

Содержание

Введение	2
Оценивание (Estimation)	3
Несмещенность (Unbiasedness)	3
Квадратичное отклонение (MSE)	3
Задачи	3
Примеры распределений	8
Распределение Бернулли, $Bern(p)$	8
Геометрическое распределение, $Geom(p)$	8
Биномиальное распределение, $Binom(n, p)$	8
Распределение Пуассона, $Poiss(\lambda)$	9
Равномерное распределение, $R[a, b]$	9
Нормальное распределение, $\mathcal{N}(a, \sigma^2)$	10
Показательное распределение, $Exp(\lambda)$	10
Математическое ожидание (Expectation)	10
Примеры вычисления EX	10
Разные виды сходимости случайных величин	12
Сходимость почти наверное	13
Сходимость по вероятности	13
Сходимость в среднем	14
Сходимость по распределению	14
Задачи	14
ЗБЧ и ЦПТ	16
Закон Больших Чисел (Law of Large Numbers)	16
Нормальное распределение (Normal Distribution)	18
Центральная Предельная Теорема (Central Limit Theorem)	18
Задачи	21
Асимптотические свойства оценок	21
Задачи	22
Методы получения оценок	26
Метод моментов (Method of moments)	26
Оценка максимального правдоподобия (Maximum likelihood estimation)	26
Оценка методом спейсингов	27
Задачи	28
Информация	33
Чисто задачи по теории вероятностей	34
Сходимости (Convergences)	34
ЦПТ и ЗБЧ (CLT and LLN)	35
Теория меры	43
Начало. Измеримое множество	43

Введение

Опишем процедуру построения теории, на которой держится математическая статистика.

В первую очередь в курсе теории вероятностей вводится понятие вероятностного пространства (probability space). Это понятие обобщает под одним термином три сущности, позволяющие в дальнейшем исследовать различные теоретические вопросы.

Итак, вероятностным пространством называют тройку

$$(\Omega, \mathcal{F}, P),$$

где Ω — это пространство элементарных исходов (sample space, все возможные результаты эксперимента), \mathcal{F} — семейство подмножеств Ω , которые называются событиями (event space), P — функция на событиях (probability function, вероятность).

Задание вероятностного пространства означает задание модели физического эксперимента.

С точки зрения теории меры, вероятность P называют еще вероятностной мерой.

В экспериментах мы хотим уметь рассчитывать вероятности событий, которые можно описать числами. Например, температуру в комнате, количество аварий, количество голов в матче и т.д. Поэтому естественно вводится понятие случайной величины (random variable).

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

Случайная величина X , заданная на пространстве (Ω, \mathcal{F}, P) , задает автоматически свою вероятностную меру, но уже на подмножествах B значений X (по-умному можно так записать $B \subseteq \text{Im } X \subseteq \mathbb{R}$):

$$P_X(B) := P(\omega : X(\omega) \in B) \equiv P(X \in B) \equiv P(X^{-1}(B))$$

То есть каждая случайная величина, как говорят, индуцирует свое вероятностное пространство

$$(\mathcal{X}, \mathcal{F}_X, P_X)$$

Задание P_X называется заданием распределения X (probability distribution).

В задаче математической статистики мы имеем результат некоторого эксперимента. В терминах вероятностного пространства у нас есть исход $\omega \in \Omega$, на котором посчитана некоторая числовая характеристика X . В нашей мат. модели X — это случайная величина (обычно, случайный вектор), про распределение которой мы точно не знаем, но имеем уверенность, что это распределение лежит в некотором семействе распределений \mathcal{P} .

Если это семейство параметризовано, т.е. каждому распределению семейства соответствует некоторый числовой параметр $\theta \in \Theta \subseteq \mathbb{R}^k$, то такое семейство называется параметризованным, и тогда задача узнать распределение сводится к задаче узнать параметр распределения.

Таким образом, статистическая модель (статистическое пространство) выглядит следующим образом

$$(\mathcal{X}, \mathcal{F}, \mathcal{P})$$

где \mathcal{X} — множество всех возможных значений X , \mathcal{F} — семейство событий на \mathcal{X} , \mathcal{P} — семейство распределений X .

Оценивание (Estimation)

Выборкой (random sample) называется набор из n случайных величин X_1, \dots, X_n . Реализацией выборки называется набор чисел $X_1(\omega), \dots, X_n(\omega)$.

Итак, задача статистики заключается в том, чтобы получить какую-либо информацию о распределении по данной выборке X_1, \dots, X_n .

Узнать что-либо о распределении это то же самое, что узнать что-то о его параметре, некую величину $h(\theta)$, где θ — параметр распределения. То есть нам надо, чтобы была функция $\hat{\theta}(x_1, \dots, x_n)$, такая что случайная величина $\hat{\theta}(X_1, \dots, X_n)$ как-то характеризовала величину $h(\theta)$. Тогда $\hat{\theta}(X_1, \dots, X_n)$ называют оценкой $h(\theta)$ (point estimator).

Несмещенность (Unbiasedness)

Встает вопрос, как определить, насколько оценка хороша. Естественно потребовать, чтобы оценка в среднем была равна оцениваемому параметру, то есть, чтобы

$$E_{\theta}\hat{\theta}(X_1, \dots, X_n) = h(\theta)$$

где E_{θ} — математическое ожидание по мере P_{θ} (иначе говоря, для распределения P_{θ}).

Если такое условие выполнено, то оценка $\hat{\theta}$ называется несмещенной для $h(\theta)$ (unbiased estimator).

Квадратичное отклонение (MSE)

Как для случайных величин важно понятие дисперсии (среднеквадратичного отклонения от среднего), так и для оценок важно понятие квадратичного смещения.

Квадратичным смещением (mean squared error или MSE) называется

$$E_{\theta}(\hat{\theta}(X_1, \dots, X_n) - h(\theta))^2$$

Для несмещенных оценок квадратичное смещение равно $D_{\theta}\hat{\theta}(X_1, \dots, X_n)$.

Задачи

Задача 1. Пусть $F_{\theta}(x) = 1 - \theta^{x+1}$, $x \in \{0, 1, \dots\}$, $\Theta = (0, 1)$. Построить несмещенную оценку для θ а) от одного наблюдения, б) с квадратичным смещением, стремящимся к нулю.

□ Пункт (а). Нужно построить оценку $\hat{\theta}(X_1)$, чтобы $E_{\theta}\hat{\theta}(X_1) = \theta$.

Посчитаем математическое ожидание такой оценки в общем виде, а затем попытаемся понять, какой должна быть сама оценка.

$$E_{\theta}\hat{\theta}(X_1) = \sum_{k=0}^{\infty} \hat{\theta}(k)P_{\theta}(X_1 = k) = (*)$$

Найдем $P_{\theta}(X_1 = k)$ из нашей функции распределения.

Так как $F_{\theta}(x) = P_{\theta}(X \leq x)$, то

$$P_{\theta}(X_1 = k) = F_{\theta}(k) - F_{\theta}(k-1) = 1 - \theta^{k+1} - 1 + \theta^k = \theta^k(1 - \theta)$$

Теперь вернемся к подсчету мат. ожидания:

$$(*) = \sum_{k=0}^{\infty} \hat{\theta}(k) \theta^k (1 - \theta) = (1 - \theta) \sum_{k=0}^{\infty} \hat{\theta}(k) \theta^k$$

(воспользовались суммой геом. прогрессии $\sum_{k=0}^{\infty} \theta = \frac{1}{1-\theta}$)

Итак, чтобы оценка $\sum_{k=0}^{\infty} \hat{\theta}(k) \theta^k$ была несмещенной, нужно, чтобы

$$(1 - \theta) \sum_{k=0}^{\infty} \hat{\theta}(k) \theta^k = \theta \Leftrightarrow \sum_{k=0}^{\infty} \hat{\theta}(k) \theta^k = \frac{\theta}{1 - \theta} = \theta + \theta^2 + \theta^3 + \dots = \sum_{k=1}^{\infty} \theta^k$$

Таким образом, мы хотим, чтобы $\hat{\theta}(0) = 0$, а $\hat{\theta}(k) = 1$ для всех $k = 1, 2, \dots$. Отсюда

$$\hat{\theta}(X_1) = I_{X_1 > 0}.$$

Пункт (б). Стандартный трюк улучшить оценку, оставив ее несмещенной, взять среднее от таких оценок. Рассмотрим

$$\hat{\theta}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n I_{X_i > 0}$$

Такая оценка тоже несмещенная, а значит ее квадратичное смещение равно ее дисперсии $D_{\theta} \hat{\theta}$. Найдем ее.

$$D_{\theta} \hat{\theta}(X_1, \dots, X_n) = \frac{1}{n^2} D_{\theta} \left(\sum I_{X_i > 0} \right) = \frac{1}{n^2} \cdot n \cdot D_{\theta} I_{X_1 > 0} = \frac{1}{n} D_{\theta} I_{X_1 > 0}$$

Вычислим

$$D_{\theta} I_{X_1 > 0} = E_{\theta} I_{X_1 > 0}^2 - (E_{\theta} I_{X_1 > 0})^2 = E_{\theta} I_{X_1 > 0} - (E_{\theta} I_{X_1 > 0})^2 = \theta - \theta^2 = \theta(1 - \theta)$$

Таким образом, квадратичное смещение равно $\frac{\theta(1-\theta)}{n}$. Оно стремится к нулю при $n \rightarrow \infty$. ■

Задача 2. а) Найти несмещенную оценку для DX_1 вида cS^2 , где $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, X_i — независимые одинаково распределенные (далее н.о.р.).

б) Найти ее квадратичное смещение в случае $X_i \sim \mathcal{N}(0, \theta)$.

□ Пункт а).

$$\begin{aligned} ES^2 &= \frac{1}{n} \sum_{i=1}^n E(X_i - \bar{X})^2 = \frac{1}{n} \sum (EX_i^2 - 2EX_i \bar{X} + E\bar{X}^2) = \\ &= \frac{1}{n} \sum EX_i^2 - \frac{2}{n} \sum EX_i \bar{X} + \frac{1}{n} \sum E\bar{X}^2 = \\ &= EX_1^2 - 2E \left(\frac{1}{n} \sum X_i \bar{X} \right) + \frac{1}{n} n E\bar{X}^2 = \\ &= EX_1^2 - 2E\bar{X}^2 + E\bar{X}^2 = EX_1^2 - E\bar{X}^2 \end{aligned}$$

Теперь посчитаем $E\bar{X}^2$:

$$\begin{aligned} E\bar{X}^2 &= E \left(\frac{X_1 + \dots + X_n}{n} \right)^2 = \frac{1}{n^2} E(X_1^2 + \dots + X_n^2 + 2 \sum_{i < j} X_i X_j) = \\ &= \frac{1}{n^2} (nEX_1^2 + 2C_n^2 (EX_1)^2) = \frac{1}{n} EX_1^2 + \frac{n-1}{n} (EX_1)^2 \end{aligned}$$

Подставляя в предыдущую формулу, получаем, что

$$ES^2 = EX_1^2 - \frac{EX_1^2}{n} - \frac{n-1}{n}(EX_1)^2 = \frac{n-1}{n}(EX_1^2 - (EX_1)^2) = \frac{n-1}{n}EX_1$$

Таким образом, оценка S^2 — смещенная, но оценка $S_0 = \frac{n}{n-1}S^2$ будет несмещенной. Теперь пункт б). Теперь мы знаем, что $X_i \sim \mathcal{N}(0, \theta)$, то есть $EX_i = 0$, $DX_i = \theta$.

