ЛИКБЕЗ
Лекция 2: Основы теории вероятностей

Дмитрий Ицыксон

ПОМИ РАН

28 сентября 2008

План

- Вероятностное пространство;
- Случайные величины;
- Averaging argument (неравенство Маркова);
- Линейность математического ожидания;
- Условные вероятности и независимость случайных величин;
- Дисперсия, неравенство Чебышева;
- Оценки Чернова;
- Марковская цепь.

Литература

1. А. А. Боровков. Теория вероятностей.
2. В. Феллер. Введение в теорию вероятностей и ее приложения.
3. Н. Алон, Дж. Спенсер. Вероятностный метод.
4. А. Шень. Вероятность: примеры и задачи.

σ-алгебра

Определение. Ω — некоторое множество, $\mathcal{A} \subseteq 2^{\Omega}$. $\mathcal{A}$ называется σ-алгеброй, если:

- $\Omega \in \mathcal{A}$
- Для последовательности $\{A_n\}_{n\in\mathbb{N}}$, если $\forall n \in \mathbb{N}, A_n \in \mathcal{A}$, то $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$ и $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.
- $A \in \mathcal{A} \iff \overline{A} = \Omega \setminus A \in \mathcal{A}$.

Определение. Ω — некоторое множество, $\mathcal{A}$ — σ-алгебра. Вероятностной мерой на $\mathcal{A}$ называется отображение $p : \mathcal{A} \to [0, 1]$:

- $p(\Omega) = 1$;
- Для последовательности попарно непересекающихся множеств $\{A_n\}_{n\in\mathbb{N}}$ выполняется $p(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$.

Вероятностное пространство

Определение. Вероятностным пространством называется тройка $(\Omega, \mathcal{A}, p)$, где

- Ω — пространство элементарных событий (некоторое множество);
- $\mathcal{A} \subseteq 2^{\Omega}$ — множество допустимых событий (σ-алгебра);
- p — вероятностная мера.

Примеры

Пример. Дискретное вероятностное пространство.
$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. $\mathcal{A} = 2^\Omega$, $p_1, p_2, \ldots, p_n \geq 0$, $\sum_{i=1}^n p_i = 1$,
$p(\omega_i) = p_i$.
$p(A) = \sum_{\omega \in A} p(\omega)$.
Замечание. Пересечение σ-алгебр - это σ-алгебра.
Пример. $\Omega = \mathbb{R}$, $\mathcal{A}$ — пересечение всех σ-алгебр, содержащих все открытые множества на $\mathbb{R}$ (борелевская σ-алгебра $\mathcal{B}$).
$\rho : \mathbb{R} \to \mathbb{R}_+$, $\int_{-\infty}^{+\infty} \rho(x)dx = 1$.
$p(A) = \int_A \rho(x)dx$.

Простейшие свойства вероятности

- $p(\emptyset) = 0$, $p(\Omega) = 1$;
- $p(A_1 \cup A_2 \cup \dots \cup A_n) \le \sum_{i=1}^{n} p(A_n)$;
- (формула включений-исключений) $p(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^{n} p(A_n) - \sum_{i<j} p(A_i A_j) + \dots + (-1)^n p(A_1 A_2 \dots A_n)$
- $p(A_1 \cup A_2 \cup \dots \cup A_n) \ge \sum_{i=1}^{n} p(A_n) - \sum_{i<j} p(A_i A_j)$

Random subsum principle

Определение. Для $x, y \in \{0,1\}^n$ определим $< x, y >= \sum_{i=1}^{n} x_i y_i (mod 2)$.
Лемма. Рассмотрим $\Omega = \{0,1\}^n, \mathcal{A} = 2^{\Omega}$, $p(x) = 2^{-n}$ для всех $x \in \Omega$. Известно, что $y \neq 0^n$. Пусть $A = \{x| < x, y >= 1\}$, тогда $p(A) = \frac{1}{2}$ (иначе $\Pr_{x \in \{0,1\}^n}\{< x, y >= 1\} = \frac{1}{2}$).
Доказательство. Пусть $y_k = 1$, каждому $x \in \Omega$ можно сопоставить $x^{(k)}$, у которого k-й элемент инвертирован, $< y, x >= 1- < y, x^{(k)} >$.

