Лекции по предмету
Теория вероятностей

Денис Беляков
Никита Попов
Алексей Хачиянц

2016/2017 учебный год

1 Лекция от 23.09.2016

Предполагаем, что простанство элементарных событий Ω дискретно, т.е. содержит не более, чем счетное количество элементов. Тогда случайная величина ξ принимает не более, чем счетное множество значений.

Пусть $X = (x_1, x_2, \ldots, x_n)$ — множество значений случайной величины ξ.

Введем следующие события:

- $A_i = \{\omega : \xi(\omega) = x_i\}$ — событие "ξ приняло значение x_i".

  Для удобства будем использовать обозначение $A_i := \{\xi = x_i\}$.

- Также введем обозначение p_i для вероятности события A_i.

  $p_i = \mathsf{P}(A_i) = \mathsf{P}(\xi = x_i)$.

Определение 1. Вместе множество значений $X = (x_1, x_2, \ldots, x_n)$ и набор вероятностей $(p_1, p_2, \ldots, p_n)$ образуют то, что называется *распределением случайной величины* ξ.

Каждому x_i сопоставлено число p_i. При этом для i, таких что $1 \leqslant i \leqslant n$, вместе A_i образуют разбиение вероятностного пространства Ω. А значит $\sum\limits_i p_i = 1$.

Приведем примеры:

1. **Распределение Бернулли.**

   Случайная величина X имеет *распределение Бернулли*, если она принимает всего два значения, 1 или 0, с заранее известными вероятностями p и q.

   Множество значений случайной величины ξ — $X = \{0, 1\}$; $\mathsf{P}(x_i = 1) = p$, $\mathsf{P}(x_i = 0) = q$.

   Принято говорить, что событие $x_i = 1$ соответствует "успеху", а событие $x_i = 0$ — "неудаче". Эти названия достаточно условные, и в зависимости от конкретной задачи могут быть заменены на противоположные.

   $\xi \sim \mathrm{Bern}(p)$.

2. **Биномиальное распределение.**

   Биномиальное распределение в теории вероятностей — распределение количества “успехов” в последовательности из n независимых случайных экспериментов, таких, что вероятность “успеха” в каждом из них постоянна и равна p.

   Пусть $x_1, \dots, x_n$ — последовательность независимых случайных величин с одинаковым распределением Бернулли. Тогда при $n \in \mathbb{N}$ и $p \in [0,1]$, $(q = 1-p)$ вероятность принятия случайной величиной значения x_k равна:

   $p_k = \mathsf{P}(\xi = x_k) = C_n^k p^k q^{n-k}$.

   Случайная величина ξ называется биномиальной.
   $\xi \sim \mathrm{Bin}(n,p)$.

3. **Пуассоновское распределение.**

   Обозначение: $\mathrm{Pois}(\lambda)$, $\lambda > 0$.

   $X = \mathbb{Z}_+ = \mathbb{N} \cup 0$[1], $p_k = \mathsf{P}(\xi = x_k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

   $\xi \sim \mathrm{Pois}(\lambda)$ — пуассоновская случайная величина.

4. **Геометрическое распределение.**

   Обозначение: $\mathrm{Geom}(p), p \in (0,1)$.

   $p_k = \mathsf{P}(\xi = k) = \mathsf{P}(1-p)^{k-1}$, $k \in \mathbb{N}$.

   $\xi \sim \mathrm{Geom}(p)$ — геометрическая случайная величина.

Определение 2. Пусть ξ, η — случайные величины. Пусть ξ принимает значения из (возможно, бесконечного) множества $(a_1, a_2, \dots, a_n, \dots)$, а η — значения $(b_1, b_2, \dots, b_n, \dots)$ [2]. Тогда ξ и η называются *независимыми*, если для любой пары $i, j \in \mathbb{N}$ события $\{\xi = a_i\}$ и $\{\eta = b_j\}$ являются независимыми.

