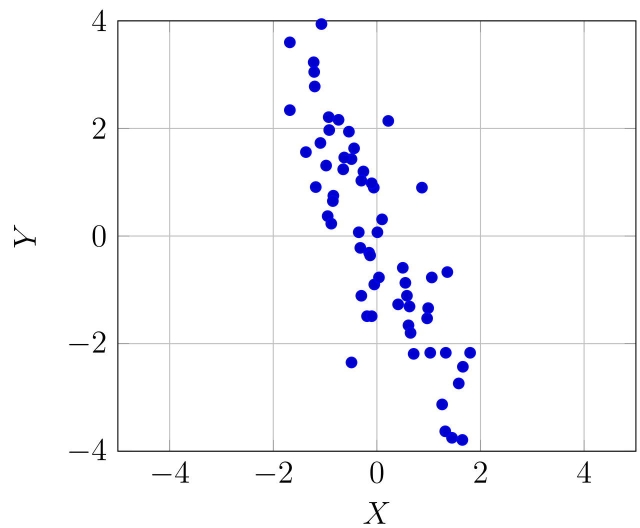
$$\Pr(Y \leq y) = \Pr(F_Y^{-1}(X) \leq y) = \Pr(X \leq F_Y(y)) = F_Y(y)$$

5 Ковариация

До сих пор мы про две с.в. могли сказать только то, зависимы ли они, или нет. Но иногда хочется определить степень этой зависимости, то есть понять, сколько информации одна с.в. дает про другую с.в. Рассмотрим пример. Пусть две с.в. X и Y имеют нулевое матожидание. Если они независимы, то $E[XY] = E[X]E[Y] = 0$. Но рассмотрим такие случаи: пусть каждая точка на графике равновероятна.



В таком случае велика вероятность, что либо X и Y оба меньше нуля, либо оба больше нуля, то есть $E[XY] > 0$. В следующем случае будет наоборот: X и Y с большой вероятностью имеют разные знаки, значит, $E[XY] < 0$.



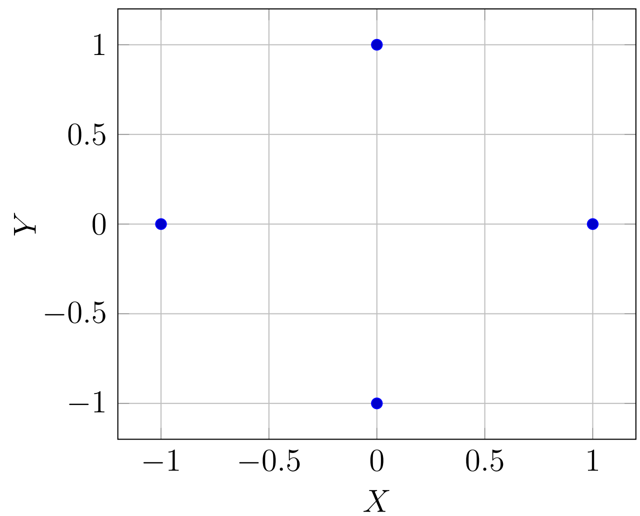
Чтобы измерить степень зависимости была введена величина, называемая *ковариацией* случайных величин X и Y.

$$\mathrm{Cov}(X, Y) = E\Big[(X - E[X])(Y - E[Y])\Big]$$

Если две с.в. независимы, то для них будет верно

$$\mathrm{Cov}(X, Y) = E(X - E(X))E(Y - E(Y)) = 0 \cdot 0 = 0$$

Правда обратное неверно: с.в. с нулевой ковариацией могут быть зависимы. Рассмотрим пример. Пусть четыре пары (X, Y) равновероятны: $(1, 0), (0, 1), (-1, 0), (0, -1)$



Заметим, что матожидания X и Y равны нулю, то есть $\text{Cov}(X, Y) = E[XY]$, но XY всегда равен нулю, поэтому ковариация нулевая. При этом X и Y явно зависимы: если $X = 1$, то Y однозначно равен нулю.

6 Свойства ковариации

Полезная формула.

Ковариация с.в. с самой собой равна ее дисперсии, причем у нас была очень милая формула:

$$\text{Cov}(X, X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Есть ли что-то подобное для ковариации? Конечно, есть.

$$\begin{aligned}\text{Cov}(X, Y) &= E\Big[(X - E[X])(Y - E[Y])\Big] = E\Big[XY - YE[X] - XE[Y] + E[X]E[Y]\Big] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y].\end{aligned}$$

$$\boxed{\text{Cov}(X, Y) = E[XY] - E[X]E[Y]}$$

Линейные преобразования

Линейное преобразование одной с.в.:

$$\begin{aligned}
\operatorname{Cov}(aX+b,Y) &= E[(aX+b)Y] - E[aX+b]E[Y] \\
&= aE[XY] + bE[Y] - aE[X]E[Y] - bE[Y] \\
&= a(E[XY] - E[X]E[Y]) = a\operatorname{Cov}(X,Y)
\end{aligned}$$

Заметим, что прибавление константы к одной с.в. ничего не меняет, так как не меняется распределение $X - E[X]$, а умножение на константу как раз увеличивает этот множитель в ту же константу.

Одна с.в. – сумма двух других:

$$\begin{aligned}
\operatorname{Cov}(X+Y,Z) &= E\Big[(X+Y-E[X+Y])(Z-E[Z])\Big] \\
&= E\Big[(X-E[X])(Z-E[Z])\Big] + E\Big[(Y-E[Y])(Z-E[Z])\Big] \\
&= \operatorname{Cov}(X,Y) + \operatorname{Cov}(Y,Z)
\end{aligned}$$

Ковариация и дисперсия суммы с.в.

Распишем дисперсию суммы двух с.в. по формуле

$$\begin{aligned}
\operatorname{Var}(X+Y) &= E\Big[(X+Y-E[X+Y])^2\Big] \\
&= E\Big[((X-E[X]) + (Y-E[Y]))^2\Big] \\
&= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E\Big[(X-E[X])(Y-E[Y])\Big] \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)
\end{aligned}$$

Что насчет более двух с.в.? Предположим, что у нас есть набор с.в. $X_1, \ldots, X_n$, и матожидание каждой с.в. равно нулю. Тогда имеем

$$\begin{aligned}
\operatorname{Var}(X_1 + \cdots + X_n) &= E((X_1 + \cdots + X_n)^2) \\
&= E\left(\sum_{i=1}^n X_i^2 + \sum_{i\neq j} X_i X_j\right) \\
&= \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i\neq j} \operatorname{Cov}(X_i, X_j)
\end{aligned}$$

Если хотим показать то же самое для с.в. с ненулевым ожиданием, то заметим, что добавление константы к с.в. не меняет ни ее дисперсию, ни ковариацию с другой с.в., поэтому

$$
\begin{aligned}
\operatorname{Var}(X_1+\dots+X_n) &= \operatorname{Var}((X_1-E[X_1])+\dots+(X_n-E[X_n]))\\
&= \sum_{i=1}^{n}\operatorname{Var}(X_i-E[X_i]) + \sum_{i\neq j}\operatorname{Cov}((X_i-E[X_i]),(X_j-E[X_j]))\\
&= \sum_{i=1}^{n}\operatorname{Var}(X_i) + \sum_{i\neq j}\operatorname{Cov}(X_i,X_j)
\end{aligned}
$$

7 Коэффициент корреляции

Ковариация – очень странная величина. Ее размер по сути зависит от разброса двух с.в., котороые являются ее аргументом. То есть по фразе "ковариация X и Y равна единице" очень сложно сказать, много это, или мало, так как мы не знаем масштаба X и Y. Например, когда мы пытаемся посчитать ковариацию температуры воздуха и давления, то в зависимости от используемых величин измерения (цельсий или фаренгейты, паскали или атмосферы) будет разная ковариация. Причем у самой ковариации всегда есть единица измерения, равная единице измерения X, умноженная на Y (например, градус цельсия * паскаль), что еще больше усложняет осознание степени зависимости пары с.в. Поэтому была введена мера зависимости с.в., называемая коэффициентом корреляции $\rho(X, Y)$.

$$\boxed{\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}}$$

NB: он определен только для с.в. с ненулевой дисперсией.

Его основные свойсва:

1. $\rho(X,Y) \in [-1,1]$
2. Если X и Y независимы, то $\rho(X,Y)=0$, как и ковариация (обратное не всегда верно, см. пример для нулевой ковариации)
3. Если $|\rho(X,Y)|=1$, то X и Y линейно зависимы, то есть $(X-E[X]) = c(Y-E[Y])$. Как следствие: $\rho(X,X) = \frac{\operatorname{Cov}(X,X)}{\sigma_X^2} = 1$.
4. Линейные преобразования любой из двух с.в. не меняют модуль коэффициента корреляции: $\rho(aX+b,Y) = \frac{a\operatorname{Cov}(X,Y)}{|a|\sigma_X\sigma_Y} = \operatorname{sign}(a)\rho(X,Y)$

Для доказательства первого и третьего свойств рассмотрим с.в. X и Y с нулевыми матожиданиями и единичной дисперсией (для общего случая легко доказать все, перейдя к с.в. $X' = \frac{X-E[X]}{\sigma_X}$ и $Y' = \frac{Y-E[Y]}{\sigma_Y}$, но мы опустим это для упрощения вычислений). Рассмотрим с.в. $Z = (X - \rho(X,Y)Y)^2$. Она неотрицательна, значит, и ее матожидание неотрицательно.

$$\begin{aligned}E[Z] = E[(X-\rho Y)^2] &= E[X^2] - 2\rho E[XY] + \rho^2 E[Y^2] \\ &= \mathrm{Var}(X) - 2\rho\cdot\rho + \rho^2\,\mathrm{Var}(Y) = 1-\rho^2 \geq 0\end{aligned}$$

Отсюда следует, что во-первых, $\rho^2 \leq 1$, а во-вторых, если $|\rho| = 1$, то $E[Z] = 0$, что возможно только если Z всегда равен нулю, что в свою очередь подразумевает, что X всегда равен $\pm Y$ (в зависимости от знака ρ).

8 Интерпретация корреляции

Если две случайных величины зависимы, то чаще всего это не значит, что одна определяет другую и наоборот. Например, проведя исследование оценок в школе, можно обнаружить, что у учеников с хорошей оценкой по математике также хорошие оценки по литературе, однако это не значит, что знания по математике помогают в изучении литературы и наоборот. Это скорее означает, что у двух этих случайных величин есть какой-то общий фактор, который влияет на них обеих. В случае с оценками это может быть, например, степень усердия, которое тот или иной ученик прилагает к учебе.

Чуть более формально, давайте представим, что есть три случайных величины Z, U и V с нулевыми матожиданиями и единичными дисперсиями, которые все независимы друг от друга. И рассмотрим две других с.в. $X = Z + U$ и $Y = Z + V$. Посчитаем $\rho(X,Y)$

$$\begin{aligned}\mathrm{Var}(X) = \mathrm{Var}(Y) &= \mathrm{Var}(Z) + \mathrm{Var}(U) = 2 \\ \sigma_X = \sigma_Y &= \sqrt(2) \\ E[XY] &= E[Z^2 + UZ + VZ + UV] = \mathrm{Var}(Z) = 1 \\ \rho(X,Y) &= \frac{E[XY]}{\sigma_X\sigma_Y} = \frac{1}{2}\end{aligned}$$

То есть коэффициент корреляции определяет вклад какой-то причины, которая влияет на обе с.в., в каждую с.в.

9 Пример важности корреляции

Допустим, вы решили инвестировать. И у вас есть 10\$, и вы вкладываете по доллару в какие-то 10 компаний. Средняя выручка с каждой компании равна 1\$ и среднеквадратичное отклонение выручки с каждой компании есть 1.3\$. Очевидно, ваша суммарная выручка равна вашему вкладу, но каково среднеквадратичное отклонение? Если выручки от разных компаний независимы, То

$$\operatorname{Var}(X_1 + \cdots + X_10) = \sum_{i=1}^{10} \operatorname{Var}(X_i) = 10 \cdot 1.3^2 = 16.9$$
$$\sigma_{X_1+\cdots+X_10} = \sqrt{16.9} \approx 4.1$$

То есть сумма довольно хорошо сконцентрирована, больше, чем при вкладе всего лишь в одну компанию. На большую прибыль надеяться не стоит, но и много проиграть шансов не так много. Теперь допустим, что у всех выручек есть какой-то общий фактор, который под ними лежит, и коэффициент корелляции между любыми X_i и X_j равен 0.9. Тогда

$$\operatorname{Var}(X_1 + \cdots + X_10) = \sum_{i=1}^{10} \operatorname{Var}(X_i) - \sum_{i \neq j} \operatorname{Cov}(X_i, X_j) = 10 \cdot 1.3^2 + 90 \cdot 0.9 \cdot (1.3)^2 \approx 154$$
$$\sigma_{X_1+\cdots+X_10} = \sqrt{16.9} \approx 12.4$$

То есть если падают акции хотя бы одной компании, то с большой вероятностью падают и акции других, что приводит к большим потерям. Примерно это и случилось в 2008 году, когда рухнул весь рынок сразу.

Лекция 8. Коэффициент корреляции. Условные матожидание и дисперсия как случайные величины

30 марта 2021 г.

1 Коэффициент корреляции

Ковариация – очень странная величина. Ее размер по сути зависит от разброса двух с.в., котороые являются ее аргументом. То есть по фразе “ковариация X и Y равна единице” очень сложно сказать, много это, или мало, так как мы не знаем масштаба X и Y. Например, когда мы пытаемся посчитать ковариацию температуры воздуха и давления, то в зависимости от используемых величин измерения (цельсий или фаренгейты, паскали или атмосферы) будет разная ковариация. Причем у самой ковариации всегда есть единица измерения, равная единице измерения X, умноженная на Y (например, градус цельсия * паскаль), что еще больше усложняет осознание степени зависимости пары с.в. Поэтому была введена мера зависимости с.в., называемая коэффициентом корреляции $\rho(X, Y)$.

$$\boxed{\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}}$$

NB: он определен только для с.в. с ненулевой дисперсией.

Его основные свойсва:

1. $\rho(X, Y) \in [-1, 1]$
2. Если X и Y независимы, то $\rho(X, Y) = 0$, как и ковариация (обратное не всегда верно, см. пример для нулевой ковариации)
3. Если $|\rho(X, Y)| = 1$, то X и Y линейно зависимы, то есть $(X - E[X]) = c(Y - E[Y])$. Как следствие: $\rho(X, X) = \frac{\mathrm{Cov}(X, X)}{\sigma_X^2} = 1$.
4. Линейные преобразования любой из двух с.в. не меняют модуль коэффициента корреляции: $\rho(aX + b, Y) = \frac{a\,\mathrm{Cov}(X, Y)}{|a|\sigma_X \sigma_Y} = \mathrm{sign}(a)\rho(X, Y)$

Для доказательства первого и третьего свойств рассмотрим с.в. X и Y с нулевыми матожиданиями и единичной дисперсией (для общего случая легко доказать все, перейдя к с.в. $X' = \frac{X-E[X]}{\sigma_X}$ и $Y' = \frac{Y-E[Y]}{\sigma_Y}$, но мы опустим это для упрощения вычислений). Рассмотрим с.в. $Z = (X - \rho(X,Y)Y)^2$. Она неотрицательна, значит, и ее матожидание неотрицательно.

$$\begin{aligned} E[Z] = E[(X-\rho Y)^2] &= E[X^2] - 2\rho E[XY] + \rho^2 E[Y^2] \\ &= \operatorname{Var}(X) - 2\rho \cdot \rho + \rho^2 \operatorname{Var}(Y) = 1 - \rho^2 \geq 0 \end{aligned}$$

Отсюда следует, что во-первых, $\rho^2 \leq 1$, а во-вторых, если $|\rho| = 1$, то $E[Z] = 0$, что возможно только если Z всегда равен нулю, что в свою очередь подразумевает, что X всегда равен $\pm Y$ (в зависимости от знака ρ).

2 Интерпретация корреляции

Если две случайных величины зависимы, то чаще всего это не значит, что одна определяет другую и наоборот. Например, проведя исследование оценок в школе, можно обнаружить, что у учеников с хорошей оценкой по математике также хорошие оценки по литературе, однако это не значит, что знания по математике помогают в изучении литературы и наоборот. Это скорее означает, что у двух этих случайных величин есть какой-то общий фактор, который влияет на них обеих. В случае с оценками это может быть, например, степень усердия, которое тот или иной ученик прилагает к учебе.

Чуть более формально, давайте представим, что есть три случайных величины Z, U и V с нулевыми матожиданиями и единичными дисперсиями, которые все независимы друг от друга. И рассмотрим две других с.в. $X = Z + U$ и $Y = Z + V$. Посчитаем $\rho(X,Y)$

$$\begin{aligned} \operatorname{Var}(X) = \operatorname{Var}(Y) &= \operatorname{Var}(Z) + \operatorname{Var}(U) = 2 \\ \sigma_X = \sigma_Y &= \sqrt{(2)} \\ E[XY] &= E[Z^2 + UZ + VZ + UV] = \operatorname{Var}(Z) = 1 \\ \rho(X,Y) &= \frac{E[XY]}{\sigma_X \sigma_Y} = \frac{1}{2} \end{aligned}$$

То есть коэффициент корреляции определяет вклад какой-то причины, которая влияет на обе с.в., в каждую с.в.

3 Пример важности корреляции

Допустим, вы решили инвестировать. И у вас есть 10\$, и вы вкладываете по доллару в какие-то 10 компаний. Средняя выручка с каждой компании равна 1\$ и среднеквадратичное отклонение выручки с каждой компании есть 1.3\$. Очевидно, ваша суммарная выручка равна вашему вкладу, но каково среднеквадратичное отклонение? Если выручки от разных компаний независимы, То

$$\mathrm{Var}(X_1 + \cdots + X_10) = \sum_{i=1}^{10} \mathrm{Var}(X_i) = 10 \cdot 1.3^2 = 16.9$$
$$\sigma_{X_1 + \cdots + X_10} = \sqrt{16.9} \approx 4.1$$

То есть сумма довольно хорошо сконцентрирована, больше, чем при вкладе всего лишь в одну компанию. На большую прибыль надеяться не стоит, но и много проиграть шансов не так много. Теперь допустим, что у всех выручек есть какой-то общий фактор, который под ними лежит, и коэффициент корелляции между любыми X_i и X_j равен 0.9. Тогда

$$\mathrm{Var}(X_1 + \cdots + X_10) = \sum_{i=1}^{10} \mathrm{Var}(X_i) - \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = 10 \cdot 1.3^2 + 90 \cdot 0.9 \cdot (1.3)^2 \approx 154$$
$$\sigma_{X_1 + \cdots + X_10} = \sqrt{16.9} \approx 12.4$$

То есть если падают акции хотя бы одной компании, то с большой вероятностью падают и акции других, что приводит к большим потерям. Примерно это и случилось в 2008 году, когда рухнул весь рынок сразу.

