

Теория вероятности и математическая статистика—1

Винер Даниил [@danya_vin](#)

Версия от 24 сентября 2024 г.

Содержание

1	Дискретное вероятностное пространство. Базовые теоремы вероятности	2
1.1	Классическое определение вероятности	2
1.2	Теорема сложения	2
1.3	Условная вероятность	2
1.4	Теорема умножения	2
2	Независимые события. Апостериорная и условная вероятности	3
2.1	Задача об авариях	3
2.2	Независимость событий	3
2.3	Пример зависимых событий в совокупности	3
2.4	Простейший вариант ЗБЧ. Неизбежность технологических катастроф	4
2.5	Формула полной вероятности	4
2.6	Формула Байеса. Апостериорные вероятности	5
2.7	Задача про неизлечимые заболевания	5
3	Схема Бернулли. Теорема Пуассона. Случайные величины	5
3.1	Схема Бернулли. Биномиальное распределение вероятности	5
3.2	Исследование Пуассона	6
3.3	Теорема Пуассона	6
3.4	Случайные величины. Математическое ожидание	7
4	Основные дискретные распределения. Математическое ожидание	7
4.1	Основные дискретные распределения	8
4.2	Свойства линейности мат.ожидания	8

1 Дискретное вероятностное пространство. Базовые теоремы вероятности

Определение. $\Omega = \{\omega_1, \dots, \omega_k, \dots\}$ называется *пространством элементарных исходов*, где ω_i — элементарный исход

Определение. A — любое подмножество Ω

Определение. Событие называется *достоверным*, если $A = \Omega$

Примечание. К A применимы те же операции, что используются с множествами

Определение. Полная группа несовместных событий — такой набор событий, для которого выполняются такие условия:

$$A_i \cap A_j = \emptyset \quad \forall i \neq j$$

$$\bigcup_i A_i = \Omega$$

Аксиома. $\forall \omega_i \exists p_i \geq 0$, при этом $\sum_i p_i = 1$

Следствие. $0 \leq p_i \leq 1$

Определение. $P(A) = \sum_{\omega_i \in A} P(\omega_i)$, где $P(\omega_i) = p_i$

(Ω, P) — вероятностное пространство в дискретном случае

Подходы к определению вероятностей

1. Априорный (предварительное знание)
2. Частотный (предел ряда частот)
3. Модельный (математическая модель)

1.1 Классическое определение вероятности

Имеет место, когда исходы равновероятны

Определение.

$$P(A) = \frac{|A|}{|\Omega|}$$

Определение. $P(A) = \sum_{\omega_i \in A} p(\omega_i)$

1.2 Теорема сложения

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

1.3 Условная вероятность

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \forall B : P(B) > 0$$

1.4 Теорема умножения

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cdot \dots \cdot A_{n-1})$$

2 Независимые события. Апостериорная и условная вероятности

Напомним классическое определение вероятности

$$P(A) = \sum_{\omega_i \in A} p(\omega_i)$$

Примечание. $P(\overline{A}) = 1 - P(A)$, $P(\Omega) = 1$

2.1 Задача об авариях

Известно, что в 40% аварий виноваты пьяные водители. Утверждается, что из этого следует, что в 60% случаев виноваты трезвые водители

Пусть есть такие события: A — водитель пьян, B — водитель трезв, C — авария случилась

Формально, задача выглядит так: $P(A|C) = 0.4 \implies P(B|C) = 0.6$

Пусть $P(B) = 0.05$

$$\frac{P(C|A)}{P(C|B)} = \frac{P(C \cap A)/P(A)}{P(C \cap B)} = \frac{P(A|C)P(C)}{P(B|C)P(C)} \cdot \frac{P(B)}{P(A)} = \frac{0.4}{0.6} \cdot \frac{0.95}{0.05} \approx 12.7$$

2.2 Независимость событий

Определение. (интуитивное) События A и B **независимы**, если $P(A|B) = P(A)$

Определение. События A и B называются попарно независимыми, если:

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A|B)P(B) &= P(A) \cdot P(B) \end{aligned} \text{ — вытекает интуитивное определение}$$

Определение. События A_1, \dots, A_n независимы в совокупности, если:

$$\begin{aligned} \forall i_1 < \dots < i_k < \dots < i_n \quad \forall k = 1, \dots, n : \\ P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) &= P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}) \end{aligned}$$