То есть вероятность того, что одновременно произошли оба события распадается в произведение их вероятностей.

$$\mathsf{P}(\{\xi = a_i\} \cap \{\eta = b_j\}) = \mathsf{P}(\{\xi = a_i\}) \cdot \mathsf{P}(\{\eta = b_j\}).$$

Примечание 1. Но для удобства написания вероятности независимых событий используют следующее обозначение.

$$\mathsf{P}(\{\xi = a_i\} \cap \{\eta = b_j\}) \sim \mathsf{P}(\xi = a_i, \eta = b_j).$$

Определение 3. Пусть $\xi_1, \dots, \xi_n$ — случайные величины. ξ_i принимает значения ($a_1^{(i)}$, $a_2^{(i)}$,…,$a_n^{(i)}$). Тогда будем говорить, что случайные величины $\xi_1, \dots, \xi_n$ *независимы (в совокупности)*, если $\forall j_1, \dots, j_n$ выполнено:

$$\mathsf{P}(\xi_1 = a_{j_1}^{(1)}, \dots, \xi_n = a_{j_n}^{(n)}) = \prod_{k=1}^{n} \mathsf{P}(\xi_k = a_{j_k}^{(k)}).$$

Задача 1. *A, B — события. Показать, что A и B независимы тогда и только тогда, когда их индикаторы I_A и I_B независимы.*

[1] Есть путаница, откуда считать $\mathbb{Z}_+$ или $\mathbb{N}$. Мы тоже будем путаться, но все же пусть $X = \mathbb{N}$.

[2] Далее множество значений случайной величины ξ будем обозначать как $\xi(\Omega)$

Задача 2. *Показать, что случайные величины $\xi_1, \ldots, \xi_n$ независимы тогда и только тогда, когда*

$$\forall x_1, \ldots, x_n \in \mathbb{R},\ \mathsf{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) = \prod_{k=1}^{n} \mathsf{P}(\xi_k = x_k).$$

2 Математическое ожидание

Определение 4. *Математическим ожиданием* случайной величины ξ называют величину

$$E[\xi] = \sum_{\omega \in \Omega} \xi(\omega)\mathsf{P}(\omega)$$

Примечание 2. Если Ω счётно, то ряд $\sum \xi(\omega)P(\omega)$ должен сходиться абсолютно; иначе сумма ряда либо равна бесконечности, либо не определена т.к. порядок перебора ω не задан (см. теорему Римана о перестановке членов условно сходящегося ряда).

Смысл определения: Математическое ожидание случайной величины можно понимать как среднее значение этой случайной величины.

Лемма (Свойства математического ожидания)**.**

1. *Линейность:*
$$E[a\xi + b\eta] = aE\xi + bE\eta$$

2. *Сохранение относительного порядка:*
$$\xi \leqslant \eta \implies E[\xi] \leqslant E[\eta]$$

3. *Модуль математического ожидания меньше математического ожидания модуля:*
$$|E[\xi]| \leqslant E\left[|\xi|\right]$$

4. *Если ξ принимает значения в $X = \{x_1, x_2, \ldots\}$ (возможно бесконечном), то*
$$E[\xi] = \sum_i x_i \mathsf{P}(\xi = x_i)$$

5. *Если ξ принимает значения в $X = \{x_1, x_2, \ldots\}$ (возможно бесконечном), то для любой функции $\varphi(y)$ выполняется*
$$E[\varphi(\xi)] = \sum_i \varphi(x_i)\mathsf{P}(\xi = x_i)$$

6. *Если $\mathsf{P}(\xi = c) = 1$ для некоторой константы c, то $E[\xi] = c$;*

7. *Если $\xi \geqslant 0$, то $E[\xi] \geqslant 0$; также, если $E[\xi] = 0$, то $\mathsf{P}(\xi = 0) = 1$;*

8. *Если ξ и η независимы, то $E[\xi\eta] = E[\xi]E[\eta]$*

Доказательство.