Так как наша оценка несмещенная, то квадратичным смещением будет просто дисперсия DS_0 .

$$DS_0 = ES_0^4 - (ES_0^2)^2 = ES_0^4 - (EX_1)^2 = ES_0^4 = E\left(\frac{n}{n-1}S^2\right)^2 = \frac{n^2}{(n-1)^2}ES^4$$

Для поиска ES^4 воспользуемся тем, что $S^2 = \overline{X^2} - \overline{X}^2$. тогда

$$\begin{aligned} ES^4 &= E\frac{1}{n^2}(\overline{X^2} - \overline{X}^2)^2 = \frac{1}{n^2}E(\overline{X^2} - 2\overline{X}^2 + \overline{X}^4) = \\ &= \frac{1}{n^2}\left(E\left(\overline{X^2}\right) - 2E\overline{X}^2E\overline{X}^2 + E\overline{X}^4\right) = (*) \end{aligned}$$

Теперь нужно отдельно посчитать мат. ожидания в скобках.

$$E\overline{X^2} = E\frac{X_1^2 + \dots + X_n^2}{n} = EX_1^2 = DX_1 + (EX_1)^2 = DX_1 = \theta$$

$$E\overline{X}^2 = \frac{EX_1^2}{n} + \frac{n-1}{n}(EX_1)^2 = \frac{EX_1^2}{n} = \frac{DX_1}{n} = \frac{\theta}{n}$$

$$\begin{aligned} E\left(\overline{X^2}\right) &= E\left(\frac{X_1^2 + \dots + X_n^2}{n}\right)^2 = \frac{1}{n^2}E\left(X_1^4 + \dots + X_n^4 + 2\sum_{i<j} X_i^2 X_j^2\right) = \\ &= \frac{1}{n^2}\left(nEX_1^4 + 2C_n^2 (EX_1^2)^2\right) = (**) \end{aligned}$$

Мы знаем, что $EX_1^2 = DX_1 = \theta$. Осталось посчитать EX_1^4 .

$$\begin{aligned} EX_1^4 &= \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2\theta}} dx = \frac{\sqrt{2}}{\sqrt{\pi\theta}} \int_0^{+\infty} x^4 e^{-\frac{x^2}{2\theta}} dx = \\ &= \theta \sqrt{\frac{2}{\pi\theta}} \int_0^{+\infty} x^3 e^{-\frac{x^2}{2\theta}} d\left(\frac{x^2}{2\theta}\right) = \sqrt{\frac{2\theta}{\pi}} \int_0^{+\infty} (2\theta u)^{3/2} e^{-u} du = \\ &= \frac{4\theta^2}{\sqrt{\pi}} \int_0^{+\infty} u^{3/2} e^{-u} du = \frac{4\theta^2}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) = \frac{4\theta^2}{\sqrt{\pi}} \cdot \frac{3}{2} \Gamma\left(\frac{3}{2}\right) = \\ &= \frac{4\theta^2}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{4\theta^2}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \sqrt{\pi} = 3\theta^2 \end{aligned}$$

Таким образом,

$$(**) = \frac{1}{n^2} (3\theta^2 n + n(n-1)\theta^2) = \frac{n+2}{n}\theta^2$$

Еще нужно вычислить

$$E\overline{X}^4 = \frac{1}{n^4} n EX_1^4 = \frac{3\theta^2}{n^3}$$

Тогда

$$(*) = \frac{1}{n^2} \left(\frac{n+2}{n} \theta^2 - 2 \cdot \theta \cdot \frac{\theta}{n} + \frac{3\theta^2}{n^3} \right) = \frac{n^3 + 3}{n^5} \theta^2$$

Окончательно,

$$DS_0 = \frac{n^2}{(n-1)^2} \frac{n^3 + 3}{n^5} \theta^2 = \frac{n^3 + 3}{(n-1)^2 n^3} \theta^2$$

■

Задача 3. Доказать, что по $X_1, \dots, X_n \sim \text{Bern}(\theta)$ нельзя построить несмещенную оценку для e^θ , $\Theta = (0, 1)$.

□ Хотим построить оценку $\hat{\theta}(X_1, \dots, X_n)$, такую, чтобы $E_\theta \hat{\theta}(X_1, \dots, X_n) = e^\theta$.

$$\begin{aligned} E_\theta \hat{\theta}(X_1, \dots, X_n) &= \sum_{(x_1, \dots, x_n) \in \{0,1\}^n} \hat{\theta}(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n) = \\ &= \sum_{(x_1, \dots, x_n) \in \{0,1\}^n} \hat{\theta}(x_1, \dots, x_n) P(X_1 = x_1) \dots P(X_n = x_n) = \\ &= \sum_{k=0}^n \hat{\theta}_k(x_1, \dots, x_n) \theta^k (1-\theta)^{n-k} = \sum_{k=0}^n c_k(x_1, \dots, x_n) \theta^k \end{aligned}$$

Мы получили разложение в ряд Тейлора, который в нашем случае всегда будет обычной суммой из конечного количества слагаемых. В то время, как ряд Тейлора для экспоненты содержит бесконечное количество ненулевых коэффициентов:

$$e^\theta = 1 + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \dots$$

Так как ряд Тейлора определяется единственным образом, то совпадения тут быть не может ни при каком $\hat{\theta}$. ■

Задача 4. Пусть $X_i \sim R[0, \theta]$, $\Theta = (0, +\infty)$. Рассмотрим оценку параметра θ вида $c \max_{i \leq n} X_i$. Найти c , при котором эта оценка будет несмещенной. Найти квадратичное отклонение.

□ Обозначим $\hat{\theta} = c \max_{i \leq n} X_i$. Тогда $E\hat{\theta} = cE(\max X_i)$. Чтобы вычислить это мат. ожидание, нам нужно найти распределение $\max X_i$.

Рассмотрим функцию распределения

$$\begin{aligned} F_{\max X_i}(x) &= P(\max X_i \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \dots P(X_n \leq x) = (P(X_1 \leq x))^n = \\ &= \left(\frac{1}{\theta} \int_0^x I_{[0,\theta]}(t) dt \right)^n \end{aligned}$$

Отсюда

$$F_{\max X_i}(x) = \begin{cases} 0, & x \leq 0, \\ \left(\frac{x}{\theta}\right)^n, & 0 < x < \theta \\ 1, & x \geq \theta \end{cases}$$

Значит, можем найти плотность

$$f_{\max X_i}(x) = \frac{n}{\theta^n} x^{n-1} I_{[0,\theta]}(x)$$

А тогда

$$\begin{aligned} E \max X_i &= \int_{\mathbb{R}} x \cdot \frac{n}{\theta^n} x^{n-1} I_{[0,\theta]}(x) dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \\ &= \frac{n}{\theta^n} \frac{x^{n+1}}{n+1} \Big|_0^\theta = \frac{n}{n+1} \theta. \end{aligned}$$

Отсюда получаем, что оценка $\hat{\theta}_0 = \frac{n+1}{n} \max_{i \leq n} X_i$ будет несмещенной.

Квадратичное смещение в случае несмещенной оценки будет просто равно дисперсии этой оценки

$$D\hat{\theta}_0 = E\hat{\theta}_0^2 - (E\hat{\theta}_0)^2 = E\hat{\theta}_0^2 - \theta^2.$$

Найдем

$$E\hat{\theta}_0^2 = \left(\frac{n+1}{n} \right)^2 E(\max X_i)^2$$

Далее,

$$\begin{aligned} E(\max X_i)^2 &= \int_{\mathbb{R}} x^2 f_{\max X_i}(x) dx = \int_{\mathbb{R}} x^2 \frac{n}{\theta^n} x^{n-1} I_{[0,\theta]}(x) dx = \\ &= \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2} \theta^2. \end{aligned}$$

Таким образом,

$$E\hat{\theta}_0^2 = \left(\frac{n+1}{n} \right)^2 \frac{n}{n+2} \theta^2 = \frac{(n+1)^2}{n(n+2)} \theta^2$$

и, значит,

$$D\hat{\theta}_0 = \frac{(n+1)^2}{n(n+2)} \theta^2 - \theta^2 = \left(\frac{(n+1)^2}{n(n+2)} - 1 \right) \theta^2 = \frac{\theta^2}{n(n+2)}.$$

■

Примеры распределений

Распределение Бернулли, $Bern(p)$

$X \sim Bern(p)$, если X принимает всего два значения (успех и неудача):

$$X = \begin{cases} 1, & p \\ 0, & 1 - p \end{cases}$$

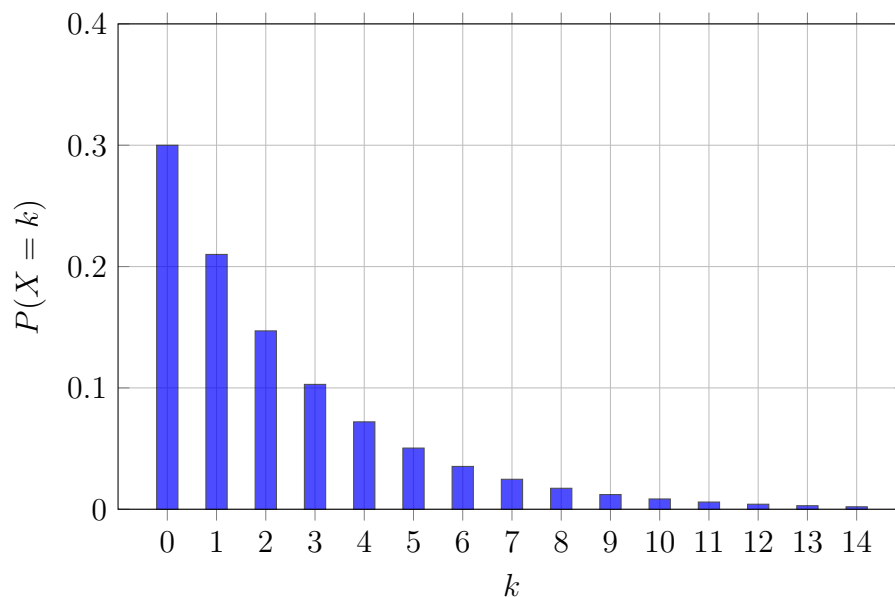
Геометрическое распределение, $Geom(p)$

Случайная величина $X \sim Geom(p)$, $p \in (0, 1)$ означает количество неудач до первого успеха при последовательном проведении эксперимента с бинарным исходом.