Случайная величина

Определение. $(\Omega, \mathcal{A}, p)$ — вероятностное пространство. Случайной величиной называется такое отображение $\xi : \Omega \to \mathbb{R}$, что для всех $A \in \mathcal{B}$ выполняется $\xi^{-1}(A) \in \mathcal{A}$.

- Если $\mathcal{A} = 2^{\Omega}$, то любое отображение $\xi : \Omega \to \mathbb{R}$ является случайной величиной.
- Случайная величина индуцирует вероятностную меру μ на $\mathbb{R}, \mathcal{B}$: $\mu(A) = p(\xi^{-1}(A))$.
- Зная меру μ можно "забыть"про вероятностное пространство. $\Pr\{\xi \in A\} = \mu(A)$.

Дискретное распределение

- $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$,
  $\xi(\omega_1) = a_1, \xi(\omega_2) = a_2, \ldots, \xi(\omega_n) = a_n$.
- $\Pr\{\xi \in A\} = \sum_{i : a_i \in A} \mu(a_i)$.

Абсолютно непрерывное распределение

Определение. Распределение называется абсолютно непрерывным, если существует такая функция $\rho : \mathbb{R} \to \mathbb{R}_+$, $\int_{-\infty}^{+\infty} \rho(x)dx = 1$, которая задает меру μ по формуле $\mu(A) = \int_A \rho(x)dx$. Функция ρ называется плотностью распределения.

Пример. $U(a; b)$ — равномерное распределение на $[a; b]$.

$$\rho(x) = \begin{cases} \dfrac{1}{b-a}, x \in [a; b] \\ 0, x \notin [a; b] \end{cases}$$

Пример. $N(\mu, \sigma^2)$ — нормальное распределение.

$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Математическое ожидание

Определение. ξ — случайная величина, μ — мера, индуцированная ξ. Математическим ожиданием ξ называется $\mathsf{E}\,\xi = \int_\Omega \xi(\omega)dp(\omega) = \int_{\mathbb{R}} xd\mu$.

- Если μ абсолютно непрерывна с плотностью $\rho(x)$, то $\mathsf{E}\,\xi = \int_{-\infty}^{+\infty} x\rho(x)dx$.
- Если μ дискретная мера, при которой $\mu(A_1) = p_1, \ldots, \mu(A_n) = p_n$, то $\mathsf{E}\,\xi = \sum_{\omega\in\Omega} \xi(\omega)p(\omega) = \sum_{i=1}^{n} p_i A_i$.

Averaging argument

- $a_1, a_2, \ldots, a_n$ — некоторые числа, их среднее арифметическое c, тогда существует $a_k \geq c$.
- ξ — случайная величина, $\mathrm{E}\,\xi = m$, тогда $\Pr\{\xi \geq m\} > 0$.
  Доказательство. Пусть ξ принимает значения A_1 с вероятностью p_1, A_2 с вероятностью p_2,..., A_n с вероятностью p_n, где $p_i > 0, \sum p_i = 1$.
  Если все $A_i < m$, то
  $m = p_1 A_1 + p_2 A_2 + \cdots + p_n A_n < p_1 m + p_2 m + \cdots + p_n m = m$.
  Противоречие!

Неравенство Маркова

Лемма. ξ — это неотрицательная случайная величина. Тогда для всех $k > 0$ выполняется неравенство $\Pr\{\xi \geq k \operatorname{E} \xi\} \leq \frac{1}{k}$.
Доказательство. Обозначим $m = \operatorname{E} \xi$ Пусть
$A_1 \leq A_2 \cdots \leq A_i < mk \leq A_{i+1} \leq \ldots A_n$.
$\Pr\{\xi \geq km\} = p_{i+1} + \cdots + p_n > \frac{1}{k}$. Тогда
$m = p_1 A_1 + p_2 A_2 + \cdots + p_n A_n \geq p_{i+1} A_{i+1} + \cdots + p_n A_n \geq mk(p_{i+1} + \cdots + p_n) > m$. Противоречие!
Пример. В лотерее на выигрыши уходит 40% стоимости билетов. Билет стоит 100 рублей. Докажите, что вероятность выиграть хотя бы 5000 не более 1%