1. Линейность:

$$E[a\xi + b\eta] = \sum_{\omega}(a\xi + b\eta)(\omega)\mathsf{P}(\omega) = \sum_{\omega} a\xi(\omega)\mathsf{P}(\omega) + \sum_{\omega} b\eta(\omega)\mathsf{P}(\omega) = aE[\xi] + bE[\eta]$$

2. Сохранение относительного порядка:

$$E[\xi] = \sum_{\omega\in\Omega}\xi(\omega)\mathsf{P}(\omega) \leqslant \{\forall\omega \implies \xi(\omega) \leqslant \eta(\omega)\} \leqslant \sum_{\omega\in\Omega}\eta(\omega)\mathsf{P}(\omega) \leqslant E[\eta]$$

3. Модуль математического ожидания меньше математического ожидания модуля:

$$E[-|\xi|] \leqslant E[\xi] \leqslant E[|\xi|]$$

$$-E[|\xi|] \leqslant E[\xi] \leqslant E[|\xi|]$$

$$|E[\xi]| \leqslant E[|\xi|]$$

4. Если ξ принимает значения в $X = \{x_1, x_2, \ldots\}$ (возможно бесконечном), то

$$E[\xi] = \sum_{\omega\in\Omega}\xi(\omega)\mathsf{P}(\omega) = \sum_{i}\sum_{w:\xi(w)=x_i}\xi(\omega)\mathsf{P}(\omega) = \sum_{i} x_i \sum_{w:\xi(w)=x_i}\mathsf{P}(\omega) = \sum_{i} x_i\mathsf{P}(\xi = x_i)$$

5. Если ξ принимает значения в $X = \{x_1, x_2, \ldots\}$ (возможно бесконечном), то для любой функции $\varphi(y)$ выполняется

$$E[\varphi(\xi)] = \sum_{\omega\in\Omega}\varphi(\xi(\omega))\mathsf{P}(\omega) = \sum_{i}\sum_{w:\xi(w)=x_i}\varphi(\xi(\omega))\mathsf{P}(\omega) =$$
$$= \sum_{i}\varphi(x_i)\sum_{w:\xi(w)=x_i}\mathsf{P}(\omega) = \sum_{i}\varphi(x_i)\mathsf{P}(\xi = x_i)$$

6. Если $P(\xi = c) = 1$ для некоторой константы c, то $E[\xi] = c$;

$$E[\xi] = \sum_{\omega\in\Omega}\xi(\omega)\mathsf{P}(\omega) = \sum_{\omega:\xi(\omega)=c}\xi(\omega)\mathsf{P}(\omega) + \sum_{\omega:\xi(\omega)\neq c}\xi(\omega)\mathsf{P}(\omega) = c\mathsf{P}(\xi = c) + \sum_{\omega:\xi(\omega)\neq c}\xi(\omega)\cdot 0 = c$$

7. Если $\xi \geqslant 0$, то $E[\xi] \geqslant 0$; также, если $E[\xi] = 0$, то $\mathsf{P}(\xi = 0) = 1$;

$$\xi \geqslant 0 \implies E[\xi] \geqslant E[0] \geqslant 0;$$

$$E[\xi] = \sum \underbrace{\xi(\omega)}_{\geqslant 0}\underbrace{\mathsf{P}(\omega)}_{\geqslant 0} = 0$$

   Значит, для каждого ω либо $\mathsf{P}(\omega)$, либо $\xi(\omega)$ равно нулю; значит, ненулевая вероятность возможна только для $\xi = 0$, а так как сумма вероятностей — 1, $\mathsf{P}(\xi = 0) = 1$.

8. Если ξ и η независимы, то $E[\xi\eta] = E[\xi]E[\eta]$

$$E[\xi\eta] = \sum_{\omega} \xi(\omega)\eta(\omega)\mathsf{P}(\omega) = \left\{ \begin{matrix} \xi(\Omega) = (a_1, a_2, \dots) \\ \eta(\Omega) = (b_1, b_2, \dots) \end{matrix} \right\} = \sum_{i,j} \sum_{\substack{\omega: \\ \xi(\omega)=a_i \\ \eta(\omega)=b_j}} \xi(\omega)\eta(\omega)\mathsf{P}(\omega) =$$

$$= \sum_{i,j} a_i b_j \mathsf{P}(\xi = a_i; \eta = b_j) = \{\text{т.к. } \xi \perp\!\!\!\perp \eta\} = \sum_{i,j} a_i b_j \mathsf{P}(\xi = a_i)\mathsf{P}(\eta = b_j) =$$

$$= \left(\sum_i a_i \mathsf{P}(\xi = a_i) \right) \cdot \left(\sum_j b_j \mathsf{P}(\eta = b_j) \right) = E[\xi] \cdot E[\eta]$$