Распределение X таково:

$$P(X = k) = p(1 - p)^k, \quad k \geq 0.$$

$$X \sim Geom(p), \quad p = 0.3$$

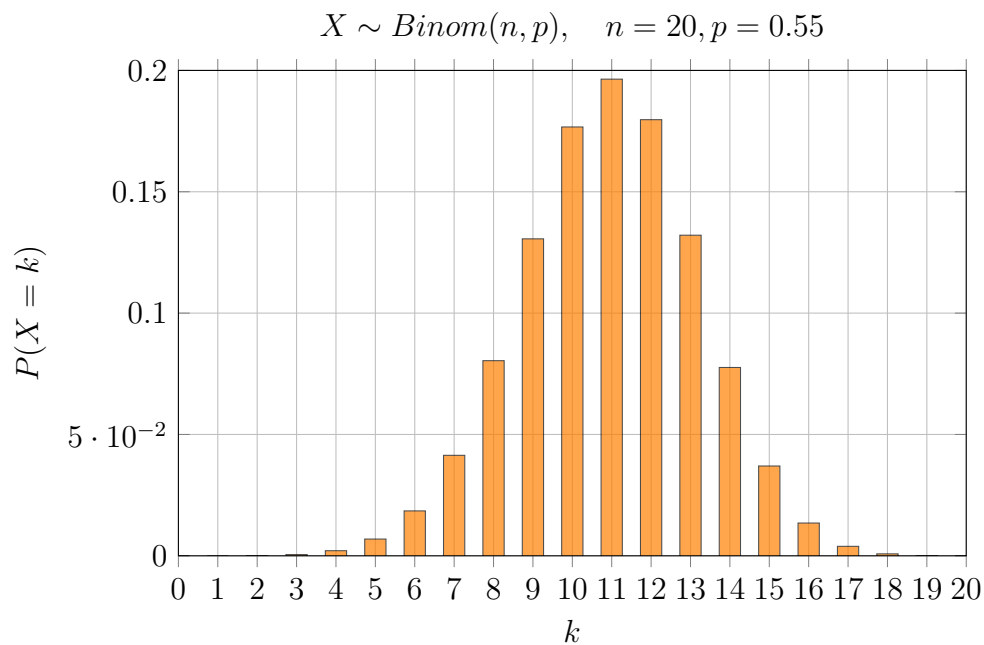


Биномиальное распределение, $Binom(n, p)$

Случайная величина $X \sim Binom(n, p)$ означает количество успехов при проведении n экспериментов с бинарным исходом, т.е. $X = \sum_{k=1}^n X_i$, где $X_i \sim Bern(p)$.

Распределение X имеет следующее:

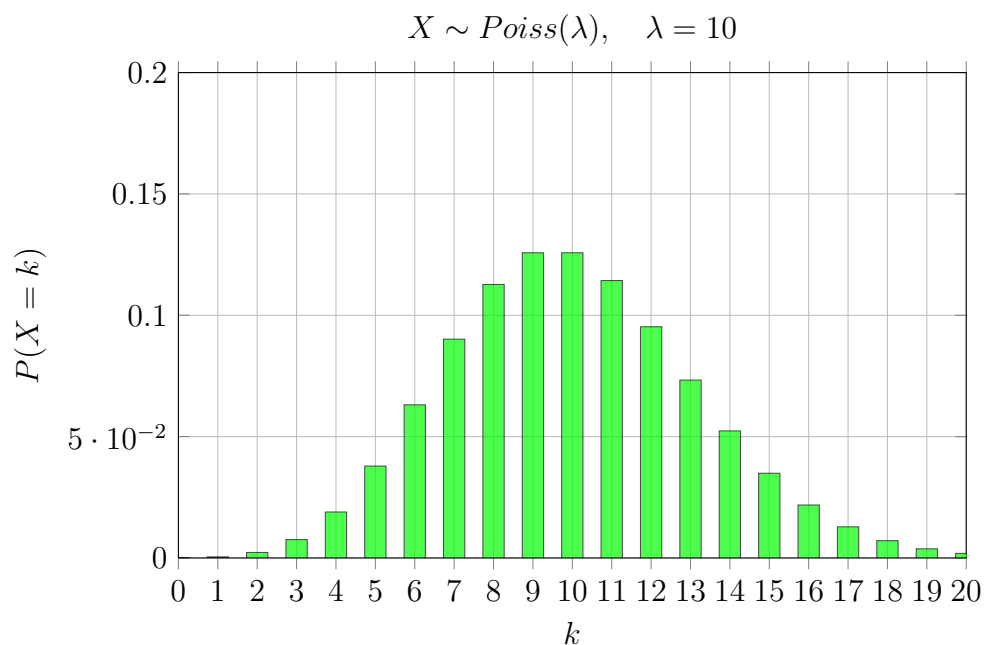
$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n.$$



Распределение Пуассона, $Poiss(\lambda)$

Случайная величина $X \sim Poiss(\lambda)$ означает количество событий, которые произойдут за некоторый промежуток времени T , про который известно, что λ — среднее число событий за время T .

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0.$$



Равномерное распределение, $R[a, b]$

$X \sim R[a, b]$, если $f_X(x) = I_{[a, b]}(x)$

Нормальное распределение, \mathcal{N}

$X \sim \mathcal{N}(a, \sigma^2)$, если X имеет плотность распределения

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Показательное распределение, $Exp(\lambda)$

$X \sim Exp(\lambda)$, $\lambda > 0$, если X имеет плотность распределения

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Полезный источник: учебник Яндекса по ML.

Математическое ожидание (Expectation)

Если у нас случайная величина X все свои возможные значения $\{x_1, x_2, \dots, x_n\}$ принимает с одинаковой вероятностью, то тогда ее среднее значение EX рассчитывается просто, как среднее арифметическое:

$$EX = \frac{x_1 + \dots + x_n}{n}$$

Но если вероятность распределена по значениям неравномерно, то тогда среднее значение случайной величины может быть отличным от среднего арифметического.

В этом случае, обозначив $p_k = P(X = x_k)$, среднее логично рассчитывать как

$$EX = x_1 p_1 + \dots + x_n p_n$$

С точки зрения физики, мы нашли центр масс (p_i — масса груза, размещенного в точке с координатой x_i).

В случае, если X принимает счетное множество значений, мат. ожидание запишется в виде ряда

$$EX = \sum_{k=1}^{\infty} x_k P(X = x_k)$$

Если же X — непрерывная случайная величина, то возникает понятие плотности распределения $f_X(x)$, с помощью которой мат. ожидание вычисляется с помощью интеграла Римана по формуле

$$EX = \int_{\mathbb{R}} x f_X(x) dx$$

Для того, чтобы подчеркнуть, что мат. ожидание берется именно по распределению P , используют обозначение $E_P X$.

Примеры вычисления EX

Пример 1. Распределение Бернулли: $X \sim Bern(p)$ — бросание монетки. X принимает два значения: 1 с вероятностью p (успех) и 0 с вероятностью $1 - p$ (неуспех).

□

$$EX = 0 \cdot (1 - p) + 1 \cdot p = p$$

Пример 2. Биномиальное распределение: $X \sim \text{Binom}(n, p)$ — число орлов при бросании монетки n раз.

□

$$\begin{aligned} EX &= \sum_{k=0}^n k \cdot P(X = k) = \sum_{k=0}^n k \cdot C_n^k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \end{aligned}$$

Но можно и проще, если заметить, что $X = X_1 + \dots + X_n$, где $X_k \sim \text{Bern}(p)$:

$$EX = EX_1 + \dots + EX_n = np.$$

Пример 3. Геометрическое распределение: $X \sim \text{Geom}(p)$ — число бросков до первого орла (успеха).

□ Сначала отметим, что $P(X = k) = p(1-p)^k$. Тогда

$$EX = \sum_{k=0}^{\infty} kp(1-p)^k.$$

Чтобы такое вычислить, нужно использовать трюк. Рассмотрим функцию

$$\varphi_X(s) = \sum_{k=0}^{\infty} s^k p(1-p)^k$$

(эта функция называется производящей).

Тогда $\varphi'_X(1) = EX$.

Считаем:

$$\begin{aligned} \varphi_X(s) &= \sum_{k=0}^{\infty} s^k p(1-p)^k = p \sum_{k=0}^{\infty} (s(1-p))^k = p \frac{1}{1-s(1-p)} \\ \varphi'_X(s) &= \frac{p(1-p)}{(1-s(1-p))^2} \\ \varphi'_X(1) &= \frac{1-p}{p} \end{aligned}$$

Пример 4. Пусть есть N ячеек, в которые бросают n шаров. Шары могут попадать в одни и те же ячейки. X — число занятых ячеек.



□ Представим $X = X_1 + \dots + X_N$, где X_k — занята ли k -я ячейка. Тогда

$$EX = \sum EX_k = NEX_1$$

так как все равно какую ячейку рассматриваем.

Далее,

$$EX_1 = 0 \cdot P(X_1 = 0) + 1 \cdot P(X_1 = 1) = P(X_1 = 1) = 1 - P(X_1 = 0) = 1 - \left(\frac{N-1}{N}\right)^n$$

Таким образом,

$$EX = N \left(1 - \left(\frac{N-1}{N}\right)^n\right)$$

■

Пример 5. (Рулетка) В комнате n человек, у каждого револьвер, по звонку каждый стреляет в случайного человека (кроме себя). Пусть X_n — число выживших. Найдите EX_n .

□ Введем Y_i :

$$Y_i = \begin{cases} 1, & \text{если } i\text{-ый выжил,} \\ 0, & \text{если } i\text{-ый не выжил} \end{cases}$$

Тогда $X_n = \sum_{k=1}^n Y_k$ и $EX_n = \sum_{k=1}^n EY_k$.

$$EY_k = P(k\text{-ый выжил})$$

Чтобы посчитать вероятность, нужно задать вероятностное пространство. Зададим

$$\Omega = \{(x_1, \dots, x_n) : x_i \neq i\}$$

Тогда

$$\begin{aligned} EY_k &= P(k\text{-ый выжил}) = \frac{(n-2) \cdots (n-2)(n-1)(n-2) \cdots (n-2)}{(n-1)^n} = \\ &= \frac{(n-2)^{n-1}(n-1)}{(n-1)^n} = \left(\frac{n-2}{n-1}\right)^{n-1} \end{aligned}$$

Значит,

$$EX_n = n \cdot \left(\frac{n-2}{n-1}\right)^{n-1}$$

Отсюда можно заметить, что доля выживших

$$\frac{EX_n}{n} = \left(\frac{n-2}{n-1}\right)^{n-1} = \left(1 - \frac{1}{n-1}\right)^{n-1} \rightarrow \frac{1}{e}$$

■

Разные виды сходимости случайных величин

Допустим, мы имеем выборку X_1, \dots, X_n и пытаемся построить разные оценки параметра распределения. Интуитивно понятно, что чем больше выборка, тем лучше можно оценить распределение. В идеале, мы бы делали бесконечно много экспериментов и тогда бы получали наилучшие оценки, но в жизни такое невозможно. Поэтому возникает логичный вопрос: какой размер выборки для нас достаточен в том смысле, что конечная выборка заданного размера даст практически столько же информации, сколько и бесконечная выборка.