- Мат. ожидание выигрыша 40 рублей.
- $\frac{40}{5000} < 1\%$

Оценка сверху

Лемма. $\xi \in [0;1]$, $m = \mathrm{E}\,\xi$. Тогда для всех $0 < c < 1$ выполняется неравенство: $\Pr\{\xi \le cm\} \le \frac{1-m}{1-cm}$.
Доказательство. Пусть $\eta = 1 - \xi$. $\mathrm{E}\,\eta = 1 - m$.Тогда

$$\Pr\{\xi \le cm\} = \Pr\{1 - \eta \le cm\} = \Pr\{\eta \ge 1 - cm\} =$$

$$\Pr\{\eta \ge \frac{1-cm}{1-m}(1-m)\} \overset{\text{н-во Маркова}}{\le} \frac{1-m}{1-cm}.$$

Линейность математического ожидания

Теорема. $\xi = \alpha\xi_1 + \beta\xi_2$, где $\alpha, \beta \in \mathbb{R}$. Тогда $\mathsf{E}\,\xi = \alpha\,\mathsf{E}\,\xi_1 + \beta\,\mathsf{E}\,\xi_2$.
Доказательство. $\mathsf{E}\,\xi = \sum_{\omega\in\Omega}\xi(\omega)p(\omega) =$
$\sum_{\omega\in\Omega}(\alpha\xi_1(\omega) + \beta\xi_2(\omega))p(\omega) = \alpha\,\mathsf{E}\,\xi_1 + \beta\,\mathsf{E}\,\xi_2$.

Турнир с большим числом гамильтоновых путей

- Турниром называется ориентированный граф между любыми двумя вершинами которых есть ровно одно ориентированное ребро.
- Гамильтонов путь — путь проходящий по всем вершинам ровно 1 раз.
- $\Omega = \{G_1, G_2, \ldots, G_{2^{C_n^2}}\}$ — множество всех турниров на n вершинах, все турниры равновероятны.
- σ — перестановка чисел от 1 до n.
  $$X_\sigma(G) = \begin{cases} 1, \text{если } \sigma \text{ задает г.п. в } G \\ 0, \text{иначе} \end{cases}$$
- $X = \sum_\sigma X_\sigma$ — число гамильтоновых путей в случайном графе.
- $\mathrm{E}\, X = \sum_\sigma \mathrm{E}\, X_\sigma = \frac{n!}{2^{n-1}}$.
- Значит, существует турнир в котором не меньше $\frac{n!}{2^{n-1}}$ гамильтоновых путей.

Двудольный подграф

- Теорема. Из любого графа $G(V, E)$ можно выбросить не более $\frac{|E|}{2}$ ребер так, чтобы он стал двудольным.
- Доказательство. $\Omega = 2^V$, все подмножества равновероятны.
- Пусть $T \in \Omega$ $(T \subseteq V)$, определим случайную величину для каждого ребра (x, y):

$$X_{xy}(T) = \begin{cases} 1, \text{если ровно одна вершина из } x, y \text{ содержится в } T \\ 0, \text{иначе} \end{cases}$$

- $X = \sum_{(x,y)\in E} X_{xy}$ — количество ребер, ровно одна из вершин которых содержится в T.
- $\mathrm{E}\,X = \sum_{(x,y)\in E} \mathrm{E}\,X_{xy} = \frac{|E|}{2}$.
- Значит, существует $T \in \Omega$, что $X(T) \geq \frac{|E|}{2}$.
- Выкинем все остальные ребра.

Условные вероятности

- Ω, $B \in \mathcal{A}$ $(B \subset \Omega)$, $\Pr\{B\} > 0$;
- $\Omega_B = \{A \cap B | A \in \mathbb{A}\}$;
- $\Pr\{A|B\} = \frac{\Pr AB}{\Pr\{B\}}$;
- Пусть $A_1, A_2, \ldots, A_n$ — полная система несовместных событий ($A_i A_j = \emptyset$, $\bigcup A_i = \Omega$).
- $C = CA_1 \cup CA_2 \cup \cdots \cup CA_n$.
- (Формула полной вероятности) $\Pr\{C\} = \sum_i \Pr\{CA_i\} = \sum_i \Pr\{C|A_i\} \Pr\{A_i\}$