Примечание. Для A_1, A_2, A_3 :

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) \cdot P(A_2) \\ P(A_2 \cap A_3) &= P(A_2) \cdot P(A_3) \\ P(A_1 \cap A_3) &= P(A_1) \cdot P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2) \cdot P(A_3) \end{aligned}$$

2.3 Пример зависимых событий в совокупности

Положим, что у нас есть тетраэдр, каждая сторона которого покрашена в некоторые комбинации цветов: A — красный, B — желтый, C — зеленый, а четвертая покрашена во все три цвета, тогда вероятности:

$$\frac{1}{4} = P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C) = \frac{1}{8}$$

Таким образом, события зависимы в совокупности

Примечание. Если A_1, \dots, A_n независимы в совокупности, то над любым из событий можно поставить знак отрицания, тогда система останется независимой

Пример. Есть события A_1, A_2, A_3 — независимы в совокупности, тогда $\overline{A_1}, A_2, A_3$ — тоже независимы в совокупности

Проверим данный пример.

$$\begin{aligned}
P(\overline{A_1} \cap A_2 \cap A_3) &= P((A_2 \cap A_3) \setminus A_1) \\
&= P(A_2)P(A_3) - P(A_1 \cap A_2 \cap A_3) \\
&= P(A_2)P(A_3) - P(A_1)P(A_2)P(A_3) \\
&= P(\overline{A_1})P(A_2)P(A_3)
\end{aligned}$$

Теперь разберемся с двумя событиями

$$\begin{aligned}
P(\overline{A_1} \cap A_2) &= P(\overline{A_1})P(A_2) \\
P(\overline{A_2} \cap A_1) &= P(A_2 \setminus A_1) \\
&= P(A_2) - P(A_2 \cap A_1) = P(A_2) - P(A_2)P(A_1) \\
&= P(A_2)(1 - P(A_1))
\end{aligned}$$

2.4 Простейший вариант ЗБЧ. Неизбежность технологических катастроф

Имеется n узлов, а A_1, \dots, A_n — события, где A_i означает, что i -ый узел вышел из строя

Очевидно, что $P(\text{хотя бы один узел выйдет из строя}) = P(A_1 \cup \dots \cup A_n)$, тогда $\forall i : 0 < \varepsilon \leq P(A_i) < 1$ выполняется:

$$1 \geq P(A_1 \cup \dots \cup A_n) = 1 - P(\overline{A_1} \cap \dots \cap \overline{A_n}) = 1 - \prod_{i=1}^n \underbrace{P(\overline{A_i})}_{\leq 1-\varepsilon} \geq \lim_{n \rightarrow \infty} 1 - (1 - \varepsilon)^n = 1$$

Это означает, что вероятность технологических катастроф при количестве узлов, стремящемся в бесконечность, равна 100%

2.5 Формула полной вероятности

Пусть $\{H_i\}$ — полная группа несовместных событий (разбиение Ω)

Некоторые свойства:

- $H_i \cap H_j = \emptyset \quad \forall i \neq j$ — несовместность

- $\bigcup_{i=1}^n H_i = \Omega$ — полнота

Теорема. Тогда, $P(A) = \sum_{i=1}^n P(A|H_i)P(H_i)$

Доказательство.

$$\begin{aligned}
P(A) &= P\left(\bigcup_{i=1}^n (A \cap H_i)\right) \\
&= \sum_{i=1}^n P(A \cap H_i) \\
&= \sum_{i=1}^n P(A|H_i) \cdot P(H_i)
\end{aligned}$$

□

2.6 Формула Байеса. Апостериорные вероятности

$$\begin{aligned} P(H_k|A) &= \frac{P(A|H_k) \cdot P(H_k)}{P(A)} \\ &= \frac{P(A|H_k) \cdot P(H_k)}{\sum_{i=1}^n P(A|H_i)P(H_i)} \end{aligned}$$

2.7 Задача про неизличимые заболевания

Требуется вычислить, какова вероятность, что человек, который получил положительный тест на СПИД, на самом деле не болен

$P(\text{СПИД}) = 0.03$ — вероятность быть носителем СПИДа

$P(+|\text{СПИД}) = 0.98$ — чувствительность теста, т.е. тест будет положительным, если у человека есть СПИД, с вероятностью 0.98