□

Примеры:

1. Пусть I_A — индикатор некоторого события A. Тогда
$$E[I_A] = 1 \cdot \mathsf{P}(A) + 0 \cdot \mathsf{P}(\overline{A}) = \mathsf{P}(A).$$

2. Если мы рассмотрим классическую модель, т.е. такую модель, где все исходы равновероятны, то
$$E[\xi] = \frac{1}{|\Omega|} \sum_{\omega} \xi(\omega)$$
Иначе говоря, математическое ожидание в классической модели равно среднему арифметическому возможных значений ξ.

3. Пусть $\xi \sim Bin(n, p)$. Тогда
$$E[\xi] = \sum_{k=0}^{n} k\mathsf{P}(\xi = k) = \sum_{k=1}^{n} kC_n^k p^k (1-p)^{n-k} =$$
$$= \sum_{k=1}^{n} nC_{n-1}^{k-1} p^k (1-p)^{n-k} = np \sum_{t=0}^{n-1} C_{n-1}^{t} p^t (1-p)^{n-1-t}$$

Определение 5. Пусть ξ — некоторая случайная величина. Тогда *дисперсией* ξ называется $D[\xi] = E[(\xi - E\xi)^2]$.

Смысл определения: Дисперсию случайной величины можно понимать как среднеквадратическое отклонение этой случайной величины от её среднего значения (математического ожидания) [3] .

Определение 6. Пусть ξ и η — две случайные величины. Тогда *ковариацией* этих величин называется $\operatorname{cov}(\xi, \eta) = E\left[(\xi - E[\xi])(\eta - E[\eta])\right]$.

Смысл определения: Ковариацию стоит воспринимать как меру зависимости двух случайных величин.

[3] Вероятность незначительного отклонения величины от математического ожидания — т.е. $\mathsf{P}(E[\xi] - 2\sqrt{D[\xi]} \leqslant \xi \leqslant E[\xi] + 2\sqrt{D[\xi]})$ — всегда велика! Это нам пока не пригодится, но на будущее стоит запомнить.

Определение 7. Две случайные величины ξ и η называют *некоррелированными*, если $\operatorname{cov}(\xi, \eta) = 0$.

Лемма (Свойства дисперсии и ковариации)**.**

1. *Ковариация билинейна:*

$$\operatorname{cov}(\xi, a\eta + b\chi) = a \cdot \operatorname{cov}(\xi, \eta) + b \cdot \operatorname{cov}(\xi, \chi)$$

2. *Выражение дисперсии через ковариацию:*

$$D[\xi] = \operatorname{cov}(\xi, \xi)$$

3. *Влияние констант на дисперсию:*

$$\begin{array}{ll} a) & D[c\xi] = c^2 D[\xi] \\ b) & D[\xi + c] = D[\xi] \end{array}$$

4. *Неотрицательность дисперсии:*

$$\forall \xi \implies D[\xi] \geqslant 0;$$

$$D[\xi] = 0 \iff \mathsf{P}(\xi = E[\xi]) = 1;$$

5. *Если ξ и η некоррелированные, то*

$$D[\xi + \eta] = D[\xi] + D[\eta]$$

6. *Связь с математическим ожиданием:*

$$D[\xi] = E[\xi^2] - (E[\xi])^2$$

$$\operatorname{cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta]$$

Доказательство.

1. Ковариация билинейна:

   Вытекает из свойств математического ожидания.

2. Выражение дисперсии через ковариацию:

$$\operatorname{cov}(\xi, \xi) = E\left[(\xi - E[\xi])(\xi - E[\xi])\right] = E\left[(\xi - E[\xi])^2\right] = D[\xi]$$

3. Влияние констант на дисперсию:

$$\begin{array}{ll} a) & D[c \cdot \xi] = \operatorname{cov}(c \cdot \xi, c \cdot \xi) = c^2 \operatorname{cov}(\xi, \xi) = c^2 D[\xi] \\ b) & \xi - E[\xi] = \xi + c - E[\xi + c] \Rightarrow D[\xi + c] = D[\xi] \end{array}$$