4 Уточнение про типы с.в.

Какие бывают с.в.:

- Сингулярные — сконцентрированные на множестве меры ноль (имеется в виду мера Лебега на $\mathbb{R}$)
- Абсолютно непрерывные — те, у которых есть плотность вероятности
- Смешанные — те, функция распределения которых является взвешенной суммой функций распределений какой-то сингулярной и какой-то непрерывной с.в. То есть для любой с.в. X сущестуют такие с.в. Y и Z (Y — сингулярная, Z — абсолютно непрерывная) и число $p \in [0, 1]$, что

  $$F_X = pF_Y + (1 - p)F_Z,$$

  что является равенством функций, то есть их значения на всех $x \in \mathbb{R}$ совпадают. То есть с вероятностью p с.в. X принимает значение сингулярной с.в., а с вероятностью $(1 - p)$ — абсолютно непрерывной

Почему мы раньше говорили о дискретных а теперь их не упоминаем, но говорим о каких-то сингулярных? Потому что дискретные с.в. (те, которые имеют функцию вероятностей) есть подмножество сингулярных, но не все сингулярные имеют функцию вероятностей. Пример сингулярного недискретного распределения — то, у которого функция распределения есть лестница Кантора — функция, неубывающая и непрерывная на отрезке $[0, 1]$, которая возрастает на нем от 0 до 1, но при этом почти всюду (то есть на всех точках, за исключением множества меры ноль) имеет нулевую производную.

Поэтому уместнее разбить сингулярные с.в. на два типа: дискретные (именно в том понимании, в котором мы имели с ними делао) и сингулярно-непрерывные (те, у которых функция распределения возрастает на множестве меры ноль, но не имеют точек разрыва). Таким образом, для любой с.в. X существуют такие с.в. Y, Z и S (Y — дискретная, Z — абсолютно непрерывная и S — сингулярно-непрерывная) и числа $p \in [0, 1], q \in [0, 1-p]$, что

$$F_X = pF_Y + qF_Z + (1-p-q)F_S.$$

Это доказывается довольно сложно и выходит за рамки нашего курса, но если интересно можете почитать про теорему Лебега о разложении меры (Lebesgue's decomposition).

В рамках нашего курса мы не будем сталкиваться с сингулярно-непрерывными распределениями, поэтому мы будем считать, что любая с.в. может быть представлена в виде комбинации дискретной и непрерывной.

5 Условное матожидание как с.в.

Мы до этого рассматривали функции от с.в. и говорили, что они также являются с.в. Например, если есть функция $g(x) : \mathbb{R} \to \mathbb{R}$, то мы можем рассмотреть композицию этой функции и с.в. X, а именно $g(X) : \Omega \to \mathbb{R} \to \mathbb{R}$.

Вспомним также, что такое условное матожидание. $E[X \mid Y = y]$ – это число, равное матожиданию с.в. X при условии, что с.в. $Y = y$ (*NB:* это условное матожидание может и не существовать или быть бесконечностью). То есть мы можем рассматривать это как функцию от y:

$$g(y) = E[X \mid Y = y]$$

А следовательно, мы можем рассмотреть и композицию этой функции с с.в. Y:

$$g(Y) = E[X \mid Y],$$

что является с.в.

А раз это с.в., то можно говорить о ее матожидании, дисперсии и прочих характеристиках.

6 Матожидание условного матожидания

Мы рассматриваем с.в. $Z = g(Y) = E[X \mid Y]$. Посчитаем ее матожидание в дискретном случае Y.

$$E[E[X \mid Y]] = E[g(Y)] = \sum_y p_Y(y)g(y) = \sum_y p_Y(y)E[X \mid Y = y] = E[X],$$

где последнее равенство — по формуле полной вероятности. То есть мы получили:

$$\boxed{E[E[X \mid Y]] = E[X]}$$

В англоязычной литературе это называется Law of iterated expectations, в русской — просто вариацией теоремы о полном матожидании (что есть правда). Аналогичное равенство можно получить и для непрерывной Y, заменив сумму интеграла, а функцию вероятностей — плотностью.

Вспомним пример про ломание палки из 6-ой лекции (непрерывные с.в, часть 3). Напомним: сначала ломаем палку длины ℓ в случайном месте $Y \sim U(0, \ell)$. Потом берем кусок $[0, Y]$ и опять ломаем его в случайном месте $X \sim U(0, Y)$. И найдем матожидание X следующим образом:

$$E[X] = E[E[X \mid Y]] = E\left[\frac{Y}{2}\right] = \frac{1}{2}E[Y] = \frac{\ell}{4}$$

Рассмотрим еще пример, чтобы лучше понять значение этого утверждения. например, вы сейчас хотите сделать прогноз о том, сколько вы заработаете на акциях в течение мая. Пусть это число будет равно X, и вас интересует $E[X]$. Но к началу мая сумарная стоимость ваших акций Y будет не такой, как сейчас, то есть она тоже будет с.в. И ваш новый прогноз на прибыл будет $E[X \mid Y = y]$. Однако если вернуться в настоящий момент, то для вас прогноз, сделанный в мае также будет являться случайной величиной. При этом $E[X] = E[E[X \mid Y]]$ можно читать так:

> *Матожидание прогноза, сделанного через месяц, равно матожиданию текущего прогноза.*

7 Матожидание условной дисперсии

Для начала ведем понятие условной дисперсии.

$$\mathrm{Var}(X \mid Y = y) = E\left[(X - E[X \mid Y = y])^2 \mid Y = y\right]$$

То есть это вариация, которая посчитана в том мире, где у нас уже произошло событие $Y = y$ и у нас соответсвтующим образом изменилась вероятностная мера для с.в. X. Но это опять функция от Y, поэтому мы можем снова определить с.в. $Z = \mathrm{Var}(X \mid Y)$ как с.в., которая принимает значение $\mathrm{Var}(X \mid Y = y)$, если Y принимает значение y.

Пример с ломанием палки, где $X \sim U(0, Y)$: $\text{Var}(X \mid Y = y) = \frac{y^2}{12}$, при этом $\text{Var}(X \mid Y) = \frac{Y^2}{2}$ — с.в. Но вот полная дисперсия в отличие от полного матожидания устроена сложнее:

$$\boxed{\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y])}$$

Давайте докажем это. Посчитаем сначала условную дисперсию при известном значении Y:

$$\text{Var}(X \mid Y = y) = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2$$

Так как это равенство выполнено для всех возможных значений y, то можем перейти к равенству с.в. (то есть при каждом элементарном исходе они принимают одинаковые значения).

$$\text{Var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

А раз равны с.в., то равны и их матожидания

$$E[\text{Var}(X \mid Y)] = E[X^2] - E[(E[X \mid Y])^2]$$

Но также мы можем посчитать вариацию с.в. $E[X \mid Y]$:

$$\text{Var}(E[X \mid Y]) = E[(E[X \mid Y])^2] - (E[X])^2$$

Остается сложить последние два равенства, тогда справа будет дисперсия X, а слева то, что указано в формуле полной дисперсии.

8 Примеры, если требуются

Первый пример: (X, Y) такие, что:

- с вероятностью $\frac{1}{2}$ с.в. $Y = 1$, а $X \sim U(0, 1)$
- с вероятностью $\frac{1}{2}$ с.в. $Y = 2$, а $X \sim U(1, 3)$

Можем посчитать:

- $E[X \mid Y]$
- $E[X] = E[E[X \mid Y]]$
- $\text{Var}(X \mid Y)$
- $E[\text{Var}(X \mid Y)]$

- $\mathrm{Var}(E[X \mid Y])$

Более глобальный пример: Y — номер группы, известны размеры групп, X — оценка каждого элемента в группе. Известны матожидания и дисперсии X внутри каждой группы.

Можем посчитать:

- $E[X \mid Y]$
- $E[X] = E[E[X \mid Y]]$
- $\mathrm{Var}(E[X \mid Y])$
- $\mathrm{Var}(X \mid Y)$
- $E[\mathrm{Var}(X \mid Y)]$

Можно интерпретировать формулу полной дисперсии как средняя дисперсия внутри групп плюс дисперсия между группами.

9 Матожидание и дисперсия суммы случайного числа с.в.

Рассмотрим ситуацию: вы заходите в N магазинов, где N — некоторая с.в. В i-ом магазине вы тратите X_i рублей, причем

- Все X_i независимы друг от друга и от N (но N может зависеть от $\{X_i\}$)
- Все X_i имеют одинаковое распределение, такое же, как какая-то с.в. X.

Сколько вы ожидаемо потратили? Пусть $Y = \sum_{i=1}^{N} X_i$, тогда

$$E[Y \mid N = n] = nE[X]$$

Это функция, то есть мы можем задать с.в. $Z = E[Y \mid N] = NE[X]$. А тогда мы знаем, что

$$E[Y] = E[E[Y \mid N]] = E[NE[X]] = E[N]E[X].$$

Заметим, что это верно только если матожидания N и X конечны. Это часто называется равенством Вальда, но его мы еще рассмотрим подробнее позже.

$$\boxed{E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X]}$$

Посчитаем также дисперсию.

$$\operatorname{Var}(E[Y \mid N]) = \operatorname{Var}(NE[X]) = (E[X])^2 \operatorname{Var}(N),$$

так как $E[X]$ не зависит от N и считается константой.

$$\operatorname{Var}(Y \mid N = n) = \operatorname{Var}\left(\sum_{i=0}^{N} X_i \mid N = n\right) = \operatorname{Var}\left(\sum_{i=0}^{n} X_i\right) = n \operatorname{Var}(X) \tag{1}$$

$$\operatorname{Var}(Y \mid N) = N \operatorname{Var}(X) \tag{2}$$

$$E[\operatorname{Var}(Y \mid N)] = E[N \operatorname{Var}(X)] = E[N] \operatorname{Var}(X) \tag{3}$$

Теперь сложим $\operatorname{Var}(E[Y \mid N])$ и $E[\operatorname{Var}(Y \mid N)]$, получим дисперсию Y:

$$\boxed{\operatorname{Var}\left(\sum_{i=1}^{N} X_i\right) = (E[X])^2 \operatorname{Var}(N) + E[N] \operatorname{Var}(X)}$$

То есть дисперсия Y есть среднее число слагаемых умножить на дисперсию каждого, плюс она увеличивается дисперсией числа слагаемых, умноженных на квадрат среднего значения каждого слагаемого.

Лекция 9. Полезные инструменты

15 апреля 2021 г.

На этой лекции мы пройдем разные неравенства, которые позволяют нам делать некоторые выводы о с.в., когда информация об этих с.в. ограничена

1 Неравенство Маркова

Самый простой случай — когда мы знаем только матожидание с.в. В этому случае нам может помочь неравенство Маркова. Если с.в. X неотрицательно, тогда для всех $a \in \mathbb{R}^+$ верно, что

$$\boxed{\Pr(X \geq a) \leq \frac{E[X]}{a}}$$

Трактовка: X вряд ли сильно больше своего матожидания. Заметим, что это неравенство несет хоть какую-то смысловую нагрузку только для $a \geq E[X]$.

Докажем это неравенство. Для этого рассмотрим другую с.в. Y, такую, что

$$Y = \begin{cases}0, \text{ если } X < a, a, \text{ иначе.}\end{cases}$$

Заметим, что так как $Y \leq X$, то и $E[Y] \leq E[X]$. Надо ли пояснить?

Поэтому

$$E[X] \geq E[Y] = a \Pr(Y = a) + 0 = a \Pr(X \geq a)$$

Разделим обе части на a, получим неравнество Маркова.

Посмотрим теперь, насколько оно точное. Возьмем с.в. $X \sim \mathrm{Exp}(1)$ и вспомним, что $E[X] = 1$ и $\Pr(X \geq a) = e^{-a}$. Неравенство Маркова дает нам куда более слабую оценку: $\Pr(X \geq a) \leq \frac{1}{a}$.

Посмотрим также, как применять это неравенство на другом примере. Пусть $X \sim U(-4, 4)$. Эта с.в. не неотрицательная, поэтому к ней нельзя применить неравнество Маркова. Однако есть три способа его использовать. Попробуем вычислить вероятность того, что $X \geq 3$. Из простых соображений она равна $\frac{1}{8}$

Первый способ. Возьмем с.в. $Y = |X|$. Несложными вычислениями можно показать, что $Y \sim U(0,4)$. (Для этого рассмотрим $F_Y(y) = F_X(x) - F_X(-x)$ и возьмем производную). Теперь заметим, что событие $Y \geq 3$ есть надсобытие для $X \geq 3$, поэтому

$$\Pr(X \geq 3) \leq \Pr(Y \geq 3) \leq \frac{E[Y]}{3} = \frac{2}{3}.$$

Второй способ. Возьмем точно такой же $Y = |X|$. Заметим, что так как X симметричен относительно нуля, то $\Pr(Y \geq 3) = \Pr(X \leq -3) + \Pr(X \geq 3) = 2\Pr(X \geq 3)$, поэтому оценка уменьшается в два раза.

$$\Pr(X \geq 3) = \frac{\Pr(Y \geq 3)}{2} \leq \frac{E[Y]}{6} = \frac{1}{3}.$$

Третий способ. Снова преобразуем X, но теперь просто сдвинем его в неотрицательную часть. Возьмем $Y = X + 4 \geq 0$. Теперь

$$\Pr(X \geq 3) = \Pr(Y \geq 7) \leq \frac{E[Y]}{7} = \frac{E[X]+4}{7} = \frac{4}{7}.$$

Заметим, что второй способ дал наилучшую оценку, третий — похуже, и первый — самую плохую. Однако все оценки получились весьма неточными. Причина этому — мы используем очень мало информации о с.в., а именно — только знаем про ее матожидание. Если мы имеем только эту информацию, то можно легко построить с.в., для которой неравенство Маркова будет строгим (а именно ту, которую мы использовали в доказательстве неравенства).

2 Неравенства для целочисленных с.в.

Для с.в. X, котороая принимает значения только из $\mathbb{N}$ мы можем использовать следующую полезную формулу:

$$\boxed{E[X] = \sum_{i=1}^{+\infty} \Pr(X \geq i)}$$

Выводится она очень просто:

$$E[X] = \sum_{j=1}^{+\infty} j \Pr(X = j) = \sum_{j=1}^{+\infty} \sum_{i=1}^{j} \Pr(X = j) = \sum_{i=1}^{+\infty} \sum_{j=i}^{+\infty} \Pr(X = j) = \sum_{i=1}^{+\infty} \Pr(X \geq i)$$

При этом для лбой с.в. X, принимающей значения в $(-\infty, 0] \cup \mathbb{N}$ выполняется следующее неравенство:

$$E[X] \le \sum_{i=1}^{+\infty} \Pr(X \ge i).$$

Доказывается по аналогии:

$$E[X] = \sum_{j=-\infty}^{+\infty} j \Pr(X = j) \le \sum_{j=1}^{+\infty} j \Pr(X = j) = \sum_{i=1}^{+\infty} \Pr(X \ge i).$$

С ее помощью можно производить оценки матожидание, когда мы уже имеем какие-то оценки на вероятность того, что с.в. больше или меньше какого-то числа. Это довольно частый случай, так как матожидание с.в. бывает не так легко найти.

Пусть есть какие-то $\alpha, \beta > 0$ и $T \in \mathbb{N}_0$. Пусть также X — какая-то целочисленная с.в.. Тогда верны следующие утверждения.

- Если для всех $\lambda \in \mathbb{N}$ верно $\Pr(X \ge T + \lambda) \le \alpha \exp(-\frac{\lambda}{\beta})$, то $E[X] \le T + \alpha\beta$.
- Если $X \ge 0$ и для всех $\lambda \in [1..T]$ верно $\Pr(X \le T - \lambda) \le \alpha \exp(-\frac{\lambda}{\beta})$, то $E[X] \ge T - \alpha\beta$.
- Если для всех $\varepsilon > 0$ верно $\Pr(X \ge (1+\varepsilon)T) \le \alpha \exp(-\frac{\varepsilon}{\beta})$, то $E[X] \le (1 + \alpha\beta)T$.
- Если для всех $\varepsilon \in (0, 1]$ верно $\Pr(X \le (1-\varepsilon)T) \le \alpha \exp(-\frac{\varepsilon}{\beta})$, то $E[X] \ge (1-\alpha\beta)T$.

Докажем первые два, остальные аналогично. Так как $X \in (-\infty, 0] \cup \mathbb{N}$, то можем применить неравенсвтво:

$$\begin{aligned} E[X] &\le \sum_{i=1}^{+\infty} \Pr(X \ge i) \le T + \sum_{i=T+1}^{+\infty} \alpha \exp\left(-\frac{i-T}{\beta}\right) \\ &= T + \alpha \frac{e^{-1/\beta}}{1 - e^{-1/\beta}} = T + \alpha \frac{1}{e^{1/\beta} - 1} \le T + \alpha\beta, \end{aligned}$$

где последнее неравенство следует из того, что $e^x - 1 \ge x$ для любого x.