$P(+|\overline{\text{СПИД}}) = 0.01 = 1 - \text{специфичность}$, т.е. тест будет положительным, если у человека нет СПИДа, с вероятностью 0.01

$$\begin{aligned} P(\text{СПИД}|+) &= \frac{P(+|\text{СПИД})P(\text{СПИД})}{P(+|\text{СПИД})P(\text{СПИД}) + P(+|\overline{\text{СПИД}})P(\overline{\text{СПИД}})} \\ &= \frac{0.98 \cdot 0.03}{0.98 \cdot 0.03 + 0.01 \cdot 0.97} \\ &\approx 0.75 \end{aligned}$$

То есть, **каждый четвертый** человек с положительным СПИД-тестом **здоров**

3 Схема Бернулли. Теорема Пуассона. Случайные величины

Рассмотрим задачу проведения социологических опросов. Допустим, проводится опрос «Употребляет ли человек наркотики?». Предполагается, что не все респонденты, даже на условиях анонимности, дадут честный ответ. Для решения этой проблемы иногда задаются косвенные вопросы, мы же рассмотрим следующий подход.

Пусть человек перед ответом берет из урны случайный шарик. В урне лежит N шариков трех цветов: белый, черный и красный. Если человек достал белый, то его ответ на вопрос просто число «0», если черный — «1», если красный — респондент должен сказать правду: употребляет ли он наркотики. Тогда, все ответы разделятся на три категории, которые можно описать, как

$$N = N_{\text{к}} + N_{\text{ч}} + N_{\text{б}}$$

Пусть p — вероятность употребления наркотиков респондентом, тогда по формуле полной вероятности получаем

$$P(1) = P(1|\text{б})P(\text{б}) + P(1|\text{ч})P(\text{ч}) + P(1|\text{к})P(\text{к})$$

$$\text{Отсюда следует, что } p = \frac{P(1)N - N_{\text{ч}}}{N_{\text{к}}}$$

Пример. Если оказалось, что $P(1) = 0.6$, это означает, что $p = \frac{1}{2}$

3.1 Схема Бернулли. Биномиальное распределение вероятности

Допустим, что мы проводим n независимых опытов, каждый из которых может закончиться успехом (1) или неудачей (0). Пусть вероятность успеха $P(1) = p$, тогда $P(0) = 1 - p = q$

Упорядочим результаты экспериментов в последовательность длины n , тогда всего исходов $|\Omega| = 2^n$, то есть последовательности длины n , состоящие из 0 и 1. Заметим, что они **НЕ равновероятны**

Пользуясь тем, что события независимы, получаем, что

$$P(\underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k}) = p^k (1-p)^{n-k}$$

Заметим, что исходов, в которых k успехов, существует C_n^k , при этом

$$P(k \text{ успехов в } n \text{ испытаниях}) = C_n^k p^k (1-p)^{n-k}$$

Определение. Набор чисел $\{C_n^k p^k (1-p)^{n-k} | k = 0, 1, \dots, n\}$ называется **биномиальным распределением вероятности**

3.2 Исследование Пуассона

Французский математик Симеон Дени Пуассон проводил исследование, в котором пытался понять насколько справедливы обвинительные приговоры, которые выносят судьи присяжных во Франции.

Присяжных $n = 12$ человек. Для того, чтобы человек был признан виновным, нужно 7 голосов «ЗА». Пуассон знал лишь процент обвинительных приговоров и количество обвинительных приговоров, вынесенных ровно 7 присяжными заседателями. Пусть p — вероятность верного решения присяжного (неизвестный нам параметр). Верное решение — когда улики достаточно для обвинения — присяжный голосует за *виновность*, не достаточно — за *невиновность*

Пусть есть событие A — улики достаточно для обвинения. Заметим, что $P(A)$ нам неизвестно.