Здесь появляется потребность в изучении различных предельных свойств. Видов сходимостей в теории вероятностей аж четыре, и все они важны, а текста и новых значков будет много. Но если разобраться идейно, то ничего сложного в самих определениях нет. В любом случае, без этого никуда, так что выхода нет.

Сходимость почти наверное

Мы неплохо знаем из обычного курса математического анализа про поточечную сходимость, когда последовательность функций $f_n(x)$ сходится к функции $f(x)$ в каждой точке x . А что, если в каких-то точках сходимости нет, но таких точек "мало"? Вот такая сходимость называется сходимостью почти наверное.

Более строго, пусть $\{X_n\}_{n=1}^\infty$ и X — случайные величины на пространстве (Ω, \mathcal{F}, P) . Говорят, что $X_n \rightarrow X$ почти наверное, если вероятность того, что $X_n(w) \rightarrow X(w)$ равна 1:

$$P(\omega : X_n(w) \rightarrow X(\omega)) = 1.$$

Пример. Пусть $\Omega = [0, 1]$, $X_n(\omega) = \frac{1}{n}$. Тогда $X_n \rightarrow 0$ п.н.

Сходимость по вероятности

Если сходимость почти наверное ослабляет поточечную сходимость, используя вероятность, то сходимость по вероятности ослабляет равномерную сходимость.

Напомню, что равномерная сходимость функций требует от последовательности функций "равномерного" приближения к предельной функции сразу по всей области определения.

Итак, снова пусть $\{X_n\}_{n=1}^\infty$ и X — случайные величины на пространстве (Ω, \mathcal{F}, P) . Говорят, что $X_n \xrightarrow{P} X$, если

$$\forall \varepsilon > 0 \quad P(w : |X_n(w) - X(w)| > \varepsilon) \rightarrow 0$$

В чем разница между сходимостью по вероятности и сходимостью почти наверное? Чтобы ответить на этот вопрос, перепишем определение сходимости почти наверное в эквивалентной форме (без доказательства):

$$X_n \xrightarrow{\text{п.н.}} X \Leftrightarrow \forall \varepsilon > 0 \quad P\left(\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) \rightarrow 0.$$

Отсюда сразу видно четкое различие и то, что из сходимости почти наверное следует сходимость по вероятности. Но хочется иметь пример сходимости по вероятности, при которой не выполняется сходимость почти наверное.

Ясно, что проблема должна заключаться в том, что вероятность отклонения с ростом n стремится к нулю, но в будущем (суммарно) мы не можем гарантировать (с вероятностью 1), что отклонение будет малым.

Классический пример — бегающая ступенька Рисса

$$Y_{n,k} = I_{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]}$$

Эта последовательность сходится к нулю по вероятности, но не сходится к нулю ни в одной точке.

С практической точки зрения эти две сходимости не так уж различимы, потому что мы не умеем генерировать бесконечные последовательности. А весь смысл различия именно в поведении на бесконечности, которое проверить в эксперименте невозможно.

Сходимость в среднем

Название говорит само за себя. Эта сходимость говорит о сходимости $E|X_n - X|$ к нулю.

Еще говорят о сходимостях в среднем порядка $p \geq 1$, когда $E|X_n - X|^p \rightarrow 0$.

С точки зрения теории функций эта сходимость в пространстве Лебега функций L_p , поэтому обозначение для такой сходимости в контексте случайных величин таково: $X_n \xrightarrow{L_p} X$.

Пример. $\Omega = [0, 1]$, $X_n(\omega) = n^{\frac{1}{p}} I_{[0, n^{-1/p}]}$. Тогда $X_n \xrightarrow{\text{п.н.}} 0$, но $X_n \not\xrightarrow{L_p} 0$.

Сходимость по распределению

Говорят, что $X_n \xrightarrow{d} X$ (по распределению), если сходятся функции распределения $F_{X_n}(x) \rightarrow F_X(x)$.

Задачи

Задача 1. Доказать, что если ряд из математических ожиданий и ряд из дисперсий независимых случайных величин X_n сходится, то ряд из X_n сходится а) в L_2 б) в L_1 .

□ Воспользуемся критерием Коши сходимости в L_2 .

Пусть задано произвольно $\varepsilon > 0$. Рассмотрим

$$E \left(\sum_{i=1}^n X_i - \sum_{i=1}^m X_i \right)^2 = E \left(\sum_{i=m+1}^n X_i \right)^2 = D \left(\sum_{i=m+1}^n X_i \right) + \left(E \sum_{i=m+1}^n X_i \right)^2$$

Из-за сходимости ряда из мат. ожиданий и дисперсий, оба слагаемых при больших $m, n > N$ будут сколько угодно малы, поэтому и $E \left(\sum_{i=1}^n X_i - \sum_{i=1}^m X_i \right)^2 < \varepsilon$. То есть по критерию Коши, имеем сходимость ряда в L_2 .

Сходимость в L_1 следует из сходимости в L_2 (общий факт). ■

Задача 2. Пусть $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$. Верно ли, что $aX_n + Y_n \xrightarrow{d} aX + Y$, если а) $Y = \text{const}$, б) $Y \neq \text{const}$?

□ Пункт а) Верно, так как, если $Y = c = \text{const}$, то для точки непрерывности x сл.в. $aX + c$ и $\forall \delta > 0$

$$\begin{aligned} F_{aX_n + Y_n}(x) &= P(aX_n + Y_n \leq x) = P(aX_n + Y_n \leq x | |Y_n - c| \leq \delta) \underbrace{P(|Y_n - c| \leq \delta)}_{\rightarrow 1, \text{ при } n \rightarrow \infty} + \\ &+ P(aX_n + Y_n \leq x | |Y_n - c| > \delta) \underbrace{P(|Y_n - c| > \delta)}_{\rightarrow 0, \text{ при } n \rightarrow \infty} \end{aligned}$$

Нужно оценить

$$\begin{aligned} P(aX_n + Y_n \leq x | |Y_n - c| \leq \delta) &\leq P(aX_n + c \leq x + \delta) = \\ &= F_{X_n} \left(\frac{x - c + \delta}{a} \right) \rightarrow F_X \left(\frac{x - c + \delta}{a} \right) = F_{aX+c}(x + \delta) \end{aligned}$$

Аналогично получается, что

$$P(aX_n + Y_n \leq x | |Y_n - c|) \geq F_{X_n} \left(\frac{x - c - \delta}{a} \right) \rightarrow F_X \left(\frac{x - c - \delta}{a} \right) = F_{aX+c}(x - \delta)$$

Так как x — точка непрерывности, а δ — любое, то

$$F_{aX_n+Y_n}(x) \rightarrow F_{aX+c}(x), \quad n \rightarrow \infty.$$

Пункт б). Неверно. Возьмем $X_n \sim \mathcal{N}(0, 1)$ и $X \sim \mathcal{N}(0, 1)$. Очевидно, что по распределению последовательность X_n сходится к X . Но при этом она сходится по распределению и к $-X$. Но $X_n + X_n = 2X_n$ не сходится к $X + (-X) = 0$. ■

ЗБЧ и ЦПТ

Закон Больших Чисел (Law of Large Numbers)

Мы уже знакомы с понятием математического ожидания, которое было введено, по сути, исходя чисто из интуитивных представлений о том, как должно быть. Интуиция, напомним, следующая.

Пусть случайная величина X принимает n значений $\{x_1, \dots, x_n\}$ с вероятностями p_1, \dots, p_n . Чему в среднем равняется X ? Хочется взять большую выборку из N экспериментов, сложить в сумму выпавшие значения и поделить на N (то бишь посчитать среднее арифметическое):

$$\frac{X_1 + \dots + X_N}{N} \approx \frac{x_1 \cdot Np_1 + \dots + x_n \cdot Np_n}{N} = x_1p_1 + \dots + x_np_n$$

Оказывается, что эта интуиция подкрепляется математическим обоснованием.

Якоб Бернулли (1655-1705) в своем труде *Ars Conjectandi* (Искусство Предположений), опубликованном в 1713 году его племянником Николаем Бернулли, доказал "Теорему Бернулли", ставшую началом истории развития теорем, объединенных названием "Закон больших чисел". Бернулли исследовал вопрос о том, насколько наше формальное математическое представление о вероятности соответствует действительности. А именно, что вероятность события означает частоту этого события при очень большом повторении одного эксперимента.

Название "Закон больших чисел" принадлежит Пуассону, который сформулировал ЗБЧ в более общем виде, предполагая, что вероятность события может быть разной для разных испытаний, но не доказал его.

Доказательство теореме Пуассона (в более общей формулировке) дал уже Чебышев в 1866 году, которое оказалось поразительно простым, использующим метод моментов (математических ожиданий).

*Использована речь А.А. Маркова на 200-летний юбилей теоремы Бернулли (из книги **Бернулли — О законе больших чисел, 1986**).*

Далее приведем доказательство ЗБЧ в простой его форме. Это доказательство строится на неравенствах Маркова и Чебышева, но и сами по себе эти неравенства полезно рассмотреть подробно, чтобы понимать логику рассуждений, приближающих к главному утверждению.

Неравенство Маркова. Пусть $X \geq 0$, $EX < \infty$. Тогда

$$\forall x > 0 \quad P(X \geq x) \leq \frac{EX}{x}.$$

□ Распишем EX , учитывая, что $X = XI_{X \geq x} + XI_{X < x}$:

$$\begin{aligned} EX &= E(XI_{X \geq x}) + E(XI_{X < x}) \geq E(XI_{X \geq x}) \geq xEI_{X \geq x} = \\ &= xP(X \geq x). \end{aligned}$$

Отсюда получаем $P(X \geq x) \leq \frac{EX}{x}$. ■

Пример. Пусть студенты никогда не приходят вовремя, они всегда опаздывают. В среднем они опаздывают на 3 минуты. Какова вероятность того, что студент опоздает на 15 минут и более? Дать грубую оценку сверху. (отсюда).

□

$$P(X \geq 15) \leq \frac{3}{15} = 0.2.$$
■

О чем говорит неравенство Маркова по сути? Можно заметить, что при $x < EX$ утверждение не несет никакой полезной информации, вероятность по определению не больше единицы. Содержательность кроется в том, что при увеличении $x > EX$ вероятность получить результат больше этого числа убывает обратно пропорционально росту числа x .