Во втором случае $X \in \mathbb{N}$, поэтому

$$\begin{aligned} E[X] &= \sum_{i=1}^{+\infty} \Pr(X \ge i) \ge \sum_{i=1}^{T} \Pr(X \ge i) \\ &= \sum_{i=1}^{T} (1 - \Pr(X \le i - 1)) \ge T - \sum_{i=1}^{T} \alpha \exp\left(-\frac{T - i + 1}{\beta}\right) \\ &= T - \alpha e^{-1/\beta} \frac{1 - e^{-T/\beta}}{1 - e^{-1/\beta}} \ge T - \alpha\beta. \end{aligned}$$

3 Неравенство Чебышева

Допустим, у нас есть еще чуть больше информации про с.в., а именно мы знаем ее матожидание и дисперсию. Рассмотрим вероятность события $(|X - E(X)| \geq a)$. Данное событие эквивалентно событию $((X - E(X))^2 \geq a^2)$?, поэтому их вероятности равны. Заметим также, что $Y = ((X - E(X))^2$ — неотрицательная с.в., то есть к ней можно применить неравнество Маркова.

$$\Pr(|X - E(X)| \geq a) = \Pr((X - E(X))^2 \geq a^2) \leq \frac{E((X - E(X))^2)}{a^2} = \frac{\mathrm{Var}(X)}{a^2} = \frac{\sigma^2}{a^2}.$$

$$\boxed{\Pr(|X - E(X)| \geq a) \leq \frac{\sigma^2}{a^2}}$$

Или, если вместо a поставить $a\sigma$, получим

$$\boxed{\Pr(|X - E(X)| \geq a\sigma) \leq \frac{1}{a^2}}$$

Проверим, что оно нам дает на том же экспоненциальном распределении, на котором мы тестили неравенство Маркова. Пусть $X \sim \mathrm{Exp}(1)$, тогда $E[X] = \mathrm{Var}[X] = 1$. Возьмем какой-нибудь $a > 2$ и посчитаем

$$\Pr(X \geq a) = \Pr((X - 1) \geq (a - 1)) = \Pr(|X - E(X)| \geq a - 1) \leq \frac{1}{(a-1)^2}.$$

Оценка, конечно, получше, но до e^{-a} все равно не дотягивает.

Есть также односторонняя версия неравенства Чебышева, которая называется неравенстовом Кантелли, правда многие приписывают эти неравенства Чебышеву (в том числе Хёфдинг, чье неравенство мы посмотрим следующим). Пусть $a > 0$.

$$\boxed{\Pr(X \geq E(X) + a\sigma) \leq \frac{1}{a^2 + 1}}$$

$$\boxed{\Pr(X \leq E(X) - a\sigma) \leq \frac{1}{a^2 + 1}}$$

Докажем первое из них. Возьмем какоt-нибудь число $u > 0$

$$\begin{aligned}\Pr(X \geq E(X) + a\sigma) &= \Pr(X - E(X) + u \geq a\sigma + u) \\ &\leq \Pr((X - E(X) + u)^2 \geq (a\sigma + u)^2) \\ &\leq \frac{E[(X - E(X) + u)^2]}{(a\sigma + u)^2} = \frac{\sigma^2 + u^2}{(a\sigma + u)^2}.\end{aligned}$$

И это верно для всех u, поэтому можем спокойно взять такой u, при котором получается самая строгая оценка. Возьмем $u = \frac{\sigma}{a}$ (можно получить это значение, продифференцировав по u).

$$\Pr(X \geq E(X) + a\sigma) \leq \frac{\sigma^2 + \frac{\sigma^2}{a^2}}{(a\sigma + \frac{\sigma}{a})^2} = \frac{1 + \frac{1}{a^2}}{\left(a + \frac{1}{a}\right)^2} = \frac{a^2 + 1}{(a^2 + 1)^2} = \frac{1}{a^2 + 1}.$$

4 Неравенство Хёфдинга

Это неравенство для оценки вероятности того, что сумма n независимых ограниченных случайных величин сильно отклоняется от своего матожидания. Выглядит оно так. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, причем каждая из них ограничена. То есть для любого $i \in [1..n]$ есть такие $a_i, b_i \in \mathbb{R}$, что $\Pr(X_i \in [a_i, b_i]) = 1$. Пусть также $X = \sum_{i=1}^{n} X_i$. Тогда

$$\boxed{\Pr(X - E[X] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)}$$

$$\boxed{\Pr(|X - E[X]| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)}$$

Чем полезно это неравенство? Во-первых, для с.в. $X_i \in [0, 1]$ оно выглядит куда милее (так как сумма в знаменателе экспоненты равна n):

$$\Pr(X - E[X] \geq t) \leq e^{2t^2/n}$$
$$\Pr(|X - E[X]| \geq t) \leq 2e^{2t^2/n}$$

Во-вторых, оно позволяет получить хорошую оценку на вероятность того, что с.в., следующая биномиальному распределению сильно отклоняетя от своего матожидания. Пусть $X \sim \mathrm{Bin}(n, p)$. Тогда его можно представить как сумму n независимых с.в., следующих распределению Бернулли, то есть ограниченных нулем и единицей. Раньше мы не могли ничего сделать с такой вероятностью:

$$\Pr(X \geq np + t) = \sum_{i=\lceil np+t \rceil}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

А теперь можем ее ограничить сверху:

$$\Pr(X \geq np + t) \leq e^{-2t^2/n}$$

Аналогично:

$$\Pr(X \leq np - t) = \Pr(n - X \geq n(1-p) + t) \leq e^{-2t^2/n}$$

Но последние два неравенства чаще представляют в мультипликативном виде:

$$\boxed{\begin{aligned} \Pr(X \geq (p+\delta)n) \leq e^{-2\delta^2 n} \\ \Pr(X \geq (p-\delta)n) \leq e^{-2\delta^2 n} \end{aligned}}$$

Пробежимся по доказательству общего неравенства (без модуля). Оно основано на лемме Хёфдинга.

Лемма 1 *Пусть с.в. X такова, что $E[X] = \mu$ и $\Pr(X \in [a,b]) = 1$. Тогда для любого $s \in \mathbb{R}$*

$$E[e^{s(X-\mu)}] \leq \exp\left(\frac{1}{8} s^2 (b-a)^2\right)$$

Ее доказательство основано на применение неравенства Йенсена к экспоненте (пока опустим его).

Вернемся к доказательству неравенства Хёфдинга. Возьмем какой-нибудь $s \in \mathbb{R}$ и применим неравенство Маркова

$$\begin{aligned} \Pr(X - E[X] \geq t) &= \Pr(e^{s(X-E[X])} \geq e^s t) \\ &\leq e^{-st} E[e^{s(X-E[X])}] \\ &= e^{-st} \prod_{i=1}^{n} E[e^{s(X_i - E[X_i])}] \\ &\leq e^{-st} \prod_{i=1}^{n} \exp\left(\frac{1}{8} s^2 (b_i - a_i)^2\right) \\ &= \exp\left(-st + \frac{1}{8} s^2 \sum_{i=1}^{n} (b_i - a_i)^2\right) \end{aligned}$$

Аргумент экспоненты – это квадратичная функция от s, поэтому легко найти ее минимум при $s = \frac{4t}{\sum_{i=1}^{n}(b_i^2 - a_i^2)}$. Подставим и получим

$$\Pr(X - E[X] \geq t) \leq \exp\left(\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

5 Границы Чернова

Самая общая граница Чернова выглядит как просто преобразование события аналогично тому, что мы сделали в доказательстве неравенства Хёфдинга и применении неравенства Маркова. Для всех $t > 0$

$$\Pr(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}}$$

Для всех $t < 0$ аналогично имеем

$$\Pr(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}}$$

Однако гораздо чаще под границами Чернова понимаются следующие "хвостовые оценки"в мультипликативной форме. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, принимающие только значения из $[0, 1]$. Пусть $X = \sum_{i=1}^{n} X_i$ и $\delta > 0$. Тогда

$$\begin{aligned}
\Pr(X \geq (1+\delta)E[X]) &\leq \left(\frac{1}{1+\delta}\right)^{(1+\delta)E[X]} \left(\frac{n - E[X]}{n - (1+\delta)E[X]}\right)^{n-(1+\delta)E[X]} \\
&\leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[X]} = \exp\left(-((1+\delta)\ln(1+\delta) - \delta)E[X]\right) \\
&\leq \exp\left(-\frac{\delta^2 E[X]}{2 + \frac{2}{3}\delta}\right) \\
&\leq \exp\left(-\frac{\min\{\delta, \delta^2\} E[X]}{3}\right).
\end{aligned}$$

Самая строгая из этих границ была показана Хёфдингом, а также он показал, что это лучшее, чего можно добиться с помощью общих границ Чернова для этих с.в. Нижние границы Чернова при всех тех же условиях выглядят очень похоже, только мы предполагаем, что $\delta \in [0, 1]$.

$$
\begin{aligned}
\Pr(X \geq (1-\delta)E[X]) &\leq \left(\frac{1}{1-\delta}\right)^{(1-\delta)E[X]} \left(\frac{n-E[X]}{n-(1-\delta)E[X]}\right)^{n-(1-\delta)E[X]} \\
&\leq \left(\frac{e^{\delta}}{(1-\delta)^{1-\delta}}\right)^{E[X]} \\
&\leq \exp\left(-\frac{\delta^2 E[X]}{2}\right).
\end{aligned}
$$

Вообще главное доказать первую верхнюю оценку, а из нее следуют все остальные простыми махинациями с фукнциями, первая нижняя — тоже из нее путем рассмотрения $Y_i = 1 - X_i$, а остальные нижние так же из нее следуют. Доказывать их несколько муторно, поэтому я просто приложу статью Хёфдинга, где он это делает.

Последняя форма границ Чернова, которую мы рассмотрим — с использованием дисперсии. Пусть у нас есть независимые $X_1, \ldots, X_n$, причем все они не превосходят свое матожидание более, чем на 1 с вероятностью 1 (то есть $\Pr(X_i \leq E[X_i] + 1) = 1$). Пусть X — их сумма, а $\sigma^2 = \mathrm{Var}(X)$. Тогда для любой $\lambda > 0$

$$
\begin{aligned}
\Pr(X \geq E[X] + \lambda) &\leq \left(\left(1+\frac{\lambda}{\sigma^2}\right)^{-(1+\frac{\lambda}{\sigma^2})\frac{\sigma^2}{n+\sigma^2}} \left(1-\frac{\lambda}{n}\right)^{-(1-\frac{\lambda}{n})\frac{n}{n+\sigma^2}}\right)^{n} \\
&\leq \exp\left(-\lambda\left(\left(1+\frac{\sigma^2}{\lambda}\right)\ln\left(1+\frac{\lambda}{\sigma^2}\right)-1\right)\right) \\
&\leq \exp\left(-\frac{\lambda^2}{2\sigma^2+\frac{2}{3}\lambda}\right) \\
&\leq \exp\left(-\frac{1}{3}\min\left\{\frac{\lambda^2}{\delta^2}, \lambda\right\}\right).
\end{aligned}
$$

Лекция 10. Границы Чернова, сходимость и законы больших чисел.

20 апреля 2021 г.

1 Границы Чернова

Следующие неравенства чаще всего имеются в виду под границами Чернова. Это так называемые "хвостовые оценки"на отклонение суммы случайных величин от ее матожидания.

Теорема 1 (Хёфдинг)**.** *Пусть $X_1, \ldots, X_n$ — независимые с.в., причем для всех i верно $\Pr[X_i \in [0,1]] = 1$. Пусть $X = \sum_{i=1}^n X_i$ и $\mu = E[X]$. Тогда для любого $\delta \in (0, \frac{n-\mu}{\mu})$ выполняются следующие неравенства.*

$$
\begin{aligned}
\Pr(X \geq (1+\delta)\mu) &\leq \left(\frac{1}{1+\delta}\right)^{(1+\delta)\mu} \left(\frac{n-\mu}{n-(1+\delta)\mu}\right)^{n-(1+\delta)\mu} \\
&\leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mu} = \exp\left(-((1+\delta)\ln(1+\delta)-\delta)\mu\right) \\
&\leq \exp\left(-\frac{\delta^2\mu}{2+\frac{2}{3}\delta}\right) \\
&\leq \exp\left(-\frac{\min\{\delta,\delta^2\}\mu}{3}\right).
\end{aligned}
$$

Для $\delta = \frac{n-\mu}{\mu}$ левая часть есть $\Pr(X \geq n) = \prod_{i=1}^n \Pr(X_i = 1)$, а для еще больших значений δ левая часть равна нулю.

Последние утверждения для $\delta \geq \frac{n-\mu}{\mu}$ очевидны, поэтому докажем неравенства для $\delta < \frac{n-\mu}{\mu}$ (заметьте, что это подразумевает, что $(1+\delta)\mu < n$)). Сначала докажем первое неравенство, а потом кратко пробежимся по доказательству остальных. Предварительно докажем следующую лемму.

Лемма 1. *Пусть X — с.в., такая, что $\Pr[X \in [0,1]] = 1$. Тогда для любого $t \in \mathbb{R}$*

$$E[e^{tX}] \leq 1 + (e^t - 1)E[X].$$

Доказательство. Так как экспонента — выпуклая вниз функция, то любое для любого $x \in [0,1]$ верно неравенство:

$$e^{tx} \le e^{0t} + \frac{e^{1t} - e^{0t}}{1-0} x = 1 + (e^t - 1)x.$$

Подставим вместо x с.в. X и возьмем матожидание от обеих частей, неравенство останется верным, что и токазывает утверждение Леммы. □

Теперь мы готовы доказать верхнюю границу Чернова.

Доказательство Теоремы 1. Для всех i обозначим $\mu_i = E[X_i]$. Для любого $t > 0$ верно, что

$$\begin{aligned}
\Pr[X \ge (1+\delta)\mu] &\le \frac{E[e^{tX}]}{e^{(1+\delta)\mu t}} \text{ (по обобщенным границам Чернова)} \\
&= e^{-(1+\delta)\mu t} \prod_{i=1}^{n} E[e^{tX_i}] \text{ (по независимости)} \\
&\le e^{-(1+\delta)\mu t} \prod_{i=1}^{n} (1 + (e^t - 1)\mu_i) \text{ (по Лемме 1)} \\
&\le e^{-(1+\delta)\mu t} \left(\frac{n + (e^t - 1)\mu}{n} \right)^n \text{ (среднее геом. } \le \text{ среднее арифм.)} \\
&= e^{-(1+\delta)\mu t} \left(1 + (e^t - 1)\frac{\mu}{n} \right)^n .
\end{aligned}$$

Теперь мы хотим минимизировать это выражение, выбрав наилучший $t \in (0, +\infty)$. Для этого продифференцируем его по t.

$$\begin{aligned}
\left(e^{-(1+\delta)\mu t} \left(1 + (e^t - 1)\frac{\mu}{n} \right)^n \right)' &= -(1+\delta)\mu e^{-(1+\delta)\mu t} \left(1 + (e^t - 1)\frac{\mu}{n} \right)^n \\
&\quad + e^{-(1+\delta)\mu t} n \left(1 + (e^t - 1)\frac{\mu}{n} \right)^{n-1} \frac{\mu}{n} e^t \\
&= \left(-(1+\delta)(\left(1 + (e^t - 1)\frac{\mu}{n} \right)) + e^t \right) \cdot \mu e^{-(1+\delta)\mu t} \left(1 + (e^t - 1)\frac{\mu}{n} \right)^{n-1} .
\end{aligned}$$

Приравнивая это нулю и решая уравнение относительно e^t получаем, что единственный ноль производной в

$$e^{t_0} = \frac{(1+\delta)(n-\mu)}{n - (1+\delta)\mu},$$

причем при меньших значениях t производная отрицательная, а при больших — положительная. Значит, это точка минимума. Подставляя это значение в наше неравенство, получаем

$$\Pr[X \ge (1+\delta)\mu] \le \left(\frac{(1+\delta)(n-\mu)}{n-(1+\delta)\mu}\right)^{-(1+\delta)\mu} \cdot \left(1 + \left(\frac{(1+\delta)(n-\mu)}{n-(1+\delta)\mu} - 1\right)\frac{\mu}{n}\right)^{n}$$
$$= \left(\frac{1}{1+\delta}\right)^{(1+\delta)\mu} \left(\frac{n-\mu}{n-(1+\delta)\mu}\right)^{n-(1+\delta)\mu}.$$

Второе неравенство получается за счет наблюдения

$$\left(\frac{n-\mu}{n-(1+\delta)\mu}\right)^{n-(1+\delta)\mu} = \left(1 + \frac{\delta\mu}{n-(1+\delta)\mu}\right)^{\frac{n-(1+\delta)\mu}{\delta\mu}\cdot\delta\mu} \le e^{\delta\mu}.$$

Следующее равенство получается путем логарифмирования: $x = e^{\ln(x)}$. Следующее неравенство – через неравенство, получаемое аккуратным анализом двух фукнций:

$$(1+\delta)\ln(1+\delta) - \delta \ge \frac{\delta^2}{2 + \frac{2}{3}\delta},$$

и последнее неравенство — аккуратным рассмотрением двух случаев: $\delta \ge 1$ и $\delta < 1$. Мы опускаем все эти подробности, так как к теорверу они не имеют особого отношения, это больше про матан. □

Также есть аналогичные нижние границы Чернова, которые выглядят так (в том же сеттинге, что и верхние, только $\delta \in (0,1)$).

$$\Pr(X \le (1-\delta)\mu) \le \left(\frac{1}{1-\delta}\right)^{(1-\delta)\mu} \left(\frac{n-\mu}{n-(1-\delta)\mu}\right)^{n-(1-\delta)\mu}$$
$$\le \left(\frac{e^{\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$
$$\le \exp\left(-\frac{\delta^2\mu}{2}\right).$$

Первое неравенство доказывается тем, что мы рассматриваем $Y_i = 1 - X_i$ и $Y = n - X$. Тогда

$$\Pr(X \le (1-\delta)\mu) = \Pr((n-X) \ge n-(1-\delta)\mu) = \Pr\left(Y \ge (n-\mu)\left(1 + \delta\frac{\mu}{n-\mu}\right)\right)$$
$$\le \left(\frac{1}{1+\delta\frac{\mu}{n-\mu}}\right)^{(1+\delta\frac{\mu}{n-\mu})(n-\mu)} \left(\frac{n-(n-\mu)}{n-\left(1+\delta\frac{\mu}{n-\mu}\right)(n-\mu)}\right)^{n-(1+\delta\frac{\mu}{n-\mu})(n-\mu)}$$
$$= \left(\frac{n-\mu}{n-(1-\delta)\mu}\right)^{n-(1-\delta)\mu} \left(\frac{1}{1-\delta}\right)^{(1-\delta)\mu}.$$

Последняя форма границ Чернова, которую мы рассмотрим — с использованием дисперсии. Пусть у нас есть независимые $X_1, \ldots, X_n$, причем все они не превосходят свое матожидание более, чем на 1 с вероятностью 1 (то есть $\Pr(X_i \leq E[X_i] + 1) = 1$). Пусть X — их сумма, а $\sigma^2 = \mathrm{Var}(X)$. Тогда для любой $\lambda > 0$

$$\begin{aligned}\Pr(X \geq E[X] + \lambda) &\leq \left(\left(1 + \frac{\lambda}{\sigma^2}\right)^{-(1+\frac{\lambda}{\sigma^2})\frac{\sigma^2}{n+\sigma^2}} \left(1 - \frac{\lambda}{n}\right)^{-(1-\frac{\lambda}{n})\frac{n}{n+\sigma^2}}\right)^n \\ &\leq \exp\left(-\lambda\left(\left(1 + \frac{\sigma^2}{\lambda}\right)\ln\left(1 + \frac{\lambda}{\sigma^2}\right) - 1\right)\right) \\ &\leq \exp\left(-\frac{\lambda^2}{2\sigma^2 + \frac{2}{3}\lambda}\right) \\ &\leq \exp\left(-\frac{1}{3}\min\left\{\frac{\lambda^2}{\delta^2}, \lambda\right\}\right).\end{aligned}$$

2 Еще пара неравенств

Неравенство Беннета

Пусть $X_1, \ldots X_n$ — независимые с.в. с нулевыми матожиданиями ($\forall i \in [1..n] E[X_i] = 0$) и каждый X_i с вероятностью 1 не превосходит какое-то a. Пусть $X = \sum_{i=1}^n X_i$ и $\sigma = \sqrt{\mathrm{Var}(X)}$. Тогда для всех $t > 0$

$$\Pr(X > t) \leq \exp\left(-\frac{\sigma^2}{a^2} h\left(\frac{at}{\sigma^2}\right)\right),$$

где $h(x) = (1 + x)\ln(1 + x) - x$ — возрастающая по x функция.