Независимость

- $A, B \subset \Omega$ называются независимыми, если $\Pr\{AB\} = \Pr\{A\} \Pr\{B\}$;
- $\Pr\{A|B\} = \Pr\{A\}$, $\Pr\{B|A\} = \Pr\{B\}$;
- Пример. $\Omega = \{00, 01, 10, 11\}$, все исходы равновероятны. A — первый бит равен 0, B — сумма битов четна. $\Pr\{A\} = \frac{1}{2}$, $\Pr\{B\} = \frac{1}{2}$, $\Pr\{AB\} = \frac{1}{4}$;
- События $\{A_i\}_{i \in I}$ называются взаимно независимыми, если для всех $T \subseteq I$ выполняется $\Pr\{\bigcap_{i \in T}\} = \prod_{i \in T} \Pr\{A_i\}$.
- (Дискретные) случайные величины ξ и η называются независимыми, если для всех $a, b \in \mathbb{R}$ выполняется $\Pr\{\xi = a, \eta = b\} = \Pr\{\xi = a\} \Pr\{\eta = b\}$.

Произведение матожиданий

Теорема. $X_1, X_2, \ldots, X_n$ — взаимно независимы. Тогда $E[X_1 X_2 \ldots X_n] = E[X_1]\,E[X_2]\ldots E[X_n]$.

Доказательство.

$$E[X_1 X_2 \ldots X_n] = \sum_x x \Pr\{X_1 X_2 \ldots X_n = X\} =$$

$$\sum_{x_1, x_2, \ldots, x_n} x_1 x_2 \ldots x_n \Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \overset{\text{(независимость)}}{=}$$

$$\sum_{x_1, x_2, \ldots, x_n} x_1 x_2 \ldots x_n \Pr\{X_1 = x_1\} \Pr\{X_2 = x_2\} \ldots \Pr\{X_n = x_n\} =$$

$$\Big(\sum_{x_1} x_1 \Pr\{X_1 = x_1\}\Big)\Big(\sum_{x_2} x_2 \Pr\{X_2 = x_2\}\Big) \ldots \Big(\sum_{x_n} x_n \Pr\{X_n = x_n\}\Big) =$$

$$\prod_{i=1}^{n} E[X_i]$$

Определение. Ковариация: $Cov(X, Y) = E[XY] - E[X]\,E[Y]$.

Дисперсия

Определение. Дисперсией случайной величины ξ называется величина $\mathsf{D}\,\xi = \mathsf{E}(\xi - \mathsf{E}\,\xi)^2$.

$\mathsf{D}\,\xi = \mathsf{E}(\xi - \mathsf{E}\,\xi)^2 = \mathsf{E}[\xi^2 - 2\xi\,\mathsf{E}\,\xi + (\mathsf{E}\,\xi)^2] = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 \geq 0$.

Лемма. (Неравенство Чебышева). $\mathsf{D}\,\xi = \sigma^2$, тогда

$$\Pr\{|\xi - \mathsf{E}\,\xi| \geq k\sigma\} \leq \frac{1}{k^2}.$$

Доказательство. $\eta = (\xi - \mathsf{E}\,\xi)^2$, $\mathsf{E}\,\eta = \sigma^2$.

$$\Pr\{|\xi - \mathsf{E}\,\xi| \geq k\sigma\} = \Pr\{\eta \geq k^2\sigma^2\} \overset{\text{(нер-во Маркова)}}{\leq} \tfrac{1}{k^2}.$$

Линейность дисперсии

Теорема. Если $\xi_1, \xi_2, \dots, \xi_n$ **попарно** независимы, то $\mathrm{D}(\xi_1 + \xi_2 + \dots + \xi_n) = \mathrm{D}\,\xi_1 + \mathrm{D}\,\xi_2 + \dots + \mathrm{D}\,\xi_n$.

Доказательство.

$$\mathrm{D}\Big(\sum_i \xi_i\Big) = \mathrm{E}\Big[\Big(\sum_i \xi_i - \sum_i \mathrm{E}\,\xi_i\Big)^2\Big] =$$

$$\mathrm{E}\Big[\Big(\sum_i \xi_i\Big)^2 - 2\Big(\sum_i \xi_i\Big)\Big(\sum_j \mathrm{E}\,\xi_j\Big) + \Big(\sum_i \mathrm{E}\,\xi_i\Big)^2\Big] =$$

$$\sum_i \mathrm{E}\,\xi_i^2 + 2\sum_{i<j} \mathrm{E}\,\xi_i\,\mathrm{E}\,\xi_j - 2\sum_i (\mathrm{E}\,\xi_i)^2 - 4\sum_{i<j} \mathrm{E}\,\xi_i\,\mathrm{E}\,\xi_j$$

$$+ \sum_i (\mathrm{E}\,\xi_i)^2 + 2\sum_{i<j} \mathrm{E}\,\xi_i\,\mathrm{E}\,\xi_j =$$

$$= \sum_i \mathrm{E}\,\xi_i^2 - \sum_i (\mathrm{E}\,\xi_i)^2 = \sum_i \mathrm{D}[\xi_i].$$