Это фундаментальное свойство, которое интуитивно должно выполняться, и выполняется. Но это еще не ЗБЧ, хотя и важный шаг к нему.

Неравенство Чебышева. Пусть X — случайная величина, $DX \leq \infty$. Тогда

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

□ Положим $Y := |X - EX|$ и применим к Y неравенство Маркова:

$$P(|X - EX| \geq \varepsilon) = P((X - EX)^2 \geq \varepsilon^2) \leq \frac{E(X - EX)^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}.$$

■

Здесь опять, как в неравенстве Маркова, утверждение об отклонении от математического ожидания. Но утверждение более конкретное: вероятность отклониться от математического ожидания можно оценить сверху числом зависящим от дисперсии. Что вполне логично, ведь мы интуитивно и думали о дисперсии, как о мере разброса.

Следствие (Правило трех сигм). Если $EX^2 < \infty$, то

$$P(|X - EX| \leq 3\sqrt{DX}) \geq \frac{8}{9}.$$

□

$$P(|X - EX| > 3\sqrt{DX}) \leq \frac{DX}{(3\sqrt{DX})^2} = \frac{1}{9}.$$

■

Теорема (Закон Больших Чисел). Пусть X_1, \dots, X_n, \dots — н.о.р., $EX_1^2 < \infty$. Тогда

$$\forall \varepsilon > 0 \quad P\left(\left|\frac{S_n}{n} - EX_1\right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty$$

или, на языке сходимостей,

$$\frac{S_n}{n} \xrightarrow{P} EX_1.$$

□ Пусть $X = \left|\frac{1}{n}S_n - EX_1\right|$. Тогда

$$P\left(\left|\frac{1}{n}S_n - EX_1\right| > \varepsilon\right) \leq \frac{D\left(\frac{1}{n}S_n - EX_1\right)}{\varepsilon^2} = \frac{D\left(\frac{1}{n}S_n\right)}{\varepsilon^2} = \frac{nDX_1}{\varepsilon^2 n^2} = \frac{DX_1}{\varepsilon^2 n} \rightarrow 0.$$

■

Усиленный ЗБЧ (Колмогоров, 1930) Пусть X_1, \dots, X_n, \dots — н.о.р., и $E|X_1| < \infty$. Тогда

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = EX_1\right) = 1.$$

Нормальное распределение (Normal Distribution)

Опр. Нормальным распределением называется распределение, задаваемое плотностью

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где a — мат. ожидание распределения, а σ^2 — дисперсия.

В случае $a = 0$, $\sigma^2 = 1$ говорят, что случайная величина имеет стандартное нормальное распределение, а ее плотность упрощается до

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

В жизни мы ощущаем нормальное распределение повсюду, этим никого не удивить. Считается, что есть норма, а есть отклонение от нормы и подсознательно даже понятно, что распределение симметрично и так далее. Но вот откуда возникает такое распределение, почему именно оно и в чем его секрет? На этот вопрос отвечает центральная предельная теорема.

Центральная Предельная Теорема (Central Limit Theorem)

Пример. Среди людей 1% левшей. Выбрали 100 человек. Какая вероятность, что среди них ≤ 2 левшей?

□ Пусть X — количество левшей. Тогда

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) = \\ &= 0.99^{100} + 0.99^{99} \cdot 0.01 \cdot 100 + C_{100}^2 \cdot 0.99^{98} \cdot 0.01^2. \end{aligned}$$

Но как такое считать? Очень сложно. ■

Теорема (Центральная предельная теорема). Пусть X_1, \dots, X_n, \dots — н.о.р., $EX_1^2 < \infty$. Тогда для любых $a < b$

$$P\left(a < \left(\frac{S_n}{n} - EX_1\right) \frac{\sqrt{n}}{\sqrt{DX_1}} < b\right) \rightarrow \frac{1}{2\pi} \int_a^b e^{-\frac{x^2}{2}} dx, \quad n \rightarrow \infty$$

или, на языке сходимостей,

$$\left(\frac{S_n}{n} - EX_1\right) \frac{\sqrt{n}}{\sqrt{DX_1}} \xrightarrow{d} Z \sim \mathcal{N}(0, 1)$$

или, переписав чуть иначе,

$$\sqrt{n} \left(\frac{S_n}{n} - EX_1\right) \xrightarrow{d} Z \sim \mathcal{N}(0, DX_1).$$

□ Заметим, что

$$\frac{S_n - na}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \widetilde{X}_k, \quad \text{где} \quad \widetilde{X}_k = \frac{X_k - na}{\sigma}.$$

Тогда $E\widetilde{X}_k = 0$, $D\widetilde{X}_k = 1$.

Вспомним, что характеристической функцией случайной величины X называется функция

$$\varphi_X(t) = Ee^{itX}, \quad t \in \mathbb{R}.$$

И для характеристической функции справедливо разложение

$$\varphi_X(t) = 1 + itEX + \dots + \frac{(it)^n}{n!} EX^n + o(t^n), \quad t \rightarrow 0.$$

Поэтому

$$\varphi_{\widetilde{X}_k}(t) = 1 + itE\widetilde{X}_k + \frac{(it)^2}{2!} E\widetilde{X}_k^2 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2), \quad t \rightarrow 0.$$

Тогда по свойству характеристической функции

$$\varphi_{\frac{\widetilde{X}_1 + \dots + \widetilde{X}_n}{\sqrt{n}}}(t) = \varphi_{\widetilde{X}_1}\left(\frac{t}{\sqrt{n}}\right) \cdots \varphi_{\widetilde{X}_n}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{\widetilde{X}_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Итого получаем, что

$$\varphi_{\frac{S_n - na}{\sigma\sqrt{n}}}(t) = \left(\varphi_{\widetilde{X}_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{-\frac{t^2}{2}}.$$

По теореме Леви

$$\frac{S_n - na}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

■

Следствие. Если $EX_1 = 0$, $DX_1 = 1$, то ЦПТ упрощается:

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d} \mathcal{N}(0, 1)$$

Сделаем несколько пояснений по поводу центральной предельной теоремы.

Во-первых, обычно, когда хотят объяснить ребенку, что такое ЦПТ, говорят о том, что распределение суммы случайных величин стремится к нормальному. Но, если посмотреть на формулировку теоремы, то встает логичный вопрос: а зачем все эти нормировки, переусложнения? написали бы, что $S_n \xrightarrow{d} \mathcal{N}$ с какими-то параметрами, и все!

Но, немножко подумав, выясняется, что такой сходимости-то не будет вовсе. Получается, врут детям? Впрочем, вернемся к этому вопросу чуть позже.

Во-вторых, говорят о некоторой скорости сходимости, мол, сумма сходится к нормальному распределению со скоростью сходимости порядка \sqrt{n} . А это как понимать?

Здесь речь идет о порядке слабой сходимости или, что то же, сходимости по распределению, которая нас здесь и интересует. Из ЗБЧ мы знаем, что если $EX_1 = 0$, то $\frac{S_n}{n} \xrightarrow{P} 0$ (а мы помним, что сходимость по вероятности к константе равносильна сходимости по распределению к той же константе), или же что $S_n = o(n)$ в смысле сходимости по распределению. А ЦПТ дает гораздо более точную асимптотическую оценку, она утверждает, что если поделить S_n на нечто меньшее по порядку, чем \sqrt{n} , то получим расходямость, а если поделить на большее по порядку, то получим ноль. На языке o -символики, из ЦПТ следует, что $S_n = O(\sqrt{n})$.

По поводу сходимости частичных сумм S_n проблема возникает следующая. Если $DX_1 = \sigma^2$, то $DS_n = n\sigma^2$, то есть дисперсия стремится к бесконечности, и никакой нормальной случайной величины мы в пределе не получим. Но это еще не повод расстраиваться. Распределение S_n при достаточно больших n все равно будет близко к нормальному, просто предела не будет.

Допустим надо найти вероятность того, что $F_{S_n}(m) = P(S_n \leq m)$, при этом n считается достаточно большим, тогда

$$\begin{aligned} P(S_n \leq m) &= P\left(\frac{\sqrt{n}}{\sqrt{DX_1}} \left(\frac{S_n}{n} - EX_1\right) \leq \frac{\sqrt{n}}{\sqrt{DX_1}} \left(\frac{m}{n} - EX_1\right)\right) \approx \\ &\approx \Phi\left(\frac{\sqrt{n}}{\sqrt{DX_1}} \left(\frac{m}{n} - EX_1\right)\right) \end{aligned}$$

Вот в каком смысле S_n стремится к нормальной случайной величине.

Теперь по поводу скорости сходимости. Закон больших чисел говорит о том, что среднее арифметическое стремится по вероятности к математическому ожиданию, но ничего не говорит о том, как оно стремится. ЦПТ дает сильно больше.

Рассмотрим

$$\begin{aligned} P\left(-\varepsilon < \frac{\sqrt{n}}{\sqrt{DX_1}} \left(\frac{S_n}{n} - EX_1\right) < \varepsilon\right) &= P\left(-\varepsilon \frac{\sqrt{DX_1}}{\sqrt{n}} < \frac{S_n}{n} - EX_1 < \varepsilon \frac{\sqrt{DX_1}}{\sqrt{n}}\right) \approx \\ &\approx 2\Phi(\varepsilon) - 1 \end{aligned}$$

То есть при увеличении n радиус области уменьшается со скоростью \sqrt{n} , а вероятность попасть в эту область сохраняется постоянной.

Еще одно изложение сути центральной предельной теоремы. Запишем ее утверждение через O -символику:

$$\left(\frac{S_n}{n} - EX_1\right) \frac{\sqrt{n}}{\sqrt{DX_1}} = Z + o(1), \quad n \rightarrow \infty,$$

тогда можно переписать

$$\frac{S_n}{n} - EX_1 = \frac{\sqrt{DX_1}}{\sqrt{n}} Z + o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty,$$

в целом уже неплохо, но перепишем еще и вот так:

$$\frac{S_n}{n} = EX_1 + \frac{\sqrt{DX_1}}{\sqrt{n}} Z + o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty,$$

Теперь отлично видно, что ЦПТ дает асимптотическую формулу для $\frac{S_n}{n}$, которая складывается из суммы математического ожидания, нормально распределенной случайной величины $\mathcal{N}\left(0, \frac{DX_1}{n}\right)$ и остаточной величины, стремящейся к нулю быстрее, чем $\frac{1}{\sqrt{n}}$.

Пример. Какая вероятность того, что при 1600 подбрасываниях монеты, выпадет менее 700 орлов?

□

$$P(X_1 + \dots + X_n \leq 700) = P\left(\frac{X_1 + \dots + X_n - nEX_1}{\sqrt{nDX_1}} \leq \frac{700 - 800}{\sqrt{\frac{1}{4} \cdot 1600}}\right) \approx \Phi(-5) \approx 0$$

■

Пример. Какова вероятность того, что мальчиков не больше, чем девочек среди $n = 1000$ детей?