Очевидно, что частота обвинительных приговоров = $P(\text{обвинение})$, тогда

$$P(\text{обвинение}) = P(\text{обвинение} | A)P(A) + P(\text{обвинение} | \bar{A})P(\bar{A})$$

Теперь введем вероятность, что человек был обвинен, когда за это проголосовало ровно 7 присяжных:

$$\begin{aligned} P(\text{обв. } 7) &= P(\text{обв. } 7 | A)P(A) + P(\text{обв. } 7 | \bar{A})P(\bar{A}) \\ &= C_{12}^7 p^7 (1-p)^5 P(A) + C_{12}^5 (1-p)^7 p^5 (1-P(A)) \end{aligned}$$

$$\text{Тогда, } P(\text{обвинение}) = \left(\sum_{k=7}^{12} C_{12}^k p^k (1-p)^{12-k} \right) P(A) + \left(\sum_{k=7}^{12} C_7^{12-k} (1-p)^k p^{12-k} \right) (1-P(A))$$

Пуассон решил, что $p \approx \frac{2}{3}$. Однако, это было ошибочным решением. Математик считал, что присяжные выносят решения независимо друг от друга, что в корне неверно. А так как Пуассон опирался на схему Бернулли, которая работает только для *независимых событий*, то и результат его исследования получился ложным

При этом, Пуассона не смутило, что в Англии для вынесения обвинительного приговора нужно, чтобы все 12 присяжных проголосовали за обвинение. Если бы решение Пуассона было верным, тогда вероятность обвинительных приговоров должна быть $\approx \left(\frac{2}{3}\right)^{12}$, что не соответствовало действительности, ведь в Англии было больше обвинительных приговоров чем во Франции

3.3 Теорема Пуассона

Теорема. Пусть имеются серии испытаний. В каждой серии n испытаний и p_n — вероятность успеха. Также, пусть $n \rightarrow \infty$ и $p_n \rightarrow 0$, причем $np_n \rightarrow \lambda$

Тогда,

$$\forall k \ P(k \text{ успехов в } n \text{ испытаниях}) = C_n^k p_n^k (1-p_n)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

Доказательство. Пусть μ_n — число успехов в n испытаниях Бернулли с вероятностью успеха p_n , а также $\lambda_n = n \cdot p_n$. По условию, $\lambda_n^k \rightarrow \lambda > 0$. Подставим $p_n = \frac{\lambda_n}{n}$ в формулу Бернулли. Тогда,

$$\begin{aligned}
P(\mu_n = m) &= \frac{n!}{m!(n-m)!} p_n^m (1-p_n)^{n-m} \\
P(\mu_n = m) &= \frac{n \cdot \dots \cdot (n-m+1)}{n^k} \left(\frac{\lambda_n}{n}\right)^m \frac{(1 - \frac{\lambda_n}{n})^n}{(1 - \frac{\lambda_n}{n})^m} \\
&= \frac{n \cdot \dots \cdot (n-m+1)}{n^k} \left(\frac{\lambda_n}{n}\right)^m \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \frac{1}{(1 - \frac{\lambda_n}{n})^m} \\
&= \underbrace{\frac{n \cdot \dots \cdot (n-m+1)}{n^k}}_{\lim_{n \rightarrow \infty} = 1} \underbrace{\frac{\lambda_n^m}{m!} \left(1 - \frac{\lambda_n}{n}\right)^n}_{e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-m}}_{\lim_{n \rightarrow \infty} = 1} \\
&= \frac{\lambda^k}{k!} e^{-\lambda}
\end{aligned}$$

Определение. Набор чисел $\left\{ \frac{\lambda^k}{k!} e^{-\lambda} : k = 0, 1, \dots, n \right\}$ называется **распределением Пуассона**

Примечание. Величина $\frac{\lambda^k}{k!} e^{-\lambda}$ устойчива при зависимых событиях, а также при ошибках предположений (таких, какие допустил Пуассон)

Примечание. $\forall B$ — множество значений, которые может принимать число успехов

$$\left| P(\mu \in B) - \sum_{m \in B} e^{-\lambda} \frac{\lambda^m}{m!} \right| \leq \min(p, np^2)$$