Неравенство Бернштейна

Те же условия, только требуем $|X_i| \leq a$ с вероятностью 1. Тогда

$$\Pr(X > t) \leq \exp\left(\frac{t^2/2}{\sigma^2 + at/3}\right).$$

Очевидно, оба эти неравенства можно обобщить на X_i с ненулевыми ожиданиями и с разными границами путем линейных преобразований и приведению всех X_i к условиям неравенств. В таком случае оба неравенства Бернштейна сравнимы с неравенством Хёфдинга, но могут давать лучшие границы при высокой концентрации с.в. X_i (когда σ получается меньше, чем сумма длин допустимых интервалов для всех X_i).

3 Сходимость в вероятностных пространствах.

В этом разделе мы будем рассматривать бесконечные последовательности с.в. $\{X_n\}_{n\in\mathbb{N}}$ и определим, что такое сходимость такой последовательности, пытаясь создать аналогию с пределом числовой последовательности.

Сходимость по распределению.

Говорят, что $X_n \to X$ по распределению (все X_n и X – с.в. на одном и том же вероятностном пространстве), если функции распределения $F_{X_i}(x)$ поточечно сходятся к функции распределения $F_X(x)$, кроме, возможно, счетного числа точек разрыва $F_X(x)$. Последнее уточнение важно, так как возьмем, например, последовательность $X_n \sim U(0, \frac{1}{n})$. В пределе, казалось бы, они дадут с.в. $X = 0$. Но все $F_{X_n}(0) = 0$, в то время как $F_X(0) = 1$.

Также заметим, что сходимость по распределению не подразумевает сходимость плотностей вероятности. Например, можно рассмотреть $X_n \in [0, 1]$ с функцией распределения $f_{X_n} = 1 - \cos(2\pi n x)$. Они сходятся по распределению к $U(0, 1)$, но их плотности вообще не сходятся. Но обратное верно: если плотности сходятся почти всюду, то почти всюду сходятся и функции распределения.

Пример: запустили фабрику по выпуску костей $d6$. Но так как оборуждование новое и ненастроенное, кости получаются нечестными. Но со временем, они становятся все более и более честными, и их функция распределения сходится поточечно к функции распределения честной $d6$.

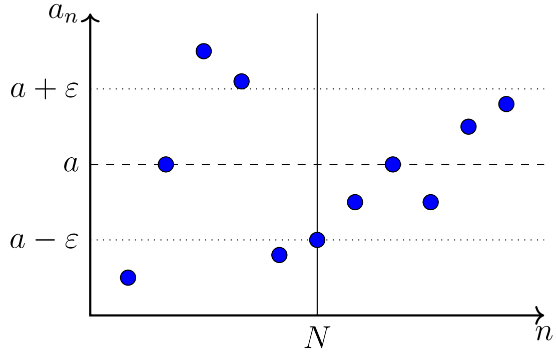
Сходимость по вероятности.

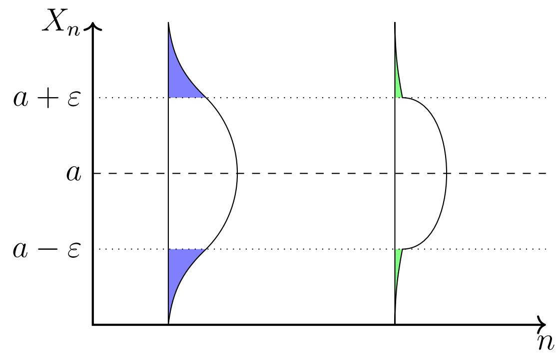
Говорят, что $X_n \to X$ по вероятности (все X_n и X – с.в. на одном и том же вероятностном пространстве), если для любого $\varepsilon > 0$

$$\boxed{\lim_{n\to+\infty} \Pr(|X - X_n| > \varepsilon) = 0}$$

То есть в данном случае это означает, что X_n с росто n становится все более и более зависимой от X, точнее даже почти равной. Особый случай, когда X — константа. Тогда это означает, что последовательность X_n становится все более и более сконцентрированной вокруг этой константы.

Проведем аналогию с сходимостью последовательностей. Для последовательности $a_n \to a$ значит, что для любого чилсла $\varepsilon > 0$ мы можем нарисовать ε-трубку вокруг a, и в нее, начиная с какого-то N попадут все члены последовательности. В случае, когда с.в. $X_n \to a$ по вероятнсоти, это значит, что для любого $\delta > 0$ начиная с какого-то N почти вся вероятностная масса X_n (а именно хотя бы $1 - \delta$) будет сконцентрирована в ε-трубке.





Сходимость по вероятности подразумевает сходимость по распределению (доказать — упражнение).

Важные свойства:

- Если $g(x)$ — непрерывная функция, то $g(X_n)$ также сходятся к $g(X)$ по вероятности.
- Если $X_n \to X$ и $Y_n \to Y$, то $(X_n + Y_n) \to (X + Y)$ (везде сходимость по вероятности).

Пример: вы смотрите на то, какие числа выдает генератор псевдослучайных чисел. Пусть X_n — ваше предположение о том, что он выдаст следующим числом после того, как вы пронаблюдали уже n числе, выданных им. Чем больше n, тем больше вероятность правильно угадать паттерн, по которому действует генератор. Поэтому вероятность, что $X_n = X$ (где X — число, выданное генератором после вашего предположения), стремится к 1.

Еще пример: Пусть X_i – независимые с.в., все $\sim U(0,1)$. Пусть $Y_n = \min_{i\in[1..n]}(X_i)$. Тогда $\Pr(|Y_n - 0| \geq \varepsilon) = \varepsilon^n \to 0$.

NB: сходимость по вероятности не подразумевает сходимость матожиданий. Пусть X_n такие, что

$$\Pr(X_n = 0) = 1 - \frac{1}{n}$$
$$\Pr(X_n = n^2) = \frac{1}{n}$$

Тогда $X_n \to 0$ по вероятности, но $E[X_n] = n^2 \cdot \frac{1}{n} = n$ — расходится.

Сходимость почти наверное. Говорят, что $X_n \to X$ по вероятности (все X_n и X – с.в. на одном и том же вероятностном пространстве), если

$$\boxed{\Pr(\lim_{n\to+\infty} X_n = X) = 1}$$

Это означает, что при всех возможных элементарных исходах из Ω (кроме, возможно, какого-то множество исходов с суммарной вероятностью ноль) X_n образуют такую последовательность, которая сходится к значению X на этом же элементарном исходе:

$$\Pr\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

Это наиболее сильный вид сходимости, который влечет за собой сходимость по вероятности (и следовательно, сходимость по распределению). Доказать это — тоже упражнение.

Пример: Пусть продолжительность жизни какого-то животного — случайная величина. Пусть X_i — количество еды, которое съело это животное в i-й день своей жизни. При любом исходе с какого-то момента X_i станет нулем, то есть последовательность сходится к нулю почти наверное.

4 Законы больших чисел

Слабый закон больших чисел.

Пусть X_i — независимые одинаково распределенные с.в. с матожиданием μ и дисперсией σ^2. Пусть $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ — среднее первых n случайных величин. Тогда

- $E[M_n] = E[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n} \cdot n\mu = \mu$
- $\mathrm{Var}(M_n) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$

То есть с ростом n среднее значение M_n сохраняется, а вот дисперсия уменьшается (то есть увеличивается концентрация вокруг матожидания). Поэтому давайте применим неравенство Чебышева:

$$\Pr(|M_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

То есть M_n сходится по вероятности к μ, что и называется слабым законом больших чисел.

$$\forall \varepsilon > 0 \text{ выполняется } \lim_{n\to+\infty} \Pr(|M_n - \mu| > \varepsilon) = 0$$

Возможные применения.

- Для точного измерения при наличии шумов. Пусть хотим узнать какую-то величину μ, но при ее измерении у нас всегда есть какая-то случайная погрешность W (причем $E(W_i) = 0$). Тогда чем больше измерений мы проведем, тем меньше будет вероятность, что среднее наших измерений сильно отклоняется от реального значения.

- Среднее по экспериментам. Пусть есть нечестная монета с вероятностью p, что выпадет орел. Тогда чем больше экспериментов мы проведем, тем лучше сможем определить этот p.

Сильный закон больших чисел (закон Колмогорова).
В тех же условиях, что и для слабого закона, верно и более сильное утверждение:

$$\Pr(\lim_{n \to +\infty} M_n = \mu) = 1$$

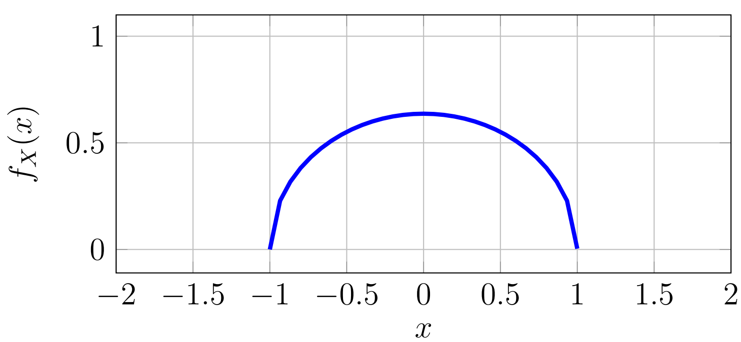
То есть M_n сходится к μ почти наверное. Также существует версия и для случая, когда X_i имеют разные распределения (но с конечными матожиданиями и дисперсиями такими, что $\sum_{i=1}^{+\infty} \frac{1}{i^2} \operatorname{Var}(X_i)$ — сходится). В таком случае $(M_n - E[M_n])$ сходится к нулю почти наверное.

Лекция 11. Центральная предельная теорема, равенство Вальда.

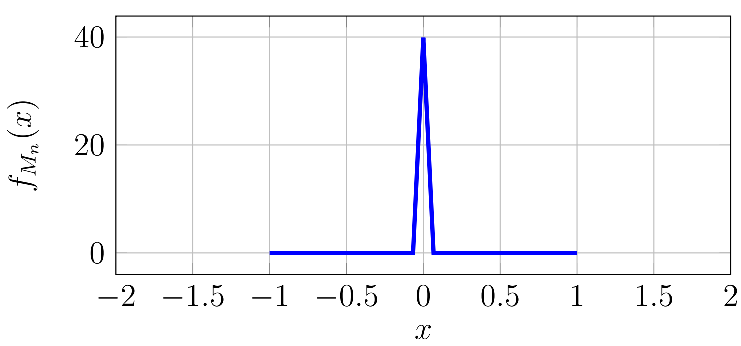
28 апреля 2021 г.

1 Центральная предельная теорема

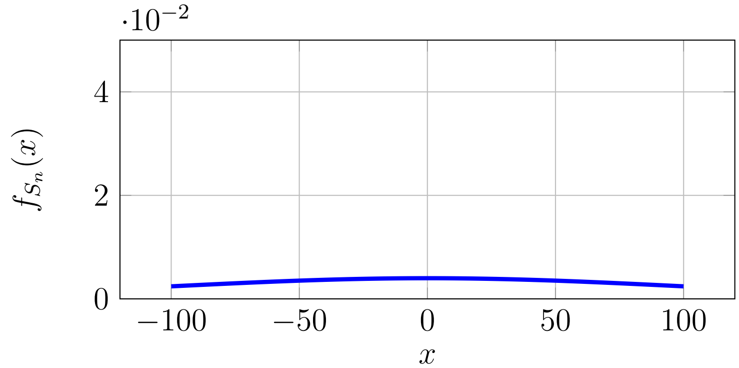
ЗБЧ давал нам только понять, что среднее первых n с.в. из какой-то последовательности сходится к своему матожиданию (во всех смыслах сходимости). Однако, он ничего не говорил про то, какое распределение у суммы с.в. Рассмотрим пример. Пусть есть последовательность одинаково распределенных, независимых с.в. X_n, имеющих плотность вероятности $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$. У всех них матожидание $\mu = 0$ и дисперсия $\sigma^2 = \frac{1}{4}$ (легко посчитать).



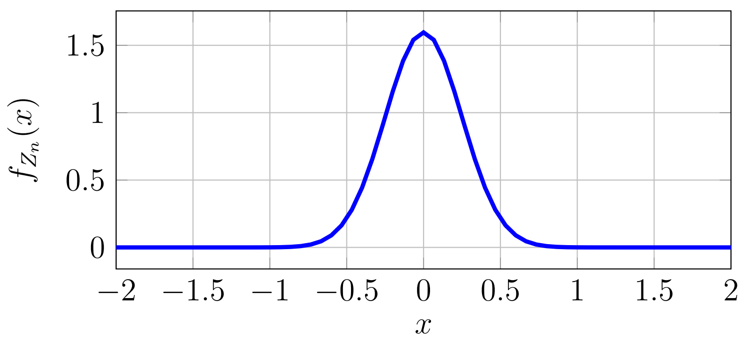
Пусть $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ — среднее первых n с.в. Из ЗБЧ мы знаем, что она с большой вероятностью несильно отклоняется от 1, то есть имеет примерно такую плотность вероятности:



Можно легко посчитать матожиание $E[M_n] = 0$ и дисперсию $\mathrm{Var}(M_n) = \frac{\sigma^2}{n}$. Однако это мало о чем нам говорит, если мы хотим узнать что-то о распределении суммы этих величин. Пусть $S_n = \sum_{i=1}^{n} X_i$. Тогда $E[S_n] = 0$, а дисперсия очень большая, $\mathrm{Var}(S_n) = n\sigma^n$.



Однако что если мы рассмотрим другую нормировку, а именно $Z_n = \frac{S_n}{\sqrt{n}}$? Тогда дисперсия Z_n будет такой же, как у любого X_i, а именно σ^2.



И тут можно заметить, что распределение Z_n подозрительно похоже на нормальное. В этом на самом деле и заключается центральная предельная теорема в ее классическом виде.

Теорема 1 (Центральная предельная теорема). *Пусть $\{X_n\}_{n\in\mathbb{N}}$ — последовательность независимых одинаково распределенных с.в. с конечными матожиданиями μ и дисперсиями σ^2. Пусть $S_n = \sum_{i=1}^{n} X_i$ и $Z_n = \frac{S_n - \mu n}{\sqrt{n}\sigma}$. Пусть также Z — это с.в., следующая стандартному нормальному распределению $N(0,1)$. Тогда Z_n сходится к Z по распределению.*

$$\forall z \lim_{n\to\infty} \Pr(Z_n \le z) = \Pr(Z \le z)$$

Часто в эту классическую формулировку включают и то, что сходимость эта — равномерная.

Мы опустим доказательство, так как оно основано на характеристических функциях и теореме Леви о непрерывности, а это немного выходит за рамки нашего курса.

Существует несколько других вариантов ЦПТ, которые накладывают меньше требований на последовательность X_n.

1.1 Теоремы, не требующие одинакового распределения

Теорема 2 (Теорема Ляпунова). *Пусть $\{X_n\}_{n\in\mathbb{N}}$ — последовательность независимых с.в. с конечными матожиданиями и дисперсиями. Причем $E[X_n] = \mu_n$ и $\mathrm{Var}(X_n) = \sigma_n^2$. Пусть $S_n = \sum_{i=1}^n X_i$ и $s_n^2 = \mathrm{Var}(S_n) = \sum_{i=1}^n \sigma_i^2$. Пусть также для какого-то $\delta > 0$ выполнено условие Ляпунова:*

$$\lim_{n\to+\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\left[|X_i - \mu_i|^{2+\delta}\right] = 0.$$

Тогда

$$\frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \to N(0,1),$$

где сходимость — по распределению.

Теорема 3 (Теорема Линдеберга). *Теорема Ляпунова также верна, если вместо условия Ляпунова для любого $\varepsilon > 0$ выполнено*

$$\lim_{n\to+\infty} \frac{1}{s_n^2} \sum_{i=1}^n E\left[(X_i - \mu_i)^2 \mathbb{1}_{\{|X_i-\mu_i|>\varepsilon s_n\}}\right] = 0.$$

Существуют также версии ЦПТ для последовательностей зависимых с.в. Например, когда будем проходить мартингалы – пройдем ЦПТ для них.