□

$$P(S_n \leq 500) = P\left(\left(\frac{S_n}{n} - np\right) \frac{\sqrt{n}}{\sqrt{np(1-p)}} \leq \left(\frac{500}{n} - np\right) \frac{\sqrt{n}}{\sqrt{np(1-p)}}\right) \approx \\ \approx \Phi(-3) \approx 0.0015$$

■

Задачи

Задача 1. Складывается 10^6 чисел, округленных с точностью до 10^{-m} . Предполагая, что ошибки округления независимы и $\sim R(-0.5 \cdot 10^{-m}, 0.5 \cdot 10^{-m})$, найти пределы, в которых с вероятностью ≥ 0.95 будет лежать суммарная ошибка.

□ Дано $n = 10^6$ н.о.р.с.в. $\sim R(a, b)$, где $a = -0.5 \cdot 10^{-m}$, $b = 0.5 \cdot 10^{-m}$.

Мы хотим найти число c : $P(-c < S_n < c) \geq 0.95 = p$.

Заметим, что $EX_1 = \frac{a+b}{2} = 0 \Rightarrow ES_n = 0$, и $\sigma^2 = DX_1 = \frac{(b-a)^2}{12} = \frac{10^{-2m}}{12}$. Поэтому

$$\left(\frac{S_n}{n} - EX_1\right) \frac{\sqrt{n}}{\sqrt{DX_1}} = \frac{S_n}{\sigma\sqrt{n}}.$$

Тогда, используя ЦПТ, получим

$$P(-c < S_n < c) = P\left(\frac{|S_n|}{\sigma\sqrt{n}} < \frac{c}{\sigma\sqrt{n}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-c/\sigma\sqrt{n}}^{c/\sigma\sqrt{n}} e^{-\frac{x^2}{2}} dx = \\ = \Phi(c/\sigma\sqrt{n}) - (1 - \Phi(c/\sigma\sqrt{n})) = \\ = 2\Phi(c/\sigma\sqrt{n}) - 1 = 0.95 \Rightarrow \Phi(c/\sigma\sqrt{n}) = 1.95/2 = 0.975$$

Где $\Phi(x)$ — табличное значение. $\Phi(x) = 0.975$ при $x = 1.96$. Таким образом,

$$c = 1.96 \cdot \sigma\sqrt{n} = 1.96 \cdot \frac{10^{-m}}{\sqrt{12}} \cdot 10^3 \approx 0.57 \cdot 10^{3-m}$$

■

Асимптотические свойства оценок

Опр. Последовательность оценок $\{\hat{\theta}_n(X_1, \dots, X_n)\}_{n=1}^\infty$ называется состоятельной оценкой (consistent estimator) параметра $g(\theta)$, если

$$\hat{\theta}_n(X_1, \dots, X_n) \xrightarrow{P} g(\theta) \quad \forall \theta \in \Theta$$

и обозначается просто $\hat{\theta}(X_1, \dots, X_n)$.

При этом если

$$\hat{\theta}_n(X_1, \dots, X_n) \xrightarrow{\text{п.н.}} g(\theta) \quad \forall \theta \in \Theta,$$

то $\hat{\theta}$ называется сильно состоятельной (strongly consistent).

Пример. Если $X_i \sim \text{Bern}(\theta)$, то $\bar{X} \xrightarrow{\text{п.н.}} \theta$ в силу УЗБЧ. Значит, \bar{X} — сильно состоятельная оценка. Аналогично и в произвольной модели, где $EX_i < \infty$, \bar{X} будет сильно состоятельно оценкой для мат. ожидания.

Опр. Если существует последовательность $\{\sigma_n(\theta)\}$, сходящаяся к нулю, такая, что

$$\frac{\hat{\theta}_n(X_1, \dots, X_n) - g(\theta)}{\sigma_n(\theta)} \xrightarrow{d} Z \sim \mathcal{N}(0, 1),$$

то оценка $\hat{\theta}$ называется асимптотически нормальной оценкой $g(\theta)$ (asymptotically normal). Мы будем рассматривать оценки вида

$$\sqrt{n} \frac{\hat{\theta}_n(X_1, \dots, X_n) - g(\theta)}{\sigma(\theta)} \xrightarrow{d} Z \sim \mathcal{N}(0, 1).$$

Величина $\sigma^2(\theta)$ называется асимптотической дисперсией оценки $\hat{\theta}$ (asymptotic variance).

Лемма. Если $\exists a \in \mathbb{R}$ и последовательность $g_n \rightarrow \infty$, такие, что

$$g_n(X_n - a) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2),$$

то

$$\forall h \in D(\mathbb{R}) \quad g_n(h(X_n) - h(a)) \xrightarrow{d} Z \sim \mathcal{N}(0, (h'(a))^2 \cdot \sigma^2)$$

или, иначе,

$$g_n \frac{h(X_n) - h(a)}{\sigma|h'(a)|} \xrightarrow{d} Z \sim \mathcal{N}(0, 1).$$

То есть асимптотическая нормальность является инвариантным свойством относительно действия дифференцируемых функций.

Почему вообще мы хотим, чтобы оценка была асимптотически нормальной? Причина в том, что математика все-таки строится, опираясь на жизненную интуицию. А в жизни, когда мы говорим об оценивании чего-либо, всегда подразумеваем, часто не осознавая, нормальную оценку.

Иначе говоря, уже подумав об оценке какого-либо параметра (роста, возраста, пола человека), мы по тем данным, которые есть, предполагаем, что есть некая норма, а есть отклонение от этой нормы. Поэтому требовать от оценки, что она будет на множестве испытаний давать именно нормальное распределение — естественно.

Но и скорость сходимости \sqrt{n} мы требуем неслучайно. Да, можно и другую рассматривать, но \sqrt{n} , в каком-то смысле, канонична. Мы же уже имеем из ЦПТ оценку математического ожидания, которая имеет именно такую оценку

Задачи

Задача 1. $X_i \sim \mathcal{N}(\theta, 1)$, $i \leq n$. Для какого параметра составительна оценка $(\bar{X})^2 \cdot \bar{X}^2$?

□ По ЗБЧ $\bar{X} \xrightarrow{P} EX_1 = \theta$. Поэтому $\bar{X}^2 \xrightarrow{P} \theta^2$.

Также по ЗБЧ:

$$\bar{X}^2 = \frac{X_1^2 + \dots + X_n^2}{n} \xrightarrow{P} EX_1^2 = DX_1 + (EX_1)^2 = 1 + \theta^2$$

Отсюда $(\bar{X})^2 \cdot \bar{X}^2 \xrightarrow{P} \theta^2(1 + \theta^2)$. ■

Задача 2. Показать, что $(X_{(1)} + \bar{X} - 1)/2$ является состоятельной оценкой, где $X_i \sim \exp(\theta, 1)$ (плотность $e^{-(x-\theta)} I_{x>\theta}$).

□ Найдем предел по распределению для $X_{(1)} := \min X_i$. Если будет иметь место сходимость к константе, то по вероятности будет тот же предел. Итак,

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) = P(\min X_i \leq x) = \\ &= 1 - P(\forall i X_i > x) = 1 - P(X_1 > x) \cdots P(X_n > x) = \\ &= 1 - (P(X_1 > x))^n = 1 - (1 - P(X_1 \leq x))^n \end{aligned}$$

Найдем ф.р. случайной величины X_1 :

$$F_{X_1}(x) = \int_{-\infty}^x e^{-(t-\theta)} I_{x>\theta}(t) dt$$

Видно, что при $x < \theta$ имеем $F_{X_1}(x) = 0$, при $x \geq \theta$

$$F_{X_1}(x) = \int_{\theta}^x e^{-(t-\theta)} dt = -e^{-(t-\theta)} \Big|_{\theta}^x = -e^{-(x-\theta)} + 1.$$

Тогда

$$F_{X_{(1)}}(x) = \begin{cases} 0, & x < \theta \\ 1 - (1 - 1 + e^{-(x-\theta)})^n, & x \geq \theta \end{cases} = (1 - e^{-n(x-\theta)}) I_{x \geq \theta}(x) \rightarrow I_{x \geq \theta}(x)$$

Таким образом, $X_{(1)} \xrightarrow{P} \theta$.

Далее, $\bar{X} \xrightarrow{P} EX_1$ по ЗБЧ. Найдем EX_1 :

$$\begin{aligned} EX_1 &= \int_{-\infty}^{+\infty} x e^{-(x-\theta)} I_{x>\theta} dx = \int_{\theta}^{+\infty} x e^{-(x-\theta)} dx = \\ &= x (-e^{-(x-\theta)}) \Big|_{\theta}^{+\infty} + \int_{\theta}^{+\infty} e^{-(x-\theta)} dx = \\ &= \theta + (-e^{-(x-\theta)}) \Big|_{\theta}^{+\infty} = \theta + 1. \end{aligned}$$

Итого, получаем, что

$$\frac{X_{(1)} + \bar{X} - 1}{2} \xrightarrow{P} \frac{\theta + \theta + 1 - 1}{2} = \theta.$$

■

Задача 3. Показать, что при $EX_1^4 < \infty$ оценка S^2 асимптотически нормальна для дисперсии X_i а) при $EX_1 = 0$, б) в общем случае. Найти ее асимптотическую дисперсию.

□ Напомним, что $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

По определению асимптотической нормальности нам надо показать, что

$$\sqrt{n} \frac{S^2 - DX_1}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

где σ — асимптотическая дисперсия.

Рассмотрим случай, когда $EX_1 = 0$.

Распишем

$$\begin{aligned}
 \sqrt{n}(S^2 - DX_1) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + X^2 - DX_1) = \\
 &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n (X_i^2 - DX_1) - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right) = \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - DX_1) - 2\sqrt{n}\bar{X}^2 + \sqrt{n}\bar{X}^2 = \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - DX_1) - \sqrt{n}\bar{X}^2 = (*)
 \end{aligned}$$

Обратим внимание на слагаемое $\sqrt{n}\bar{X}^2$. Из ЦПТ мы знаем, что $\bar{X} = O\left(\frac{1}{\sqrt{n}}\right)$, а потому $\sqrt{n}\bar{X}^2 = O\left(\frac{1}{\sqrt{n}}\right)$ и, значит, стремится к нулю. Но в этой записи скрыто слишком много технических нюансов, поэтому сделаем все аккуратно:

$$\begin{aligned}
 \sqrt{n}\bar{X}^2 &= \sqrt{n} \left(\sum_{i=1}^n X_i \right)^2 \cdot \frac{1}{n^2} = n^{-2/3} \left(\sum_{i=1}^n X_i \right)^2 = \\
 &= \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right)^2 \xrightarrow{d} 0,
 \end{aligned}$$

так как второй множитель стремится по распределению к $\mathcal{N}(0, 1)$, первый стремится к нулю как числовая последовательность, а результат, по лемме Слущкого стремится к нулю.