Оценки Чернова

- $X_1, X_2, \ldots, X_n$ — взаимно независимые случайные величины, принимающие значения из $\{0, 1\}$;
- $X = \sum_{i=1}^{n} X_i$, $m = \mathrm{E}\, X$;
- Хочется получить $\Pr\{X \geq (1+\delta)m\} \leq$ что-то маленькое
- $\mathrm{E}\, e^{tX_i} = (1 - p_i) + e^t p_i$;

$$\mathrm{E}\, e^{tX} = \mathrm{E}\, e^{t \sum_i X_i} = \mathrm{E} \prod_i e^{tX_i} = \prod_i \mathrm{E}\, e^{tX_i} = \prod_i (1 + p_i(e^t - 1)) \leq$$

$$\overset{(1+x \leq e^x)}{\leq} \prod_i e^{p_i(e^t-1)} = e^{\sum_i p_i(e^t-1)} = e^{m(e^t-1)}$$

Оценки Чернова

- $\mathsf{E}\, e^{tX} = e^{m(e^t - 1)}$
- $\Pr\{X \geq (1+\delta)m\} = \Pr\{e^{tX} \geq e^{(1+\delta)mt}\} \leq \frac{\mathsf{E}\, e^{tX}}{e^{(1+\delta)mt}} \leq \frac{e^{m(e^t-1)}}{e^{(1+\delta)mt}}$
- $t = ln(1+\delta)$
- $\Pr\{X \geq (1+\delta)m\} \leq (\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}})^m \leq e^{(\delta-(1+\delta)\ln(1+\delta))m} \leq e^{(\delta-(1+\delta)(\delta-\delta^2/2))m} = e^{(-\delta^2/2+\delta^3/3)m} \leq e^{-\delta^2 m/6}$
- $\Pr\{X \leq (1-\delta)m\} \leq (\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}})^m \leq e^{-\delta^2 m/2}$

Оценки Чернова

- $X_1, X_2, \ldots, X_n$ — одинаково распределенные взаимно независимые случайные величины.
- $\Pr\{X_i = 1\} = p, \Pr\{X_i = 0\} = 1 - p$.
  $\mathrm{E}\, X_i = p, \mathrm{E}\, X = \mathrm{E} \sum_{i=1}^{n} X_i = np$;
- $\Pr\{|\frac{\sum X_i}{n} - p| \geq \varepsilon\} = \Pr\{|X - np| \geq np\frac{\varepsilon}{p})\} \leq 2e^{-\frac{\varepsilon^2 n}{6p}}$
- 10000 раз бросали монетку. Оценить вероятность того, что выпало больше 5500 орлов?
- $e^{-\frac{(\frac{500}{10000})^2 \cdot 10000}{3}} \leq 0.00025$.

Марковская цепь

- Дан ориентированный граф. Для каждой вершины известны вероятности переходов по ребрам. "Блуждание"по такому графу — это марковский процесс.
- Формально. Вершины: $\{1, 2, \ldots, N\}$.
- $X_0, X_1, X_2, \ldots,$ принимают значения $\{1, 2, \ldots, N\}$.
- Известны $p_{j,i} = \Pr\{X_{k+1} = j | X_k = i\}$
- $\pi^{(k)} = (\pi_1^{(k)}, \pi_2^{(k)}, \ldots, \pi_n^{(k)})$ — распределение X_k.
- $\pi_j^{(k+1)} = \sum_{i=1}^{n} \Pr\{X_{k+1} = j | X_k = i\} \Pr\{X_k = i\} = \sum_{i=1}^{n} p_{j,i} \pi_i^{(k)}$;
- $\pi^{(k+1)} = P\pi^{(k)}$, $P = (p_{j,i})$ — матрица перехода;
- $\pi^{(k)} = P^k \pi^{(0)}$.