1.2 Практическое значение ЦПТ

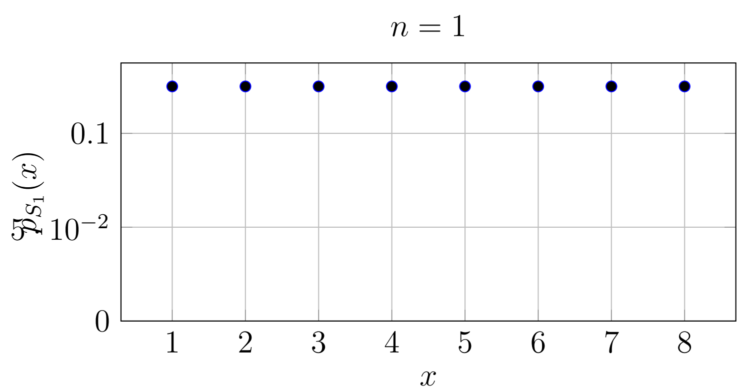
ЦПТ позволяет обращаться с Z_n как со с.в., следующей стандартному нормальному распределению. Так как $S_n = \sqrt{n}\sigma Z_n + n\mu$, мы с ней можем обращаться как со с.в., следующей $N(n\mu, n\sigma^2)$ (но с меньшей точностью). Кстати про точность. Вообще ЦПТ ничего не говорит про скорость поточечной сходимости, но на практике обычно даже для небольших n уже дает довольно хорошие приближения. На этот счет есть теорема Берри-Эссена (для случая независимых, одинаково распределенных с.в.), которая гласит, что если есть конечный тертий момент у всех с.в. последовательности (то есть $E[|X_n|^3] < \infty$), то сходимость к нормальному распределению имеет порядок как минимум $O(1/\sqrt{n})$. Более формально, она звучит так.

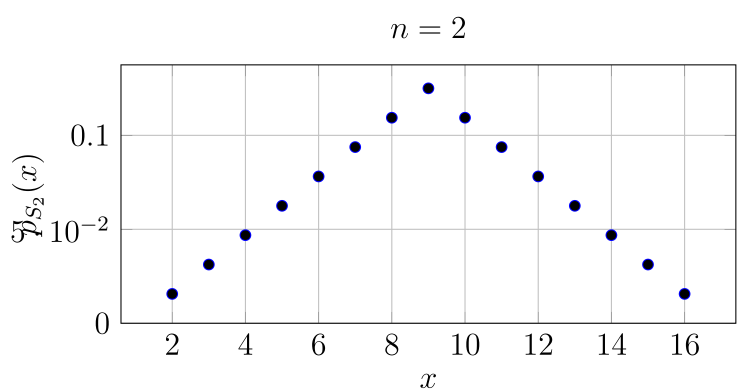
Теорема 4 (Теорема Берри-Эссена)**.** *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ *— последовательность независимых одинаково распределенных с.в. с нулевыми матожиданиями, конечными дисперсиями* σ^2 *и конечными третими моментами* $\rho = E[|X_n|^3]$*. Пусть* $S_n = \sum_{i=1}^{n} X_i$ *и* $Z_n = \frac{S_n}{\sqrt{n}\sigma}$*. Пусть также* $F_n(x)$ *— функция распределения* Z_n*, а* $\Phi(x)$ *— функция распределения стандартного нормального распределения* $N(0,1)$*. Тогда существует такая константа* C*, такая что для всех* n *и для всех* x *верно, что*

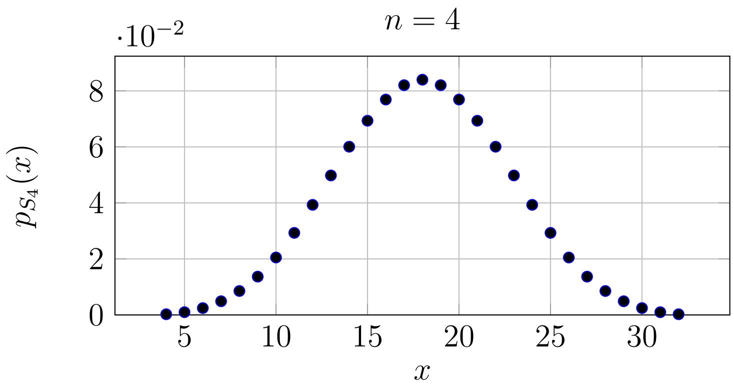
$$|F_n(x) - \Phi(x)| \le \frac{C\rho}{\sigma^3\sqrt{n}}.$$

Из не очень проверенных мной источников известно, что C может быть не очень большой, чуть больше 0.33. Это все означает, что ЦПТ дает уже довольно хорошее приближение даже при не очень больших n. Рассмотрим примеры для некоторых дискретных с.в.

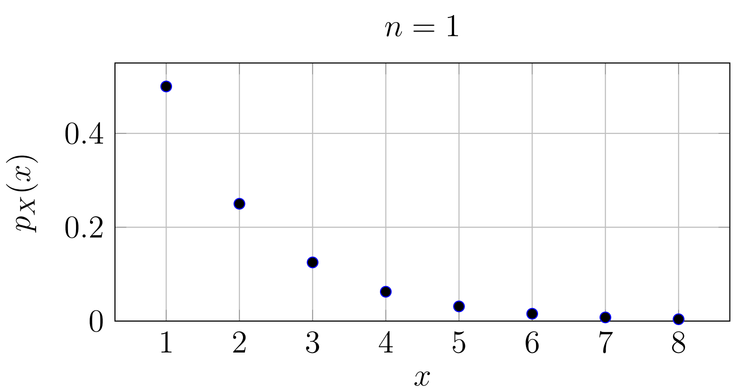
Пусть все $X_n \sim U(1,8)$ (дисркетное равномерно распределение). Посмотрим на графики функции вероятности S_n для разных n

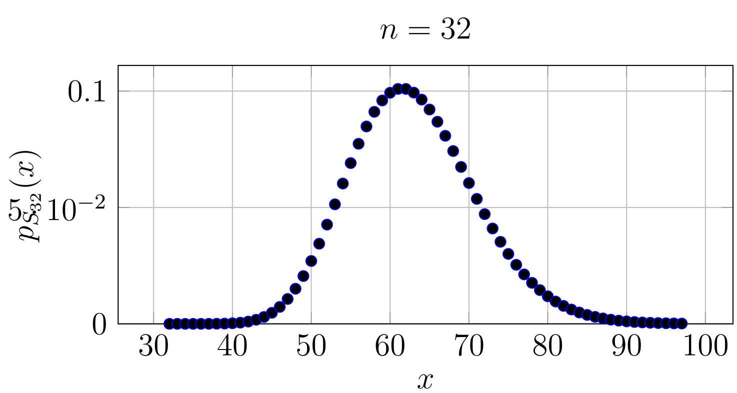






То есть уже при $n = 4$ распределение уже очень похоже на нормальное. Однако если у нас нет такой хорошей симметрии в распределнии X_n, то все не так гладко. Пусть $X_n \sim \mathrm{Geom}(1/2)$. Посмотрим на картинки в этом случае.





Тут уже даже при $n = 32$ все выглядит не так симметрично (хотя и довольно похоже на нормальное распределение).

1.3 Как пользоваться ЦПТ

ЦПТ позволяет приближать сумму с.в. нормальным распределением, а значит для оценок вероятностей мы можем пользоваться таблицами для нормального распределения. Например, рассмотрим такую задачу. На корабль грузят контейнеры, вес каждого контейнера – с.в., следующая экспоненциальному распределению с параметром $\lambda = 1/2$ (то есть $\mu = \sigma = \frac{1}{\lambda} = 2$). Загрузили 100 контейнеров. Какова вероятность, что суммарный вес превзойдет 210?

По сути мы хотим вычислить $\Pr(S_{100} \geq 210)$. Но S_n не приближается стандартным нормальным распределением, а $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ — приближается. Поэтому

$$\Pr[S_{100} > 210] = \Pr\left[\frac{S_{100} - 100 \cdot 2}{\sqrt{100} \cdot 2} \geq \frac{210 - 100 \cdot 2}{\sqrt{100} \cdot 2}\right] = \Pr[Z_n \geq 0.5] = 1 - \Phi(0.5) \approx 0.3085.$$

Дискретный случай

Посмотрим, насколько хорошо работает ЦПТ в случае дискретных распределений. Пусть $\{X_n\}_{n \in \mathbb{N}}$ — последовательность с.в. Бернулли с параметром $p = \frac{1}{2}$ (то есть $\mu = \frac{1}{2}$, $\sigma = \sqrt{p(1-p)} = \frac{1}{2}$). Тогда $S_n \sim \mathrm{Bin}(n, \frac{1}{2})$ и $Z_n = \frac{S_n - n/2}{\sqrt{n}/2}$ примерно следует стандартному нормальному распределению. Но насколько точно? Посчитаем на примере. Пусть $n = 36$. Посчитаем $\Pr(S_n \leq 21)$ точно:

$$\Pr(S_n \leq 21) = \sum_{i=0}^{n} \binom{36}{i} 2^{-36} \approx 0.8785$$

Попробуем применить ЦПТ:

$$\Pr[S_n \leq 21] = \Pr\left[Z_n \leq \frac{21 - 18}{3}\right] = \Pr[Z_n \leq 1] \approx 0.8413.$$

Получилось довольно большая погрешность. Попробуем посчитать по-другому. В дискретном случае $\Pr[S_n \leq 21] = \Pr[S_n < 22]$, поэтому

$$\Pr[S_n \leq 21] = \Pr\left[Z_n < \frac{22 - 18}{3}\right] = \Pr[Z_n < 1.33] \approx 0.9082.$$

Опять погрешность довольно велика. Поэтому часто для дискретных с.в. пользуются правилом коррекци на 1/2, а именно

$$\Pr[S_n \leq 21] = \Pr[S_n \leq 21.5] \Pr\left[Z_n < \frac{21.5 - 18}{3}\right] = \Pr[Z_n < 1.165] \approx 0.8780.$$

И это уже получается довольно точно.

С такой коррекцией можно достаточно точно аппроксимировать даже функцию вероятностей биномиального распеределения (рассмотрим другое число для того, чтобы не зацикливаться на 21)

$$\Pr[S_n = 19] = \Pr[18.5 \leq S_n \leq 19.5] \approx \Pr[0.166 \leq Z_n \leq 0.5] = \Phi(0.5) - \Phi(0.17) \approx 0.124$$

Если же посчитать эту вероятность точно, то $\Pr[S_n = 19] = \binom{36}{19} 2^{-36} \approx 0.1251$. На этот счет есть теорема Муавра-Лапласа, но работает она только для значений функции вероятностей в районе $E[S_n] = np$

Теорема 5. *Теорема Муавра-Лапласа Для любых $p \in [0,1]$ и $q = 1-p$ если $\lim_{n\to+\infty} \frac{k}{n} = p$, то*

$$\lim_{n\to+\infty} \frac{\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}} = 1.$$

В оригинале доказывается через формулу Стирлинга, но это очень громоздко, и мы этого делать не будем.

2 Равенство Вальда

У нас уже была упрощенная версия этой теоремы, однако давайте рассмотрим обобщенный случай.

Теорема 6 (Равенство Вальда)**.** *Пусть $\{X_n\}_{n\in\mathbb{N}}$ — бесконечная последовательность вещественных с.в. Пусть также N — какая-то целочисленная неотрицательная с.в. (то есть N всегда принимает значения из $\mathbb{N}_0$). Допустим также, что выполнены все условия:*

1. *Все X_n имеют конечное матожидание.*
2. *Для любого $n \in \mathbb{N}$ верно, что $E[X_n \mathbb{1}_{\{N\geq n\}}] = E[X_n] \Pr(N \geq n)$ (можно трактовать как событие $N \geq n$ не очень зависит от с.в. X_n).*
3. *Следующий ряд сходится (трактовка: просто техническое требование для абсолютной сходимости $E[S_N]$ и $E[T_N]$, но также часто встречающееся на практике, особенно когда X_n неотрицательные).*

$$\sum_{n=1}^{+\infty} E[|X_n| \mathbb{1}_{\{N\geq n\}}] < +\infty$$

Тогда, если мы обозначим

$$S_N = \sum_{n=1}^{N} X_n,$$

$$T_N = \sum_{n=1}^{N} E[X_n],$$

то верно, что $E[S_N] = E[T_N]$

Если также N имеет конечное матожидание и все X_n имеют одинаковое конечное матожидание, то

$$E[S_N] = E[X_1]E[N].$$

Именно последнее равенство часто и имеют в виду, когда говорят о равенстве Вальда. Заметим, что по сравнению с тем, что у нас было раньше, мы не требуем полной независимости X_n от N, а также допускаем, что у X_n могут быть разные распределения.

Заметим, что в общем случае все три условия необходимы. Если не выполняется первое, то суммы S_N и T_N просто неопределены.

Необходимость второго условия демонстрируется следующим примером. Возьмем последовательность X_n, которые следуют распределению Бернулли с $p = 0.5$. И возьмем довольно зависимый от них $N = 1 - X_1$. В таком случае S_N всегда равно нуль (либо это сумма из нуля слагаемых, либо это одно слагаемое, равное нулю), однако $E[X_1]E[N] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. В данном примере есть явная зависимость события $N \geq 1$ от X_1, причем

$$E[X_1 \mathbb{1}_{\{N \geq 1\}}] = \frac{1}{2} \cdot 1 \cdot 0 + \frac{1}{2} \cdot 0 \cdot 1 = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = E[X_1] \Pr[N \geq 1].$$

Для необходимости третьего (более технического) условия рассмотрим последовательность $X_n = \pm 2^n$, причем знак выбирается равновероятно. И возьмем $N = \min_n \{X_n = +2^n\}$, то есть первый раз, когда мы взяли положительную степень двойки. В такос случае выполняется первое условие, $E[X_n] = 0$, выполняется и второе, так как событие $N \geq n$ зависит только от $X_1, \ldots, X_{n-1}$, но не от X_n. При этом третье условие не выполнено:

$$E[|X_n| \mathbb{1}_{\{N \geq n\}}] = 2^n \cdot \left(\frac{1}{2}\right)^{n-1} = \frac{1}{2}.$$

В таком случае $S_N = 2$, так как независимо от N эта сумма равна $2^N - \sum_{n=1}^{N-1} 2^n = 2$. Но так как $E[X_n] = 0$, то справа в равенстве ноль.

Также пара примеров, когда все рабботает. Очевидный пример — задача про сумму очков, которая выпадает на кости до первой выпавшей единицы. Мы раньше считали матожидание этой суммы так:

$$\sum_{i=1}^{+\infty} \frac{1}{6} \left(\frac{5}{6}\right)^{i-1} (4(i-1) + 1) = 21$$

По равенству Вальда получается то же самое (только надо проверить удовлетворение всех условий)

$$E[S_N] = E[N]E[X_1] = 6 \cdot 3.5 = 21$$

Также неравенство Вальда работает при зависимых друг от друга X_n. Пусть есть какая-то с.в. Z с нулевым матожиданием, и мы определяем $X_n = (-1)^n Z$. По равенству Вальда матожидание их суммы равно нулю. Если рассуждать по-другому, то если N четное, то сумма равна нулю, а если нечетное, то $E[S_N] = E[-Z] = 0$. Но опять же надо проверить все условия.

Доказательство равенства Вальда

Сначала надо доказать абсолютную сходимость $E[S_N]$ и $E[T_N]$.

$$\begin{aligned} E[|S_N|] &= E\left[\sum_{i=1}^{+\infty} |S_i| \mathbb{1}_{N=i}\right] \le E\left[\sum_{i=1}^{+\infty}\sum_{j=1}^{i} |X_j| \mathbb{1}_{N=i}\right] \\ &= E\left[\sum_{j=1}^{+\infty} |X_j| \sum_{i=j}^{+\infty} \mathbb{1}_{N=i}\right] = E\left[\sum_{j=1}^{+\infty} |X_j| \mathbb{1}_{N\ge j}\right] \\ &= \sum_{j=1}^{+\infty} E\left[|X_j| \mathbb{1}_{N\ge j}\right] < \infty, \end{aligned}$$

где переход со второй на третью строчку легален из-за конечности финального ряда, а последнее неравенство следует из условий теоремы.

$$\begin{aligned} E[|T_N|] &= \sum_{i=1}^{+\infty} |T_i| \Pr[N=i] \le \sum_{i=1}^{+\infty} \Pr[N=i] \sum_{j=1}^{i} |E[X_j]| \\ &= \sum_{j=1}^{+\infty} |E[X_j]| \Pr[N \ge j] = \sum_{j=1}^{+\infty} |E[X_j \mathbb{1}_{N\ge j}]| \\ &\le \sum_{j=1}^{+\infty} E[|X_j| \mathbb{1}_{N\ge j}] < \infty, \end{aligned}$$

где в середине второй строчки мы воспользовались вторым условием равенства, а в конце – третьим. Также мы дважды воспользовались неравенством треугольника (в обоих нестрогих неравенствах). Теперь легально доказать само равенство следующим образом:

$$\begin{aligned} E[S_N] &= E\left[\sum_{i=1}^{+\infty} X_i \mathbb{1}_{N\ge i}\right] = \sum_{i=1}^{+\infty} E\left[X_i \mathbb{1}_{N\ge i}\right] \\ &= \sum_{i=1}^{+\infty} E[X_i] \Pr[N \ge i] = \sum_{i=1}^{+\infty} E[X_i] \sum_{j=i}^{+\infty} \Pr[N=j] \\ &= \sum_{j=1}^{+\infty} \Pr[N=j] \sum_{i=1}^{j} E[X_i] = \sum_{j=1}^{+\infty} \Pr[N=j] T_j = E[T_N]. \end{aligned}$$

Мы везде могли менять местами порядок суммирования, так как все у нас сходится абсолютно, плюс воспользовались вторым условием при переходе на вторую строчку. Наконец, если у всех X_n одинаковое матожидание, и у N конечное матожидание, то

$$E[S_N] = E[NE[X_1]] = E[N]E[X_1],$$

Так как $E[X_1]$ в матожидании расценивается как константа.