Теперь вернемся к первому слагаемому в (*). Тут хочется применить ЦПТ для суммы $Y_i = X_i^2 - DX_1$. Для этого найдем EY_1 и DY_1 .

$$\begin{aligned}
 EY_1 &= E(X_1^2 - DX_1) = EX_1^2 - DX_1 = EX_1^2 - (EX_1^2 - (EX_1)^2) = 0 \\
 DY_1 &= D(X_1^2 - DX_1) = DX_1^2 = EX_1^4 - (EX_1^2)^2 = EX_1^4 - (DX_1)^2
 \end{aligned}$$

Таким образом,

$$\sqrt{n} \frac{S^2 - DX_1}{\sqrt{EX_1^4 - (DX_1)^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

значит, асимптотическая дисперсия есть $EX_1^4 - (DX_1)^2$.

В случае $EX_1 = a \neq 0$ нужно рассмотреть $\tilde{X}_i = X_i - a$. Тогда $E\tilde{X}_i = 0$, при этом $S^2 = \tilde{S}^2$. И далее мы получим, что

$$\sqrt{n} \frac{\tilde{S}^2 - D\tilde{X}_1}{\sqrt{E\tilde{X}_1^4 - (D\tilde{X}_1)^2}} = \sqrt{n} \frac{S^2 - DX_1}{\sqrt{E(X_1 - a)^4 - (DX_1)^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Решение взято отсюда. ■

Задача 4. Пусть $X_i \sim R[0, \theta]$. Показать, что $\sqrt[k]{(k+1)\bar{X}^k}$ — асимптотически нормальная оценка θ . Найти ее асимптотическую дисперсию. Существует ли среди указанных оценок оценка с наименьшей асимптотической дисперсией?

□ Как доказывать асимптотическую нормальность? Либо применить ЦПТ напрямую, либо воспользоваться Леммой. Первый вариант не подходит, потому что интересующая оценка не является суммой. Поэтому нужно думать в сторону второго варианта.

Но нам так или иначе все равно нужно где-то применить ЦПТ, чтобы получить асимптотически нормальную последовательность случайных величин, из которой можно будет получить нужную оценку. Скорее всего, читатель быстро догадается, взглянув на оценку $\sqrt[k]{(k+1)\overline{X}^k}$, что можно применить ЦПТ тут только к \overline{X}^k . Сделаем это.

Найдем $E\overline{X}^k = EX_1^k$:

$$EX_1^k = \int_{\mathbb{R}} x^k \frac{1}{\theta} I_{[0,\theta]}(x) dx = \frac{1}{\theta} \frac{x^{k+1}}{k+1} \Big|_0^\theta = \frac{\theta^k}{k+1}.$$

Из ЦПТ получаем, что \overline{X}^k — асимптотически нормальна для $\frac{\theta^k}{k+1}$.

А дальше уже понятно, что нужно применить Лемму с функцией $h(x) = \sqrt[k]{(k+1)x}$ и получим, что оценка $\sqrt[k]{(k+1)\overline{X}^k}$ — асимптотически нормальная оценка θ .

Найдем теперь асимптотическую дисперсию этой оценки. Для этого сначала найдем дисперсию X_1^k :

$$\begin{aligned} DX_1^k &= EX_1^{2k} - (EX_1^k)^2 = \frac{\theta^{2k}}{2k+1} - \left(\frac{\theta^k}{k+1} \right)^2 = \\ &= \theta^{2k} \left(\frac{1}{2k+1} - \frac{1}{(k+1)^2} \right) = \theta^{2k} \frac{k^2 + 2k + 1 - 2k - 1}{(2k+1)(k+1)^2} = \\ &= \frac{k^2}{(2k+1)(k+1)^2} \theta^{2k} \end{aligned}$$

Теперь найдем производную $h(x)$ и подставим $x = EX_1^k$:

$$\begin{aligned} h'(x) &= \left(((k+1)x)^{\frac{1}{k}} \right)' = \frac{k+1}{k} ((k+1)x)^{\frac{1-k}{k}} \\ h' \left(\frac{\theta^k}{k+1} \right) &= \frac{k+1}{k} \left((k+1) \frac{\theta^k}{k+1} \right)^{\frac{1-k}{k}} = \frac{k+1}{k} \theta^{1-k} \end{aligned}$$

Тогда по Лемме асимптотическая дисперсия оценки $\sqrt[k]{(k+1)\overline{X}^k}$ будет равна

$$DX_1^k \cdot \left(h' \left(\frac{\theta^k}{k+1} \right) \right)^2 = \frac{k^2}{(k+1)^2(2k+1)} \theta^{2k} \cdot \frac{(k+1)^2}{k^2} \theta^{2-2k} = \frac{\theta^2}{2k+1}.$$

А значит, при увеличении k асимптотическая дисперсия будет монотонно стремиться к нулю, т.е. среди таких оценок оценки с наименьшей дисперсией не существует. ■

Задача 5. Найти функции, для которых $\sqrt{(X_1^2 + \dots + X_n^2)/n}$, $\sqrt[n]{X_1 \dots X_n}$ и $n/(1/X_1 + \dots + 1/X_n)$ асимптотически нормальные, $X_i \sim R[1, \theta]$.

□ Найдем второй момент

$$EX_i^2 = \int_1^\theta x^2 \frac{1}{\theta-1} dx = \frac{1}{\theta-1} \frac{x^3}{3} \Big|_1^\theta = \frac{\theta^2 + \theta + 1}{3}$$

Значит, по ЦПТ, \overline{X}^2 — а.н. оценка для $\frac{\theta^2 + \theta + 1}{3}$. А по Лемме, $\sqrt{\overline{X}^2}$ — а. н. оценка для $\sqrt{\frac{\theta^2 + \theta + 1}{3}}$.

Далее, $\ln \sqrt[n]{X_1 \dots X_n} = \overline{X}$, это а.н. оценка для EX_1 . Значит, сама $\sqrt[n]{X_1 \dots X_n}$ является а.н. оценкой для e^{EX_1} . Найдем первый момент:

$$EX_1 = \int_1^\theta x \frac{1}{\theta-1} dx = \frac{1}{\theta-1} \frac{x^2}{2} \Big|_1^\theta = \frac{\theta+1}{2}.$$

Итого, $\sqrt[n]{X_1 \cdots X_n}$ есть а.н. оценка для $e^{\frac{\theta+1}{2}}$.

Для третьего случая поступаем аналогично. Находим

$$E \frac{1}{X_1} = \int_1^\theta \frac{1}{x} \frac{1}{\theta-1} dx = \frac{1}{\theta-1} \ln x \Big|_1^\theta = \frac{\ln \theta}{\theta-1}.$$

Значит, $\frac{1}{\bar{X}}$ — а.н. оценка для $\frac{\ln \theta}{\theta-1}$, а тогда по Лемме, $1/\frac{1}{\bar{X}}$ — а.н. для $\frac{\theta-1}{\ln \theta}$. ■

Методы получения оценок

Мы уже порассуждали на тему того, что такое оценивание и определились со свойствами оценок, которые нас интересуют. Методы получения состоятельных и асимптотически нормальных оценок у нас очень ограничены. По сути мы имеем только ЗБЧ и ЦПТ, больше ничего. Но есть и другие способы, с которыми нам предстоит познакомиться.

Метод моментов (Method of moments)

Итак, мы хотим строить состоятельные, а лучше асимптотически нормальные оценки. При этом мы знаем, что \bar{X} является состоятельной оценкой для EX_1 , а в случае $EX_1^2 < \infty$, еще и асимптотически нормальной. Но $EX_1 = a(\theta)$, то бишь некоторый параметр распределения. И тогда, если $a(x)$ — гладкая обратимая функция, то $a^{-1}(\bar{X})$ будет оценкой θ , обладающей теми же свойствами. В этом и заключается метод моментов.

Пример 1. Рассмотрим модель сдвига экспоненциального распределения, т.е. X_i имеют плотность

$$f_\theta(x) = e^{-(x-\theta)} I_{x>\theta}(x).$$

Тогда

$$EX = \int_\theta^\infty x e^{-(x-\theta)} dx = |y = x - \theta| = \theta + \int_0^\infty y e^{-y} dy = \theta + \Gamma(2) = \theta + 1.$$

Таким образом, \bar{X} — состоятельная оценка для $\theta + 1$, следовательно, $\hat{\theta} = \bar{X} - 1$ — состоятельная оценка для θ .

Пример 2. Пусть $X_i \sim \mathcal{N}(0, \theta^2)$. Тогда $EX_1 = 0 = a(\theta)$, а значит, функция a необратима. В таких случаях надо рассмотреть следующий момент $EX_1^2 = \theta^2$. А мы знаем, что по ЦПТ \bar{X}^2 — а.н. оценка для EX_1^2 , значит, по Лемме, $\sqrt{\bar{X}^2}$ — а.н. оценка для θ .

Оценка максимального правдоподобия (Maximum likelihood estimation)

Пусть выборка приняла значение $x = (x_1, \dots, x_n)$. Тогда при каждом θ для дискретных выборок можно вычислить вероятность попадания выборки X_1, \dots, X_n в эти значения, а для абсолютно-непрерывных — плотность в точке (x_1, \dots, x_n) . Чем больше это значение, тем более вероятен результат (x_1, \dots, x_n) . Идея в том, чтобы найти θ , при котором как раз эта вероятность наибольшая.

Опр. Функцией правдоподобия выборки X_1, \dots, X_n в дискретном случае называется совместное распределение

$$L(x_1, \dots, x_n; \theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n),$$

а в абсолютно-непрерывном случае — совместную плотность

$$L(x_1, \dots, x_n; \theta) = f_{X_1, \dots, X_n; \theta}(x_1, \dots, x_n).$$

Опр. Оценкой максимального правдоподобия $\hat{\theta}(X_1, \dots, X_n)$ называется функция $\hat{\theta}(x_1, \dots, x_n)$:

$$L(x_1, \dots, x_n; \hat{\theta}(x_1, \dots, x_n)) \geq L(x_1, \dots, x_n; \theta) \quad \forall x_i, \theta.$$

Пример 1. Рассмотрим схему Бернулли с параметром θ . Функция правдоподобия будет иметь вид

$$L(x_1, \dots, x_n; \theta) = \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - x_1 - \dots - x_n}$$

тогда возьмем логарифм от нее

$$\ln L(x_1, \dots, x_n; \theta) = (x_1 + \dots + x_n) \ln \theta + (n - x_1 - \dots - x_n) \ln(1 - \theta)$$

При $x_1, \dots, x_n \in \{0, 1\}$ дифференцируем по θ и приравняем к нулю, откуда

$$\frac{x_1 + \dots + x_n}{\theta} = \frac{n - x_1 - \dots - x_n}{1 - \theta},$$

то есть $\theta = \bar{x}$. При этом, в этой точке у $\ln L$ максимум, следовательно ОМП будет \bar{X} .