4. Неотрицательность дисперсии:

$$D[\xi] \geqslant 0, \text{так как } D[\xi] = E[\xi - E[\xi]]^2, \text{ а } (\xi - E[\xi])^2 \geqslant 0$$

5. Дисперсия суммы некоррелированных величин:

   Доказывается обычным честным подсчетом. [4]

$$D[\xi+\eta] = \{\textit{см. второе свойство}\} = \operatorname{cov}(\xi+\eta,\xi+\eta) = \\ = \operatorname{cov}(\xi,\xi) + \operatorname{cov}(\eta,\eta) + 2\cdot\operatorname{cov}(\xi,\eta) = \\ = \{\textit{т.к. величины некоррелированы, } 2\cdot\operatorname{cov}(\xi,\eta) = 0\} = D[\xi] + D[\eta]$$

6. Связь с математическим ожиданием:

$$D[\xi] = E[(\xi - E[\xi])^2] = E[\xi^2 - 2\xi\cdot E[\xi] + (E[\xi])^2] = \\ = \{\textit{по линейности}\} = E[\xi^2] - 2E[\xi]\cdot E[\xi] + (E[\xi])^2 = E[\xi^2] - (E[\xi])^2$$

□

Примечание 3. Если случайные величины независимы, то они *некоррелируемы*. Обратное, вообще говоря, неверно!

Доказательство.

$\Rightarrow$

$$\operatorname{cov}(\xi,\eta) = \{\textit{по 6-му свойству ковариации}\} = E[\xi\eta] - E[\xi]\cdot E[\eta] = \\ = \{\textit{по 8-му свойству математического ожидания}\} = E[\xi]\cdot E[\eta] - E[\xi]\cdot E[\eta] = 0$$

$\not\Leftarrow$ **Контрпример:**

Пусть случайная величина ξ равновероятно принимает значения из множества $\{0, 1, -1\}$. Возьмем случайную величину $\eta = \xi^2$. По определению можно проверить, что величины η и ξ *некоррелируемы*:

$$\operatorname{cov}(\xi,\eta) = E[\xi\eta] - E[\xi]\cdot E[\eta] = E[\xi^3] - E[\xi]\cdot E[\xi^2] = 0$$

Но ξ и η *не являются независимыми*, что проверяется опять же по определению:

$$\mathsf{P}(\xi = 0, \eta = 0) = \frac{1}{3} \neq \frac{1}{9} = \mathsf{P}(\xi = 0)\cdot\mathsf{P}(\eta = 0)$$

□

Под конец лекции обсудим два неравенства, которые сами по себе являются весьма полезными.

1. **Неравенство Маркова.**

   Пусть $\xi \geqslant 0$ — неотрицательная случайная величина. Тогда:

$$\forall a > 0 \implies \mathsf{P}(\xi \geqslant a) \leqslant \frac{E[\xi]}{a}$$

[4]Вообще, при подсчете дисперсии суммы очень удобно переходить к ковариации.

Доказательство.

$$\mathsf{P}(\xi \geqslant a) = E[I\{\xi \geqslant a\}] =$$
$$= \{\textit{заметим, что } I\{\xi \geqslant a\} \leqslant \frac{\xi}{a} \cdot I\{\xi \geqslant a\}\} \leqslant$$
$$\leqslant E[\frac{\xi}{a} \cdot I\{\xi \geqslant a\}] \leqslant E[\frac{\xi}{a}] = \frac{E[\xi]}{a}$$

□

2. **Неравенство Чебышева.**

   Пусть ξ — случайная величина, $D[\xi] < +\infty$ (дисперсия принимает конечное значение). Тогда:
$$\forall a > 0 \implies \mathsf{P}(|\ \xi - E[\xi]\ | \geqslant a) \leqslant \frac{D[\xi]}{a^2}$$

   Доказательство.

$$\mathsf{P}(|\xi - E[\xi]| \geqslant a) = \mathsf{P}((\xi - E[\xi])^2 \geqslant a^2) \leqslant$$
$$\leqslant \{\textit{неравенство Маркова}\} \leqslant \frac{E[\xi - E[\xi]]^2}{a^2} = \frac{D[\xi]}{a^2}$$

□