Лекция 12. Случайные процессы. Мартингалы

6 мая 2021 г.

В остатке курса мы будем изучать случайные процессы. Вообще говоря, случайным процессом называется любое не менее, чем счетное, упорядоченное множество случайных величин $\{X_t\}$. При этом с.в. в этом множестве могут зависеть или не зависеть друг от друга. Индекс t чаще всего интерпретируется как время, и в большинстве случаев принимает либо дискретные значения из $\mathbb{N}$, либо непрерывные из $\mathbb{R}+$.

Заметьте, что в случае дискретного $t \in \mathbb{N}$ множество $\{X_t\}$ является последовательностью. Мы начнем знакомство со случайными процессами с мартингалов. Очень грубо говоря, это такие процессы, матожидание изменение которых в единицу времени равно нулю. Но сначала нам надо ввести несколько дополнительных понятий.

1 Условное матожидание

У нас уже фигурировало матожидание, условное на событии или на другой с.в. В обоих случаях это было обычное матожидание, но с учетом какой-то дополнительной информации, которая у нас есть. Иногда эту информацию "кодируют" в σ-алгебре, и обуславливаются на ней.

Пусть у нас есть вероятностное пространство $(\Omega, \mathcal{F}_0, \Pr)$[1]. Пусть также есть какая-то с.в. X (измеримая и интегрируемая по $\mathcal{F}_0$) и есть какая-то под-алгебра $\mathcal{F} \subset \mathcal{F}_0$. Тогда под $E[X \mid \mathcal{F}]$ мы понимаем любую с.в. Y, такую что

1. Y измерима по $\mathcal{F}$
2. Для любого множества $A \in \mathcal{F}$ выполняется равенство

$$\int_A X d\Pr = \int_A Y d\Pr$$

Оно существует и единственно почти наверное. Единственность почти наверное значит, что если какие-то две с.в. Y_1 и Y_2 удовлетворяют обоим условиям, то $\Pr(Y_1 \neq$

[1] Заметьте, что мы обозначили сигма-алгебру буквой $\mathcal{F}$, а не Σ, но в этом нет особого скрытого смысла. Просто литература, по которой готовилась лекция, использовала это обозначение.

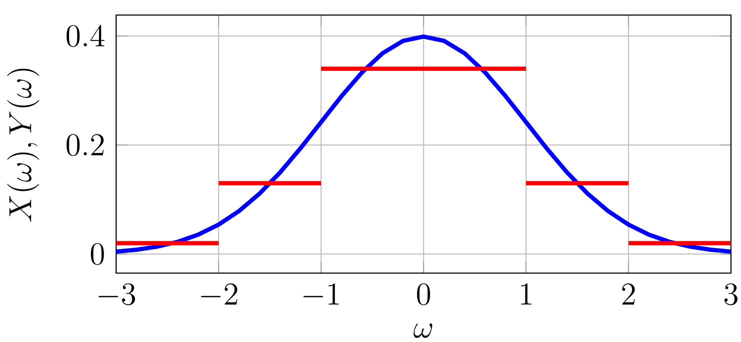
$Y_2) = 0$. Мы не будем доказывать существование и почти единственность, дабы не углубляться в теорию меры.

Как и в других условных вероятностях и матожиданиях, $\mathcal{F}$ в условии стоит воспринимать как имеющуюся у нас информацию. А именно стоит трактовать это так: мы знаем интеграл с.в. на всех множествах из $\mathcal{F}$.

1.1 Примеры

Пример 1.

Для того, чтобы понять, как это работает, рассмотрим следующий пример. Пусть у нас дано вероятностной пространство $(\Omega, \mathcal{F}_0, \Pr)$, где $\Omega = \mathbb{R}$, а $\mathcal{F}_0$ — стандартная σ-алгебра, построенная на множестве всех полуинтервалов. Пусть $\mathcal{F}$ — σ-алгебра, построенная на множестве полуинтервалов с концами в целых числах. Значит, если на исходном пространстве задана какая-то с.в. X, то когда мы говорим про $E[X \mid \mathcal{F}]$, мы говорим о какой-то с.в., которая может менять свое значение только в целых числах (иначе она будет не измерима по $\mathcal{F}$). Во всех остальных точках пространства она должна быть константой. На графике изображен пример, когда X — какая-то непрерывная с.в. (график нормального распределения взят просто для удобства, не путайте, это не плотность вероятности!), а Y, обозначенный красным, как раз является $E[X \mid \mathcal{F}]$.



Пример 2.

Пусть у нас случайная величина X измерима по $\mathcal{F}$. Это значит, что мы знаем ее значение на любом возможном событии, то есть у нас есть вся информация о ней. Поэтому $E[X \mid \mathcal{F}] = X$.

Пример 3.

Пусть с.в. X независима от $\mathcal{F}$. По определению независимости это значит, что $\mathcal{F}$ не дает никакой информации об X. Формально это можно записать так: для любого события $A \in X(\mathcal{F}_0)$ и $B \in \mathcal{F} \subset \mathcal{F}_0$ верно, что

$$\Pr(\{\omega : X(\omega) \in A\} \cap B) = \Pr(\{\omega : X(\omega) \in A\}) \Pr(B).$$

А так как мы нигде ничего не знаем о с.в., то наше лучшее предположение о с.в. — это ее среднее значение, то есть $E[X \mid \mathcal{F}] = E[X]$.

1.2 Свойства условного матожидания

Свойства все те же, что и у обычного. Мы просто перешли в мир, где нам известна какая-то новая информация.

- Линейность: $E[\alpha X + \beta Y \mid \mathcal{F}] = \alpha E[X \mid \mathcal{F}] + \beta E[Y \mid \mathcal{F}]$.
- Монотонность: $\forall \omega\ X(\omega) \le Y(\omega) \Rightarrow E[X \mid \mathcal{F}] \le E[Y \mid \mathcal{F}]$.
- Неравенство Йенсена: $\phi(x)$ — вогнута, тогда $\phi(E[X \mid \mathcal{F}]) \le E[\phi(X) \mid \mathcal{F}]$.

И пара неочевидных, хотя ничего нового для нас тут нет

- Пусть $\mathcal{F}_1 \subset \mathcal{F}_2$. Тогда $E[E(X \mid \mathcal{F}_1) \mid \mathcal{F}_2] = E[E(X \mid \mathcal{F}_2) \mid \mathcal{F}_1] = E[X \mid \mathcal{F}_1]$.
- Если X измерима по $\mathcal{F}$, то $E[XY \mid \mathcal{F}] = XE[Y \mid \mathcal{F}]$.

2 Фильтрации

Фильтрацией называется неубывающая последовательность σ-алгебр в вероятностном пространстве. То есть у нас есть $(\Omega, \mathcal{F}, \Pr)$ и последовательность $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$, причем для любого $n \in \mathbb{N}$ выполняется $\mathcal{F}_n \subset \mathcal{F}$.

Фильтрацию стоит воспринимать как процесс набора информации, количество которой не убывает со временем. Если дана последовательность с.в. $\{X_n\}_{n\in\mathbb{N}}$, то естественный фильтрацией обычно называют

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n),$$

то есть объединение σ-алгебр, задаваемых первыми n с.в. последовательности. В этом смысле запись $E[X_{n+1} \mid \mathcal{F}_n]$ означает с.в., равную X_{n+1}, когда нам уже все известно про $X_1, \dots X_n$.

3 Мартингалы

Случайный процесс (последовательность с.в. $\{X_n\}_{n\in\mathbb{N}}$) называется *Мартингалом* относительно фильтрации $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, если выполнены три условия:

- Все X_n абсолютно интегрируемы, то есть $E[|X_n|] < \infty$
- Для всех n X_n измерима по $\mathcal{F}_n$
- Для всех n выполняется $E[X_{n+1} \mid \mathcal{F}_n] = X_n$

Последнее условие иногда записывается так:

$$E[X_{n+1} - X_n \mid \mathcal{F}_n] = 0$$

Случайный процесс называется *субмартингалом*, если в последнем условии стоит $\geq$. То есть следующая с.в. ожидаемо не меньше предыдущей. Случайный процесс называется *супермартингалом*, если в последнем условии стоит $\leq$. Да, казалось бы, что названия должны быть наоборот, но на самом деле "суб" относится к текущему наблюдению, которое не превышает ожидание следующего наблюдения $E[X_{t+1} \mid \mathcal{F}_n]$.

Чаще всего используется естественная фильтрация, которая означает, что мы знаем первые n шагов процесса. В таком случае вместо второго и третьего условия пишут просто:

$$E[X_{n+1} \mid X_1, \ldots, X_n] = X_n,$$

что есть равенство с.в. То есть случайный процесс зависит только от своего прошлого.

Также возможен случай, когда один случайный процесс X_n является мартингалом относительно другого случайного процесса Y_n. Это происходит, когда вместо второго и третьего условия (в определении через фильтрацию) выполняется

$$E[X_{n+1} \mid Y_1, \ldots, Y_n] = X_n,$$

Примеры мартингалов Самый простой пример: пусть есть последовательность $\{Y_n\}_{n\in\mathbb{N}}$ независимых с.в. с нулевыми матожиданиями, и есть X_n, который задается так:

$$X_n = \begin{cases} x, & \text{if } n = 1, \\ X_{n-1} + Y_{n-1}, & \text{otherwise.} \end{cases}$$

Возьмем фильтрацию $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_{n-1})$. По свойствам условного матожидания легко проверить, что это мартингал:

$$E[X_{n+1} \mid \mathcal{F}_n] = E[X_n \mid \mathcal{F}_n] + E[Y_n \mid \mathcal{F}_n] = X_n + E[Y_n] = X_n.$$

Второй пример. Пусть есть последовательность $\{Y_n\}_{n\in\mathbb{N}}$ независимых с.в. с единичными матожиданиями, и есть X_n, который задается так:

$$X_n = \prod_{i=1}^{n} Y_i,$$

он является мартингалом относительно $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$:

$$E[X_{n+1} \mid \mathcal{F}_n] = E[X_n Y_{n+1} \mid \mathcal{F}_n] = X_n E[Y_n \mid \mathcal{F}_n] = X_n E[Y_n] = X_n.$$

Пара важных свойств (суб-/супер-)мартингалов:

Лемма 1. *Если X_n — субмартингал относительно $\mathcal{F}_n$, то для любых $m > n$ выполняется $E[X_m \mid \mathcal{F}_n] \geq X_n$*

Если X_n — супермартингал относительно $\mathcal{F}_n$, то для любых $m > n$ выполняется $E[X_m \mid \mathcal{F}_n] \leq X_n$

Если X_n — мартингал относительно $\mathcal{F}_n$, то для любых $m > n$ выполняется $E[X_m \mid \mathcal{F}_n] = X_n$

Доказательство. Первое доказывается по индукции. Пусть $m = n + k$. Для $k = 1$ утверждение верно по определению. Для больших k воспользуемся свойствами условного матожидания:

$$E[X_{n+k} \mid \mathcal{F}_n] = E[E[X_{n+k} \mid \mathcal{F}_{n+k-1}] \mid \mathcal{F}_n] \geq E[X_{n+k-1} \mid \mathcal{F}_n].$$

Для доказательства для супермартингала достаточно заметить, что $-X_n$ является субмартингалом, а для доказательства для мартингала достаточно заметить, что он является и суб-, и супер-мартингалом. □

Лемма 2. *Если X_n — мартингал относительно $\mathcal{F}_n$, а ϕ — вогнутая функция, то $\phi(X_n)$ — субмартингал. Если phi неубывающая, то достаточно, чтобы X_n был субмартингалом.*

Доказательство. По неравенству Йенсена в первом случае получаем:

$$E[\phi(X_{n+1}) \mid \mathcal{F}_n] \geq \phi(E[X_{n+1} \mid \mathcal{F}_n]) = \phi(X_n)$$

Во втором случае последнее равенство заменяется на неравенство (пользуемся неубыванием ϕ):

$$E[\phi(X_{n+1}) \mid \mathcal{F}_n] \geq \phi(E[X_{n+1} \mid \mathcal{F}_n]) \geq \phi(X_n)$$

□

3.1 Предсказуемые с.в.

Последовательность с.в. $\{H_n\}_{n\in\mathbb{N}}$ называется предсказуемой относительно фильтрации $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, если для любого n H_n измерима относительно $\mathcal{F}_{n-1}$.

Классический пример — ставки игрока в казино. Если нам известна его стратегия, то когда мы знаем результат его первых n ставок, то мы обычно можем сказать, какой будет его $n + 1$-ая ставка. Например, известная стратегия – удваивать ставку после проигрыша и прекращать игру после первой победы. Такая тактика (при бесконечном имеющимся капитале) гарантирует вам прибыль в размере первой ставки.

Для как раз таких стратегий часто используется следующая последовательность с.в.:

$$(H \cdot X)_n = \sum_{i=1}^{n} H_i(X_i - X_{i-1}),$$

Которая по сути равна балансу игрока, делающего ставки по стратегии H_i, где X_i — выигрыш в i-ой игре при условии единичной ставки.

Теорема 1. *Если X_n — супремартингал, а $H_n \geq 0$ — предсказуемая, ограниченная последовательность с.в. Тогда $(H \cdot X)_n$ — супермартингал.*

Доказательство. Так как $(H \cdot X)_n$ измеримо по $\mathcal{F}_n$, а H_n также измеримо по $\mathcal{F}_{n-1}$,

$$\begin{aligned} E[(X \cdot H)_{n+1} \mid \mathcal{F}_n] &= (H \cdot X)_n + E[H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1} E[(X_{n+1} - X_n) \mid \mathcal{F}_n] = (H \cdot X)_n. \end{aligned}$$

□

3.2 Ограничение прибыли

В данном разделе мы рассмотрим upcrossing inequality (ограничение на пересечения вверх/ограничение на прибыль), которое играет важную роль, например, в биржевых торгах.

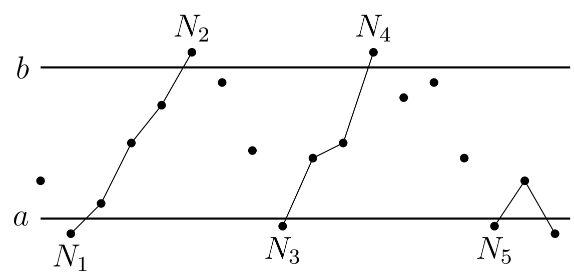
Для начала введем определение *времени останова*. Натуральная с.в. N называется временем останова относительно фильтрации $\{\mathcal{F}_n\}$, если для всех n событие $\{N = n\}$ измеримо относительно $\mathcal{F}_n$. Если мы введем индикаторную величину $H_n = [N \geq n]$, то она будет измерима относительно $\mathcal{F}_{n-1}$ (так как событие $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$), а значит, она будет предсказуема.

Таким образом, если X_n — супермартингал, то $(H \cdot X)_n = X_{n \cap N} - X_0$ — супермартингал. Мы тут обозначили $X_{\min(n,N)} = X_{n \cap N}$ для более короткой записи. Таким образом, и $X_{n \cap N}$ — супермартингал (как сумма двух супермартингалов).

Введем теперь понятие upcrossing (пересечение вверх). Пусть у нас есть случайный процесс X_n, для простоты представим, что это стоимость акций в момент времени n. И у нас следующая стратегия: мы покупаем акции, когда их цена становится ниже какого-то a и продаем, когда их цена становится выше какого-то $b > a$. Тогда за время n наша суммарная прибыль будет $(b-a)$ умножить на число случаев, когда цена акций возрастала от a до b (что есть число пересечений полосы $[a, b]$ снизу вверх до времени n). Введем также времена, когда мы покупаем и продаем акции:

$$\begin{aligned} N_{2k-1} &= \inf\{m > N_{2k-2} : X_m \leq a\} \text{ (время покупки)} \\ N_{2k} &= \inf\{m > N_{2k-1} : X_m \geq b\} \text{ (время продажи)} \end{aligned}$$

и положим $N_0 = -1$. В таком случае upcrossing — это время от покупки до продажи (на рисунке они обозначены линиями).



Введем H_n — индикаторную величину, показывающую, на руках ли у нас акции в момент времени n, или другими словами,

$$H_n = \begin{cases} 1, & \text{если } n \text{ лежит между } N_{2k-1} \text{ и } N_{2k}, \\ 0, & \text{иначе.} \end{cases}$$

Нагонец, введем с.в. $U_n = \sup\{k : N_{2k} \leq n\}$ — число upcrossing'ов, завершенных к моменту времени n. Для него известно следующее утверждение

Теорема 2 (Upcrossing inequality)**.** *Если X_m — субмартингал, тогда*

$$(b-a)E[U_n] \leq E[(X_n-a)^+] - E[(X_0-a)^+]$$

Доказательство. Пусть $Y_m = a + (X_m - a)^+$. Это субмартингал по Лемме 2. Он пересекает полосу $[a, b]$ столько же раз, сколько X_m (единственное отличие — он не уходит ниже a). Тогда наша прибыль описывается как $(H \cdot Y)_n \geq (b-a)U_n$, так как в случае с Y последний начатый upcrossing не приносит убытка.

Пусть теперь $K_m = 1 - H_m$. Тогда $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$. Так как $(K \cdot Y)_n$ – субмартингал, $(K \cdot Y)_n \geq (K \cdot Y)_0 = 0$. То есть $E[(H \cdot Y)_n] \leq E[Y_n - Y_0]$. Остается все подставить и воспользоваться линейностью матожидания. $\square$

Из этого результата следует другой, куда более общий результат о сходимости мартингалов.

Теорема 3. *Если X_n — субмартингал, причем $\sup E[X_n^+] < +\infty$, тогда X_n сходится почти наверное к какому-то X с $E[|X|] < \infty$*

Мы опустим доказательство из-за его излишней техничности, но в целом оно основано на том, что число upcrossing'ов любого отрезка $[a, b]$ конечно, откуда следует сходимость верхнего и нижнего пределов X_n.

3.3 Декомпозиция Дуба

Важным моментом, касающимся предсказуемых с.в. является декомпозиция Дуба.