3.4 Случайные величины. Математическое ожидание

Определение. Случайная величина $\xi : \Omega \rightarrow \mathbb{R}$

ξ имеет биномиальное распределение, если

$$\xi \in \{0, 1, \dots, n\} \quad P(\xi = k) = C_n^k p^k (1-p)^{n-k},$$

где $\{0, 1, \dots, n\}$ — число успехов в n испытаниях

$$P(\xi \in B) = \sum_{\omega} P(\omega : \xi(\omega) \in B)$$

Определение. Математическим ожиданием величины ξ называется величина

$$\mathbb{E}(\xi) = \sum_{\omega} P(\xi) \cdot \xi(\omega),$$

если этот ряд сходится абсолютно, то есть $\sum |\xi(\omega)| P(\omega)$

Теорема. $\mathbb{E}(\xi) = \sum_k a_k P(\xi = a_k)$

4 Основные дискретные распределения. Математическое ожидание

Напомним определение случайной величины — $\xi : \Omega \rightarrow \mathbb{R}$

Определение. Дискретное вероятностное пространство — $(\Omega, P) \rightarrow (\mathbb{R}, P(\xi))$, где $P\xi : P(\xi = a) \forall a \in \mathbb{R}$

Если $\xi \in a_1 \dots a_k$, а вероятности равны $p_1 \dots p_k$, где $p_i = P(\xi = a_i)$, тогда

$$P(\xi = a) = P\left(\underbrace{\{\omega : \xi(\omega) = a_i\}}_{A\text{-образ } \{\xi=a_i\}}\right) \text{ — распределение вероятностей случайной величины } \xi$$

Пример. $\xi : 0, 1, \dots, n, p : P(\xi = k) = C_n^k p^k (1-p)^{n-k}$ — биномиальное распределение

Обозначение: $Bi(n, p)$

Определение. $\mathbb{E}(\xi) = \sum_{\omega \in \Omega} \xi(\omega)P(\omega)$ — если ряд сходится абсолютно

Если ряд не сходится, то математического ожидания **не существует**

4.1 Основные дискретные распределения

1. Вырожденное распределение. $\exists c : P(\xi = c) = 1, c \in \mathbb{R}$
2. Распределение Бернулли: $Bi(1, p)$. $\xi \in [0, 1]$ с вероятностями p и $1 - p$ соответственно

Индикатор события $A - I\{A\} = \begin{cases} 1, & \text{если } A \\ 0, & \text{если } \bar{A} \end{cases}$

Пример. 10 кубиков, A — хотя бы одна 6

$$I\{A\} = \begin{cases} 1, & P = 1 - \left(\frac{5}{6}\right)^{10} \\ 0, & P = \left(\frac{5}{6}\right)^{10} \end{cases}$$

3. Биномиальное распределение $Bi(n, p)$, где μ_n — число успехов в n испытаниях
 $P(\mu_n = k) = C_n^k p^k (1 - p)^{n-k}, k = \overline{0, n}$

$$\mathbb{E}(\mu_n) = \sum_{k=0}^n k C_n^k p^k (1 - p)^{n-k}$$

4. Распределение Пуассона: $\Pi(\lambda)$

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k \in \mathbb{Z}$$

Иногда это распределение называют *распределением редких событий*

$$\mathbb{E}(\xi) = \lambda \quad (\lambda = np), \text{ где } \mathbb{E}(\xi) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

5. Геометрическое распределение

Пример. Число неудач до первого успеха = $\xi - 1$ **или** номер первого успеха = ξ

При этом, p — вероятность успеха

$$\xi : 1, \dots \quad P(\xi = k) = (1 - p)^{k-1} p$$

$$\mathbb{E}(\xi) = \sum_{k=1}^{\infty} k (1 - p)^{k-1} p$$

Пусть $\mu = \mathbb{E}(\xi)$ — мат.ожидание номера первого успеха

$$\text{Тогда, } \mu = 1 \cdot p + (1 - p)(1 + \mu) \implies \mu = \frac{1}{p}$$

6. Отрицательное биномиальное распределение: $\overline{Bi}(r, p)$. Номер r -го успеха, k -е место

$$P(\xi = k) = C_{k-1}^{r-1} p^{r-1} (1 - p)^{k-r} \cdot p$$

7. Гипергеометрическое распределение

$$P(\xi = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_n^k}$$

4.2 Свойства линейности мат.ожидания

Пусть $\mu_n = \sum_{i=1}^n I\{A_i\}$, где A_i — в i -ом опыте был успех

$$\mathbb{E}(I\{A_i\}) = P(A_i) = p, \text{ тогда, } \mathbb{E}(\mu_n) = np$$

$$np - q \leq (\text{наиболее вероятное число успехов}) \leq np + p$$