Теорема 4 (Декомпозиция Дуба)**.** *Пусть X_n — субмартингал. Тогда он может быть представлен как $X_n = M_n + H_n$, где M_n — мартингал, а H_n — предсказуемая неубывающая последовательность.*

Доказательство. Возьмем $A_0 = 0$ и каждый следующий A_n такой, что

$$A_{n+1} - A_n = E[X_{n+1} \mid \mathcal{F}_n] - X_n$$

Правая часть равенства измерима по $\mathcal{F}_n$, значит, по индукции можно доказать, что A_{n+1} измерима по $\mathcal{F}_n$, то есть является предсказуемой. Также заметим, что правая часть неотрицательна (так как X_n — субмартингал), а значит, A_n — возрастающая последовательность.

Остается показать, что M_n — мартингал, то есть

$$\begin{aligned} E[M_n \mid \mathcal{F}_{n-1}] &= E[X_n - A_n \mid \mathcal{F}_{n-1}] \\ &= E[X_n \mid \mathcal{F}_{n-1}] - A_n = X_{n-1} - A_{n-1} = M_{n-1}. \end{aligned}$$

□

4 Ветвящиеся процессы

Ветвящиеся процессы в общем случае это процессы, описывающие бесполое размножение каких-то организмов с помощью генеалогических деревьев. Одним из простейших примеров является следующий процесс, называемый процессом Гальтона-Ватсона.

Пусть у нас есть двумерная последовательность независимых, одинаково распределенных с.в. $X_i^n \in \mathbb{N}$. И также есть процесс $\{Z_n\}_{n\in\mathbb{N}}$, описываемый следующим образом: $Z_0 = 1$ и

$$Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} X_i^{n+1}, \text{ if } Z_n > 0, 0, \text{else}. \end{cases}$$

В данном случае X_i^n — это число потомков, которых породит i-ая особь из поколения n, если она вообще будет существовать. А сам Z_n отображает размер поколения.

Лемма 3. *Пусть $\mathcal{F}_n = \sigma(X_i^m : i \geq 1, 1 \leq m \leq n)$ (фильтрация, знающая все про первые n поколений). И пусть $\mu = EX_i^n$. Тогда $\frac{Z_n}{\mu^n}$ мартингал относительно фильтрации $\mathcal{F}_n$.*

Доказательство.

$$E[Z_{n+1} \mid \mathcal{F}_n] = E[X_1^{n+1} + \cdots + X_k^{n+1} \mid \mathcal{F}_n] = k\mu = \mu Z_n.$$

Поэтому $E[\frac{Z_{n+1}}{\mu^{n+1}} \mid \mathcal{F}_n] = \frac{Z_n}{\mu_n}$. □

Отсюда понятно, что Z_n ведет себя примерно как μ^n, откуда следуют следующие наблюдения.

Лемма 4. *Если $\mu < 1$, то $Z_n = 0$ начиная с какого-то n почти наверное, то есть $\frac{Z_n}{\mu^n} \to 0$.*

Доказательство.

$$\Pr[Z_n > 0] \leq E[Z_n] = \mu^n \to 0.$$

□

Лемма 5. *Если $\mu = 1$ и $\Pr[X_i^n = 1] < 1$, тогда $Z_n = 0$ начиная с какого-то n почти наверное.*

Доказательство. Z_n является мартингалом, значит, по теореме о сходимости он сходится к какой-то с.в. Так как Z_n целочисленна, то и предел целочисленный.

Пусть этот предел не ноль. Тогда

$$\Pr[Z_n = k \text{ for all } n \geq N] = 0,$$

поэтому единственный вариант предела есть ноль. □

Для следующей леммы мы введем понятие производящей функции. Пусть $\phi(s) = \sum_{k \geq 0} p_k s^k$, где $p_k = \Pr[X_i^n = k]$

Лемма 6. *Если $\mu > 1$ и $Z_0 = 1$, тогда $\Pr[Z_n = 0$ начиная с какого-то $N] = \rho$, где ρ — единственное решение уравнения $\phi(\rho) = \rho \in [0, 1)$.*

Оставим без доказательства (слишком сложно).

5 Полезные неравенства

5.1 Неравенство Азумы

Теорема 5. *Пусть $\{X_n\}_{n \in \mathbb{N}}$ — супермартингал. Пусть также его изменение ограничено, то есть существует такая последовательность $\{c_n\}_{n \in \mathbb{N}}$, что для любого n верно, что $|X_{n+1} - X_n| \leq c_n$. Тогда для любой $\delta > 0$ верно, что*

$$\Pr[X_n - X_0 \geq \delta] \leq \exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Очень легко переделать это в неравенство для субмартингала и в двустороннее неравенство для мартингала:

Теорема 6. *Пусть $\{X_n\}_{n \in \mathbb{N}}$ — субмартингал. Пусть также его изменение ограничено, то есть существует такая последовательность $\{c_n\}_{n \in \mathbb{N}}$, что для любого n верно, что $|X_{n+1} - X_n| \leq c_n$. Тогда для любой $\delta > 0$ верно, что*

$$\Pr[X_n - X_0 \leq -\delta] \leq \exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Теорема 7. *Пусть $\{X_n\}_{n \in \mathbb{N}}$ — мартингал. Пусть также его изменение ограничено, то есть существует такая последовательность $\{c_n\}_{n \in \mathbb{N}}$, что для любого n верно, что $|X_{n+1} - X_n| \leq c_n$. Тогда для любой $\delta > 0$ верно, что*

$$\Pr[|X_n - X_0| \geq \delta] \leq 2\exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Данное неравенство доказывается через общие границы Чернова, примененные к $X_n - X_0 = \sum_{i=1}^{n-1}(X_{i+1} - X_i)$, затем выделению каждого члена суммы в отдельный множитель, который ограничивается через лемму Хёффдинга.

5.2 Неравенство МакДиармида

Теорема 8. *Пусть $X_1, \dots, X_n$ — независимые с.в., определенные на $\Omega_1, \dots, \Omega_n$ соответственно. Пусть $\Omega = \Omega_1 \times \dots \times \Omega_n$. Пусть также есть функция $f : \Omega \to \mathbb{R}$, для которой существует такой набор чисел $c_1, \dots, c_n$, что для любых двух векторов элементарных исходов $\omega, \nu \in \Omega$, отличных только в i-ом компоненте верно, что $|f(\omega) - f(\nu)| \le c_i$. Пусть также $X = f(X_1, \dots, X_n)$. Тогда*

$$\Pr[X - E[X] \ge \lambda] \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n-1} c_i}\right),$$
$$\Pr[X - E[X] \le -\lambda] \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n-1} c_i}\right).$$

Лекция 13. Мартингалы (продолжение). Процесс Бернулли

12 мая 2021 г.

1 Декомпозиция Дуба

Важным моментом, касающимся предсказуемых с.в. является декомпозиция Дуба.

Теорема 1 (Декомпозиция Дуба)**.** *Пусть X_n — субмартингал. Тогда он может быть представлен как $X_n = M_n + H_n$, где M_n — мартингал, а H_n — предсказуемая неубывающая последовательность.*

Доказательство. Возьмем $A_0 = 0$ и каждый следующий A_n такой, что

$$A_{n+1} - A_n = E[X_{n+1} \mid \mathcal{F}_n] - X_n$$

Правая часть равенства измерима по $\mathcal{F}_n$, значит, по индукции можно доказать, что A_{n+1} измерима по $\mathcal{F}_n$, то есть является предсказуемой. Также заметим, что правая часть неотрицательна (так как X_n — субмартингал), а значит, A_n — возрастающая последовательность.

Остается показать, что M_n — мартингал, то есть

$$\begin{aligned} E[M_n \mid \mathcal{F}_{n-1}] &= E[X_n - A_n \mid \mathcal{F}_{n-1}] \\ &= E[X_n \mid \mathcal{F}_{n-1}] - A_n = X_{n-1} - A_{n-1} = M_{n-1}. \end{aligned}$$

□

2 Ветвящиеся процессы

Ветвящиеся процессы в общем случае это процессы, описывающие бесполое размножение каких-то организмов с помощью генеалогических деревьев. Одним из простейших примеров является следующий процесс, называемый процессом Гальтона-Ватсона.

Пусть у нас есть двумерная последовательность независимых, одинаково распределенных с.в. $X_i^n \in \mathbb{N}$. И также есть процесс $\{Z_n\}_{n\in\mathbb{N}}$, описываемый следующим образом: $Z_0 = 1$ и

$$Z_{n+1} = \Big\{ \sum_{i=1}^{Z_n} X_i^{n+1}, \text{ if } Z_n > 0, 0, \text{else}.$$

В данном случае X_i^n — это число потомков, которых породит i-ая особь из поколения n, если она вообще будет существовать. А сам Z_n отображает размер поколения.

Лемма 1. *Пусть* $\mathcal{F}_n = \sigma(X_i^m : i \geq 1, 1 \leq m \leq n)$ *(фильтрация, знающая все про первые n поколений). И пусть* $\mu = EX_i^n$. *Тогда* $\frac{Z_n}{\mu^n}$ *мартингал относительно фильтрации* $\mathcal{F}_n$.

Доказательство.

$$E[Z_{n+1} \mid \mathcal{F}_n] = E[X_1^{n+1} + \cdots + X_k^{n+1} \mid \mathcal{F}_n] = k\mu = \mu Z_n.$$

Поэтому $E[\frac{Z_{n+1}}{\mu^{n+1}} \mid \mathcal{F}_n] = \frac{Z_n}{\mu_n}$. □

Отсюда понятно, что Z_n ведет себя примерно как μ^n, откуда следуют следующие наблюдения.

Лемма 2. *Если* $\mu < 1$, *то* $Z_n = 0$ *начиная с какого-то n почти наверное, то есть* $\frac{Z_n}{\mu^n} \to 0$.

Доказательство.

$$\Pr[Z_n > 0] \leq E[Z_n] = \mu^n \to 0.$$

□

Лемма 3. *Если* $\mu = 1$ *и* $\Pr[X_i^n = 1] < 1$, *тогда* $Z_n = 0$ *начиная с какого-то n почти наверное.*

Доказательство. Z_n является мартингалом, значит, по теореме о сходимости он сходится к какой-то с.в. Так как Z_n целочисленна, то и предел целочисленный.

Пусть этот предел не ноль. Тогда

$$\Pr[Z_n = k \text{ for all } n \geq N] = 0,$$

поэтому единственный вариант предела есть ноль. □

Для следующей леммы мы введем понятие производящей функции. Пусть $\phi(s) = \sum_{k \geq 0} p_k s^k$, где $p_k = \Pr[X_i^n = k]$

Лемма 4. *Если* $\mu > 1$ *и* $Z_0 = 1$, *тогда* $\Pr[Z_n = 0$ *начиная с какого-то* $N] = \rho$, *где* ρ *— единственное решение уравнения* $\phi(\rho) = \rho \in [0, 1)$.

Оставим без доказательства (слишком сложно).

3 Полезные неравенства

3.1 Неравенство Азумы

Теорема 2. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ *— супермартингал. Пусть также его изменение ограничено, то есть существует такая последовательность* $\{c_n\}_{n\in\mathbb{N}}$*, что для любого n верно, что* $|X_{n+1} - X_n| \le c_n$*. Тогда для любой* $\delta > 0$ *верно, что*

$$\Pr[X_n - X_0 \ge \delta] \le \exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Очень легко переделать это в неравенство для субмартингала и в двустороннее неравенство для мартингала:

Теорема 3. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ *— субмартингал. Пусть также его изменение ограничено, то есть существует такая последовательность* $\{c_n\}_{n\in\mathbb{N}}$*, что для любого n верно, что* $|X_{n+1} - X_n| \le c_n$*. Тогда для любой* $\delta > 0$ *верно, что*

$$\Pr[X_n - X_0 \le -\delta] \le \exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Теорема 4. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ *— мартингал. Пусть также его изменение ограничено, то есть существует такая последовательность* $\{c_n\}_{n\in\mathbb{N}}$*, что для любого n верно, что* $|X_{n+1} - X_n| \le c_n$*. Тогда для любой* $\delta > 0$ *верно, что*

$$\Pr[|X_n - X_0| \ge \delta] \le 2\exp\left(-\frac{\delta^2}{2\sum_{i=1}^{n-1} c_i}\right).$$

Данное неравенство доказывается через общие границы Чернова, примененные к $X_n - X_0 = \sum_{i=1}^{n-1}(X_{i+1} - X_i)$, затем выделению каждого члена суммы в отдельный множитель, который ограничивается через лемму Хёффдинга.

3.2 Неравенство МакДиармида

Теорема 5. *Пусть* $X_1, \dots, X_n$ *— независимые с.в., определенные на* $\Omega_1, \dots, \Omega_n$ *соответственно. Пусть* $\Omega = \Omega_1 \times \dots \times \Omega_n$*. Пусть также есть функция* $f : \Omega \to \mathbb{R}$*, для которой существует такой набор чисел* $c_1, \dots, c_n$*, что для любых двух векторов элементарных исходов* $\omega, \nu \in \Omega$*, отличных только в i-ом компоненте верно, что* $|f(\omega) - f(\nu)| \le c_i$*. Пусть также* $X = f(X_1, \dots, X_n)$*. Тогда*

$$\Pr[X - E[X] \ge \lambda] \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n-1} c_i}\right),$$

$$\Pr[X - E[X] \le -\lambda] \le \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n-1} c_i}\right).$$

4 Процесс Бернулли

После сложных мартингалов мы рассмотрим более простой процесс Бернулли (для подготовки к рассмотрению более сложного процесса Пуассона). Процессом Бернулли называется случайный процесс $\{X_t\}_{t\in\mathbb{N}}$, в котором все X_t независимы и следуют одинаковому распределению Бернулли с одинаковым параметром p. С его помощью часто моделируют разные реальные штуки:

- Браки на производстве
- Прибытие посетителя в магазин в интервалы времен
- Поступление запросов на сервер

Как вообще стоит рассматривать случайный процесс, в том числе процесс Бернулли? В случае, когда у нас конечное множество с.в. $\{X_t\}$, то нам достаточно задать совместную функцию вероятности или совместную плотность вероятности для всех возможных подмножеств $\{X_{t_1}, \ldots, X_{t_k}\}$. Это необходимо, так как с.в. могут быть и зависимы. Но в случае с бесконечными последовательностями даже если мы определим совместные функции/плотности вероятности на всех конечных подмножествах, мы все равно не сможем определить, например, вероятность того что все $X_t = 1$.

Поэтому в случае со случайными процесами стоит рассматривать Ω как множество всех возможных последовательностей, которые могут получиться из значений случайных величин. В случае с процессом Бернулли — это все возможные последовательности из нолей и единиц. В таком случае легко найти вероятность события, что все $X_t = 1$ как в случае, когда $p < 1$:

$$\begin{aligned} &\Pr(\forall t \in \mathbb{N}\ X_t = 1) \le \Pr(\forall t \in [1..n]\ X_t = 1) = p^n \\ \Rightarrow &\Pr(\forall t \in \mathbb{N}\ X_t = 1) = 0, \end{aligned}$$

так и для $p = 1$:

$$\Pr(\forall t \in \mathbb{N}\ X_t = 1) \ge 1 - \sum_{t=1}^{+\infty} \Pr[X_t = 0] = 1.$$

В дальнейшем случай с $p = 1$ (как и с $p = 0$) мы рассматривать не будем, так как ничего интересного в этом нет.

4.1 Свойства процесса Бернулли

Рассмотрим несколько свойств процесса Бернулли. Напомним, что так как какждый X_t следует распределению Бернулли независимо от других с.в., то

- $E[X_t] = p$
- $\mathrm{Var}(X_t) = p(1-p)$

Рассмотрим следующие две случайные величины.

S_n — **число единиц в первые n единиц времени.** Так как X_t независимы, то $S_n \sim \mathrm{Bin}(n, p)$, причем

- $\Pr[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$
- $E[S_n] = np$
- $\mathrm{Var}(S_n) = np(1-p)$

T_1 — **время первой единицы.** Так как X_t независимы, то $T_1 \sim \mathrm{Geom}(p)$, причем

- $\Pr[T_1 = k] = p(1-p)^{k-1}$
- $E[T_n] = \frac{1}{p}$
- $\mathrm{Var}(T_n) = \frac{1-p}{p^2}$

Беспамятство. Если мы рассматриваем процесс $Y_t = X_{t+n}$ для какого-то фиксированного n, то это опять будет процесс Бернулли, так как:

- Все Y_t независимы
- Все Y_t имеют распределение $\mathrm{Bern}(p)$

Более того, если мы говорим, что $Y_t = X_{t+N}$, где N — какая-то случайная величина, то в некоторых случаях мы также получаем процесс Бернулли. Например:

- Пусть N — время первого успеха в процессе X_t. Тогда все X_{t+N} независимы друг от друга и от N, поэтому являются процессом Бернулли.
- Пусть теперь N — время перед первым успехом. Тогда X_{N+1} уже не независим от N, а точно равен единице.

В общем случае Y_t будет процессом Бернулли, если N является временем останова, то есть событие $N \geq n$ зависит только от $X_1, \ldots, X_{n-1}$.

Беспамятство дает нам много других полезных свойств процесса Бернулли.

Продолжительность занятого периода. Рассмотрим сервер, на который в каждую единицу времени приходит или не приходит запрос. Это можно описать процессом Бернулли, где событие "во время t пришел запрос" соответсвует $X_t = 1$. Скажем, что сервер занят во время t, если в это время на него пришел запрос. Занятый период — это отрезок $[t_1..t_2]$, такой, что во все $t \in [t_1..t_2]$ сервер занят, но в моменты $t_1 - 1$ и $t_2 + 1$ — свободен или еще не начался. Нас интересует продолжительность первого занятого периода.

Пусть T_1 — время первого запроса на сервер. Благодаря беспамятству мы можем сказать,что X_{t+T_1} задают новый процесс Бернулли с тем же параметром p. В этом процессе первый ноль появится во время $T_2 \sim \mathrm{Geom}(1-p)$. При этом T_2 в таком случае и будет равно длине занятого периода.

Время k-ого успеха. Пусть Y_k — время, в которое на сервер прибыл k-ый по счету запрос. Пусть $T_k = Y_k - Y_{k-1}$, то есть интервал между $k-1$-ым и k-ым запросами. При этом $Y_k = T_1 + \cdots + T_k$.

Благодаря беспамятству мы можем сказать, что в каждый момент Y_k мы запускаем процесс заново. Поэтому каждый $T_k \sim \mathrm{Geom}(p)$. Отчюда мы сразу можем сказать про Y_k, что

- $E[Y_k] = \sum_{i=1}^{k} E[T_k] = \frac{k}{p}$
- $\mathrm{Var}(Y_k) = \sum_{i=1}^{k} \mathrm{Var}(T_k) = \frac{k(1-p)}{p^2}$.

Также уже было в разборе контрольной, что функция распределения имеет следующий вид:

$$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k},$$

так как мы точно знаем, что во время t произошел успех, и нам надо расставить остальные $k-1$ успех по оставшимся $t-1$ ячейкам времени.

5 Слияние и разделение процесса Бернулли

Рассмотрим случай, когда у нас на сервер могут поступать запросы с двух источников. Поступление запросов от каждого из источников описывается процессом Бернулли, причем эти два процесса (назовем их X_t и Y_t) независимы. Рассмотрим процесс Z_t, где Z_t — индикаторная с.в. события "в момент времени t поступил запрос от хотя бы одного источника". То есть $Z_t = \max\{X_t, Y_t\}$. Заметим, что Z_t не зависит от любого $Z_{s \neq t}$, так как является функцией от двух величин X_t и Y_t, не зависящих от любых других $X_{s \neq t}$ и $Y_{s \neq t}$. А также

$$\Pr[Z_t = 1] = \Pr[X_t = 1 \cup Y_t = 1] = 1 - \Pr[X_t = 0 \cap Y_t = 0] = p + q - pq,$$

где p и q — параметры процессов X_t и Y_t соответственно. Таким образом, Z_t также является процессом Бернулли с параметром $p+q-pq$. Заметье, что вместо логического "И" мы можем применить любую другую операцию к двум поступающим запросам. Например, одновременный запрос от двух источников мы можем считать коллизией, и не считать его за запрос. Тогда Z_t также будет процессом Бернулли, но с другим параметром $p + q - 2pq$.

Мы также можем разделять процесс Бернулли на два. Пусть у нас есть процесс Бернулли Z_t с параметром p, описывающий запросы на сервер. Каждый раз, когда на сервер поступает запрос, он решает, в какой из двух выходов его отправить дальше, причем он отправляет его в первый выход с вероятностью q и во второй — с вероятностью $1-q$, принимая это решение независимо от всего остального. В таком случае запросы, отправленные в первый выход, X_t являются процессом Бернулли с параметром pq. (TODO: доказать независимость и бернулевость всех X_t). Запросы на втором выходе Y_t — тоже, но с параметром $p(1-q)$. Однако процессы X_t и Y_t не являются независимыми, так как если мы знаем, что $X_t = 1$, то мы точно знаем, что $Y_t = 0$.

6 Аппроксимация процесса Бернулли

Рассмотрим процесс Бернулли с параметром p. Чему равна вероятность того, что за время n будет k успехов? По тому, что мы считали, это будет

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Но что будет, если мы раздробим время на меньше интервалы (при этом вероятность успеха в каждый интервал будет падать обратно пропорционально длине интервала)? То есть для каждого n мы будем брать $p = \lambda n$ для какой-то фиксированной λ и устремим $n \to \infty$. В таком случае будет:

$$\begin{aligned}
p_X(k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\to 1 \cdot 1 \cdot \ldots \cdot 1 \cdot \frac{\lambda^k}{k!} e^{-\lambda} \cdot 1 = \frac{\lambda^k}{k!} e^{-\lambda}.
\end{aligned}$$

И это подводит нас к процессу Пуассона, который мы изучим на следующей паре.

Лекция 14. Процесс Пуассона

25 мая 2021 г.

1 Аппроксимация процесса Бернулли

Рассмотрим процесс Бернулли с параметром p. Чему равна вероятность того, что за время n будет k успехов? По тому, что мы считали, это будет

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Но что будет, если мы раздробим время на меньшие интервалы (при этом вероятность успеха в каждый интервал будет падать обратно пропорционально длине интервала)? То есть для каждого n мы будем брать $p = \lambda n$ для какой-то фиксированной λ и устремим $n \to \infty$. В таком случае будет:

$$\begin{aligned}
p_X(k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\to 1 \cdot 1 \cdot \ldots \cdot 1 \cdot \frac{\lambda^k}{k!} e^{-\lambda} \cdot 1 = \frac{\lambda^k}{k!} e^{-\lambda}.
\end{aligned}$$

И это подводит нас к процессу Пуассона.

2 Процесс Пуассона: определение

Процессом Пуассона называется некоторый процесс с вещественным временем, и в каждый момент времени происходит или не происходит событие. Однако в силу непрерывности вероятность, что в каждый конкретный момент времени произойдет событие равна нулю, зато на временном интервале может произойти более одного события. Для того, чтобы такой процесс назывался процессом Пуассона, мы требуем:

1. Независимость непересекающихся временных интервалов: число событий, произошедших в интервал времени τ_1 не зависит от числа событий в другом интервале τ_2.

2. Неизменяемость по времени. Вероятность $P_\tau(k)$ того, что в интервал длиной τ произойдет k событий не зависит от того, в какой именно момент времени начался интервал. Заметим, что для любого τ выполняется $\sum_{k=0}^{+\infty} P_\tau(k) = 1$.

3. Для маленьких интервалов длиною δ мы также требуем

$$P_\delta(k) \approx \begin{cases} 1-\lambda\delta, & k=0 \\ \lambda\delta, & k=1 \\ 0, & \text{иначе.} \end{cases}$$

   То есть мы хотим, чтобы для достаточно маленького интервала вероятность того, что в нем уместилось более двух событий была ничтожно маленькая, а вероятность одного события была пропорционально длине интервала. Более точное требование: при $\delta \to 0$ должно выполняться

$$P_\delta(k) = \begin{cases} 1-\lambda\delta + O(\delta^2), & k=0 \\ \lambda\delta + O(\delta^2), & k=1 \\ O(\delta^2), & \text{иначе.} \end{cases}$$

   В данном случае λ называется *интенсивностью* процесса.

Примеры из реальной жизни:

1. Смерти от удара лошади в Прусской армии (самое первое применение, причем не у Пуассона, а позже Борткевичем)

2. Испускание радиоактивной частицы

3. Детекция фотона от слабого источника

4. Потрясения рынка (финансовые кризисы и т.п.)

5. Сфера обслуживания: звонки в колл-центры, приход покупателя, ...

3 Распределение Пуассона

Пусть случайная величина N_τ — число событий в процессе Пуассона с интенсивностью λ, которые произошли в интервал длиной τ. Попробуем оценить ее распределение. Для этого разобьем весь интервал длиной τ на отрезки (слоты) длиной δ, всего получится $\frac{\tau}{\delta}$ отрезков. При этом вероятность p того, что есть хотя бы один отрезок, куда попали два или более событий не больше, чем сумма вероятностей по всем отрезкам того, что в этот отрезок попало два и более событий. По третьему свойству процесса Пуассона имеем.

$$p \le \sum_{i=1}^{\tau/\delta} \Pr[N_\delta \ge 2] = \frac{\tau}{\delta} O(\delta^2) = O(\tau\delta).$$

Это значит, что мы можем добиться сколь угодно маленькой вероятности того, что у нас есть слот с двумя и более событиями, а значит, мы можем считать, что $N_\tau \sim \mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta)$. При этом в начале лекции мы показали, что

$$\Pr[N_\tau = k] = \lim_{\delta \to 0} \Pr[X \sim \mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta + O(\delta^2)) = k] = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}.$$

Говорят, что N_τ в таком случае следует распределению Пуассона с параметром $(\lambda\tau)$:

$$N_\tau \sim \mathrm{Poisson}(\lambda\tau).$$

4 Свойства распределения Пуассона

Матожидание и дисперсия.

Матожидание распределения Пуассона можно посчитать напрямую

$$\begin{aligned} E[N_\tau] &= \sum_{k=0}^{+\infty} k \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau} \\ &= e^{-\lambda\tau} \lambda\tau \sum_{k=0}^{+\infty} \frac{(\lambda\tau)^k}{(k-1)!} \\ &= \lambda\tau. \end{aligned}$$

А можно сказать, что оно близко к биномиальному распределению $\mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta + O(\delta^2))$, то есть его матожидание есть

$$E[N_\tau] = \lim_{\delta \to 0} \frac{\tau}{\delta}(\lambda\delta + O(\delta^2)) = \lambda\tau.$$

Дисперсию тоже посчитаем этим методом:

$$\mathrm{Var}(N_\tau) = \lim_{\delta \to 0} \frac{\tau}{\delta}(\lambda\delta + O(\delta^2))(1 - \lambda\delta + O(\delta^2)) = \lambda\tau.$$

То есть для распределения Пуассона с параметром $\lambda\tau$

$$E[N_\tau] = \mathrm{Var}(N_\tau) = \lambda\tau$$

Интенсивность процесса Пуассона при этом можно интерпретировать как среднее число событий, происходящих в единицу времени:

$$\lambda = \frac{E[N_\tau]}{\tau}.$$

Время первого события T_1.

Вероятность того, что первое событие произошло через время большее, чем t, есть вероятность того, что за время t произошло ноль событий, поэтому можем посчитать функцию распределения T_1.

$$F_{T_1}(t) = \Pr[T_1 \leq t] = 1 - \Pr[T_1 > t] = 1 - P_t(0) = 1 - e^{-t\lambda}.$$

Получается, время T_1 следует распределению $\mathrm{Exp}(\lambda)$. Заметим, что в силу беспамятства экспоненциального распределения, при условии $T_1 > t$ величина $T_1 - t$ также следует $\mathrm{Exp}(\lambda)$.

Время k-ого события T_k.

Можно попытаться посчитать функцию распределения T_k тем же образом:

$$\Pr[T_k \leq t] = \sum_{i=k}^{+\infty} P_t(i),$$

но это чревато дебрями сложной математики (остаток ряда экспоненты выглядит не очень мило). Поэтому давайте попробуем посчитать плотность вероятности следующим образом:

$$f_{T_k}(t)\delta \approx \Pr[t \leq T_k \leq t + \delta] = P_t(k-1)(\lambda\delta + O(delta^2)) + (\sum_{i=0}^{k-2} P_t(i))O(\delta(2)) \approx P_t(k-1)\lambda\delta.$$

Таким образом,

$$f_{T_k}(t) \approx \lambda P_t(k-1) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-(k-1)t}.$$

Это, кстати, называется распределением Эрланга (частный случай Гамма-распределения, как некоторые из вас знают с контрольной).

Беспамятство.

Как и в случае с процессом Бернулли, у процесса Пуассона есть беспамятство. Рассмотрим случай, когда мы начинаем смотреть процесс в фиксированное время t. Тогда для наблюдаемого нами процесса по-прежнему выполняются все три свойства процесса Пуассона. То же верно, когда T — время останова.

Поэтому если мы начинаем смотреть в любой фиксированный момент времени, то время до следующего события следует экспоненциальному распределению. Также если мы начинаем смотреть в момент T_k, когда произошло событие k, то время до следующего события $Y_{k+1} = T_{k+1} - T_k$ также следует $\mathrm{Exp}(\lambda)$. Поэтому существует альтернативное определение процесса Пуассона.

Пусть есть последовательность независимых с.в. $\{Y_k\}_{k \in \mathbb{N}}$, следующих распределению $\mathrm{Exp}(\lambda)$. Тогда последовательность с.в. $T_k = \sum_{i=1}^{k} Y_k$ образует процесс Пуассона.

Отсюда следуют свойства T_k:

- $E[T_k] = \frac{k}{\lambda}$
- $\mathrm{Var}(T_k) = \frac{k}{\lambda^2}$

5 Сравнение с процессом Бернулли

	Пуассон	Бернулли
Время	Непрерывное	Дискретное
Интенсивность	λ в единицу времени	p в тайм-слот
Распределение числа событий в интервале	$\mathrm{Poisson}(\lambda\tau)$	$\mathrm{Bin}(n, p)$
Интервалы между событиями	$\mathrm{Exp}(\lambda)$	$\mathrm{Geom}(p)$
k-е событие	Эрланг	Паскаль (inverse Bin)

6 Сумма двух с.в. Пуассона

Рассмотрим процесс Пуассона с $\lambda = 1$. Рассмотрим два последовательных временных интервала длиной μ и ν. Пусть X — число событий в первом интервале, а Y — число событий во втором. Тогда $X \sim \mathrm{Poisson}(\mu)$ и $Y \sim \mathrm{Poisson}(\nu)$, причем они независимы.

С.в. $Z = X + Y$ есть число событий в интервале длиной $\mu + \nu$, то есть $Z \sim \mathrm{Poisson}(\mu + \nu)$. Таким образом, для двух независимых с.в. Пуассона выполняется

$$\mathrm{Poisson}(\mu) + \mathrm{Poisson}(\nu) = \mathrm{Poisson}(\mu + \nu),$$

что есть равенство распределений (то есть совпадение функции распределения во всех точках).

7 Слияние двух процессов Пуассона

Аналогично с процессами Бернулли, мы можем сливать процесс Пуассона из двух независимых процессов. Пусть у нас есть две лампочки разных цветов, зеленая и красная. Они независимо мигают с интенсивностью λ_1 и λ_2 (то есть мигания каждой лампочки — процессы Пуассона). У нас есть датчик, регистрирующий вспышку, но не воспринимающий цвет. Тогда регистрации на датчике тоже будут процессом Пуассона.

1. Выполняется независимость непересекающихся интервалов
2. Выполняется неизменность во времени
3. Рассмотрим вероятность того, то в интервале длины δ произошло k событий

		$1 - \lambda_1\delta$	$\lambda_1\delta$	$O(\delta^2)$
		0	1	≥ 2
$1 - \lambda_2\delta$	0	$1 - (\lambda_1 + \lambda_2)\delta$	$\lambda_1\delta$	
$\lambda_2\delta$	1	$\lambda_2\delta$		
$O(\delta^2)$	≥ 2			$O(\delta^2)$

То есть интенсивность нового процесса есть $\lambda_1 + \lambda_2$.

Также мы можем показать вероятность того, из какого процесса произошло k-е событие. Возьмем очень маленький интервал δ, в котором это событие произошло, и по формуле условной вероятности и по таблице выше почитаем, что вероятность того, что событие пришло из первого процесса есть $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Причем источники различных событий в слитом процессе независимы. То есть, последовательность источников событий образует процесс Бернулли.

Пример.

Пусть у нас есть три лампочки, время работы которых следует $\mathrm{Exp}(\lambda)$. Когда перегорит первая лампочка и когда перегорит последняя?

Насчет распределения времени перегорания первой можно посчитать в тупую тройной интеграл функции $\min\{T_1, T_2, T_3\}$, но это скучно (можно написать этот страшный интеграл на доске).

Можно посчитать через функцию распределения:

$$\Pr[\min\{T_1, T_2, T_3\} \geq t] = \Pr[T_1 \geq t] \Pr[T_2 \geq t] \Pr[T_3 \geq t] = e^{-3\lambda t}$$

А можно просто сказать, что перегорание каждой лампочки — это первое событие в процессе Пуассона с интенсивностью λ. Поэтому слитый процесс имеет интенсивность 3λ, и первая лампочка перегорает через время $T \sim \mathrm{Exp}(3\lambda)$.

Для оценки того, когда перегорает последняя, можем заметить, что если мы смотрим на процесс после с момента перегорания первой лампочки, то мы смотрим на слияние двух процессов, которое имеет интенсивность 2λ, то есть время между первой и второй следует $\mathrm{Exp}(2\lambda)$. Ну а последняя перегорает еще через $\mathrm{Exp}(\lambda)$. Таким образом, время перегорания последней лампочки следует распределению $\mathrm{Exp}(3\lambda) + \mathrm{Exp}(2\lambda) + \mathrm{Exp}(\lambda)$, где все три распределения — независимы.

8 Разделение процесса Пуассона

Рассмотрим процесс Пуассона с интенсивностью λ. Можем при каждом событии подкидывать нечестную монетку и с вероятностью q отправлять событие в поток A и с вероятностью $1 - q$ — в поток B. Каждый из потоков будет процессом Пуассона, причем поток A — с интенсивностью λq, а поток B — с интенсивностью $\lambda(1 - q)$. Будут ли они независимы? В отличие от процессов Бернулли — да, будут, так как мы находимся в ситуации непрерывного времени.

9 Парадокс наблюдателя

Допустим, вы знаете, что автобусы ходят с интенсивностью 4 автобуса в час. То есть интервал в среднем должен быть 15 минут. Вы приходите на остановку, интересуетесь у продавца в ларьке, как давно ушел последний автобус и ждете следующего. Таким образом вы замеряете интервал между автобусами и уходите. Так вы делаете много дней подряд и обнаруживаете, что средний интервал 30 минут, а не 15. Почему так?

Когда мы пришли на остановку — матожидание времени до следующего автобуса есть 15 минут из-за беспамятства процесса Пуассона. Однако ожидаемое время до предыдущего автобуса — тоже 15 минут, так как мы можем рассматривать автобусы в прошлом как независимый от будущего процесс Пуассона, в котором время течет обратно. В чем подвох?

Подсказка. Пусть есть процесс, в котором события происходят через какие-то интервалы. Но при этом каждый интервал либо 10, либо 5 минут, причем равновероятно. То есть среднее время между событиями есть 7.5 минут. Однако если мы хотим прийти в случайный момент времени и замерить длину интервала, в который мы пришли, то получится, что мы с вероятностью $\frac{2}{3}$ придем в 10-минутный интервал, так как они занимают треть от всей числовой оси. То есть для нас средняя длина отрезка будет $\frac{5}{3} + 203 \approx 8.3$.

Также и с процессом Пуассона: мы намного вероятнее приходим на остановку в длинные интервалы.

Тут также стоит сказать, что подобный эффект может влиять на опросы.

- Если мы опрашиваем случайных людей о размере их семьи, то наш средний результат будет выше, чем реальный размер семьи.

- Если мы опрашиваем случайных людей о заполненности автобуса, на котором они сегодня ехали — средний ответ тоже будет выше среднего значения.