Пример 2. Рассмотрим $X_i \sim \mathcal{N}(\theta_1, \theta_2^2)$. Тогда

$$L(x_1, \dots, x_n; \theta) = \frac{1}{(2\pi\theta_2^2)^{n/2}} \exp \left(- \sum_{i=1}^n \frac{(x_i - \theta_1)^2}{2\theta_2^2} \right)$$

$$\ln L(x_1, \dots, x_n; \theta) = -\frac{n}{2} \ln(2\pi) - n \ln \theta_2 - \sum_{i=1}^n \frac{(x_i - \theta_1)^2}{2\theta_2^2}.$$

Приравняв производные по θ_1 и θ_2 к нулю, получаем уравнения

$$\sum_{i=1}^n \frac{x_i - \theta_1}{\theta_2^2} = 0, \quad -\frac{n}{\theta_2} + \sum_{i=1}^n \frac{(x_i - \theta_1)^2}{\theta_2^3} = 0.$$

Из первого уравнения $\hat{\theta}_1 = \bar{X}$, из второго $\hat{\theta}_2 = S^2$. Это ОМП для θ_1, θ_2 , если в этой точке достигается максимум. Это вытекает из того, что

$$L(x, \theta_1, \theta_2) \leq L(x, \hat{\theta}_1, \theta_2) \leq L(x, \hat{\theta}_1, \hat{\theta}_2)$$

в силу знаков производной: максимум L по первой переменной при любом θ_2 достигается в точке $\hat{\theta}_1$, а при такой первой переменной максимум по второй достигается в $\hat{\theta}_2$.

Оценка методом спейсингов

Опр. Пусть

$$D(x_1, \dots, x_n; \theta) = F(x_{(1)}; \theta) (F(x_{(2)}; \theta) - F(x_{(1)}; \theta)) \cdots (1 - F(x_{(n)}; \theta)).$$

Тогда оценкой методом спейсингов (ОМС) называют функцию $\hat{\theta}(x_1, \dots, x_n)$, такую, что

$$D(x_1, \dots, x_n; \hat{\theta}(x_1, \dots, x_n)) \geq D(x_1, \dots, x_n; \theta) \quad \forall x_i, \theta.$$

Пример 3. Пусть $X_i \sim R[0, \theta]$. Тогда

$$D(x_1, \dots, x_n; \theta) = \frac{x_{(1)}(x_{(2)} - x_{(1)}) \cdots (\theta - x_{(n)})}{\theta^{n+1}} I_{0 < x_{(1)} < \cdots < x_{(n)} < \theta}.$$

Максимизация D равносильна максимизации $\ln(\theta - x_{(n)}) - (n+1) \ln \theta$. Дифференцируем по θ и приравняем к нулю:

$$\frac{1}{\theta - x_{(n)}} - \frac{n+1}{\theta} = 0$$

откуда получаем решение $\theta = \frac{n+1}{n} x_{(n)}$. Причем это действительно максимум. Эта оценка несмещенная в отличие от ОМП.

Задачи

Задача. Пусть $X_i \sim R[\theta, \theta + 1]$. Найти а) ОМП, б) ОМС для θ .

□ а) Имеем плотность $f_{X_i}(x) = I_{[\theta, \theta+1]}$. Тогда Функция правдоподобия

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n I_{[\theta, \theta+1]}(x_i) = \begin{cases} 1, & x_{(1)} \geq \theta, x_{(n)} \leq \theta + 1 \\ 0, & \text{иначе} \end{cases} = \quad (1)$$

$$= I_{[x_{(n)}-1, x_{(1)}]}(\theta) \quad (2)$$

Эта функция достигает своего максимального значения 1 при всех $x_{(n)} - 1 \leq \theta \leq x_{(1)}$. Все они будут ОМП. Например, можно взять

$$\hat{\theta}(X_1, \dots, X_n) = \frac{X_{(1)} + X_{(n)}}{2}.$$

б) Применим метод спейсингов. Напишем по определению функцию

$$D(x_1, \dots, x_n; \theta) = F(x_{(1)}; \theta)(F(x_{(2)}; \theta) - F(x_{(1)}; \theta)) \cdots (1 - F(x_{(n)}; \theta)) = (*)$$

В нашем случае

$$F(x; \theta) = (x - \theta) I_{[\theta, \theta+1]}$$

поэтому

$$\begin{aligned} (*) &= (x_{(1)} - \theta)(x_{(2)} - \theta - x_{(1)} + \theta) \cdots (1 - x_{(n)} + \theta) I_{\theta \leq x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \leq \theta+1} = \\ &= (x_{(1)} - \theta)(x_{(2)} - x_{(1)}) \cdots (1 + \theta - x_{(n)}) I_{\theta \leq x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \leq \theta+1} \end{aligned}$$

Тогда

$$\ln D = \ln(x_{(1)} - \theta) + \ln(x_{(2)} - x_{(1)}) + \cdots + \ln(1 + \theta - x_{(n)})$$

Продифференцируем по θ и приравняем к нулю:

$$-\frac{1}{x_{(1)} - \theta} + \frac{1}{1 + \theta - x_{(n)}} = 0 \Leftrightarrow x_{(1)} - \theta = 1 + \theta - x_{(n)}$$

Отсюда получаем, что

$$\theta = \frac{x_{(1)} + x_{(n)} - 1}{2}.$$

Это максимум, поэтому ОМС будет

$$\hat{\theta} = \frac{X_{(1)} + X_{(n)} - 1}{2}.$$



Задача. Пусть $X_i \sim \exp(\theta, 1)$. Найти а) ОМП, б) ОМС для θ .

□

$$X_i \sim \exp(\theta, 1) \Leftrightarrow f_{X_i}(x; \theta) = e^{-(x-\theta)} I_{x \geq \theta}.$$

1) Таким образом, функция правдоподобия

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n e^{\theta-x_i} I_{x_i \geq \theta} = I_{x_{(1)} \geq \theta} \prod_{i=1}^n e^{\theta-x_i}.$$

Ясно, что при $\theta > x_{(1)}$ значение будет равно нулю, а это, очевидно, не максимальное значение. Поэтому далее рассмотрим случай $\theta \leq x_{(1)}$. Тогда

$$\ln L = \sum_{i=1}^n (\theta - x_i) \Rightarrow \frac{d \ln L}{d\theta} = n,$$

то есть критических точек нет, то есть функция монотонно возрастает с ростом θ . Значит, максимум достигается при $\theta = x_{(1)}$. Тогда ОМП будет $\hat{\theta} = X_{(1)}$.

2) Теперь перейдем к методу спейсингов:

$$D(x_1, \dots, x_n; \theta) = F(x_{(1)}; \theta)(F(x_{(2)}; \theta) - F(x_{(1)}; \theta)) \cdots (1 - F(x_{(n)}; \theta))$$

Функция распределения для сдвинутого экспоненциального распределения есть

$$F(x; \theta) = (1 - e^{-(x-\theta)}) I_{x \geq \theta}.$$

Тогда

$$\begin{aligned} D &= (1 - e^{-(x_{(1)}-\theta)})(1 - e^{-(x_{(2)}-\theta)} - 1 + e^{-(x_{(1)}-\theta)}) \cdots (1 - 1 + e^{-(x_{(n)}-\theta)}) I_{x_{(1)} \geq \theta} = \\ &= (1 - e^{-(x_{(1)}-\theta)}) e^{\theta} (e^{-x_{(1)}} - e^{-x_{(2)}}) \cdots e^{\theta} e^{-x_{(n)}} I_{x_{(1)} \geq \theta} = \\ &= e^{n\theta} (1 - e^{-(x_{(1)}-\theta)}) (e^{-x_{(1)}} - e^{-x_{(2)}}) \cdots e^{-x_{(n)}} I_{x_{(1)} \geq \theta} \end{aligned}$$

Отсюда берем логарифм (далее считаем, что $x_{(1)} \geq \theta$)

$$\ln D = n\theta + \ln(1 - e^{-(x_{(1)}-\theta)}) + \ln(e^{-x_{(1)}} - e^{-x_{(2)}}) + \cdots - x_{(n)}.$$

Производная по θ всех слагаемых, кроме первых двух, равна нулю, поэтому, находя производную и приравнявая ее к нулю, получаем

$$\frac{d \ln D}{d\theta} = n - \frac{e^{-(x_{(1)}-\theta)}}{1 - e^{-(x_{(1)}-\theta)}} = 0.$$

Далее решаем уравнение:

$$\begin{aligned} n &= \frac{e^{-(x_{(1)}-\theta)}}{1 - e^{-(x_{(1)}-\theta)}} \\ n - ne^{-(x_{(1)}-\theta)} &= e^{-(x_{(1)}-\theta)} \\ n &= (n+1)e^{-(x_{(1)}-\theta)} \\ e^{-(x_{(1)}-\theta)} &= \frac{n}{n+1} \\ -x_{(1)} + \theta &= \ln \frac{n}{n+1} \\ \theta &= x_{(1)} + \ln \frac{n}{n+1} \end{aligned}$$

Это максимум (можно проверить через вторую производную), поэтому оценка $\hat{\theta} = X_{(1)} + \ln \frac{n}{n+1}$ является ОМС. ■

Задача. Углы четырехугольника $ABCD$ многократно (по n раз каждый) измерены с независимыми погрешностями $\mathcal{N}(0, \sigma^2)$. Найти ОМП для углов и σ^2 . Выразить ответ через выборочные средние и выборочные дисперсии наблюдения для каждого угла.

□ У нас есть $4n$ измерений, по n на каждый угол четырехугольника. Каждое измерение есть случайная величина $X_{i,j}$, где i — индекс угла, j — номер измерения. Они имеют вид

$$X_{i,j} = \theta_i + \mathcal{N}(0, \sigma^2) = \mathcal{N}(\theta_i, \sigma^2)$$

И тогда, казалось бы, все сводится к оценке параметров нормального распределения, которое мы умеем оценивать. Но тут есть ограничение: сказано, что θ_i — углы четырехугольника, а значит, накладывается еще условие

$$\theta_1 + \theta_2 + \theta_3 + \theta_4 = 2\pi.$$

Поэтому это нужно будет учесть. А именно, напишем функцию правдоподобия

$$L(x_{i,j}; \theta_i; \sigma^2) = \prod_{i=1}^4 \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_{i,j} - \theta_i)^2}{2\sigma^2}}$$

Возьмем логарифм:

$$\ln L = \sum_{i=1}^4 \sum_{j=1}^n \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_{i,j} - \theta_i)^2}{2\sigma^2} \right) = -2n \ln(2\pi) - 2n \ln \sigma^2 - \sum_{i=1}^4 \sum_{j=1}^n \frac{(x_{i,j} - \theta_i)^2}{2\sigma^2}$$

Дальше надо найти максимум этой функции. Но в данной задаче накладывается условие на углы, поэтому нужно искать условный максимум. Для этого есть метод Лагранжа. По этому методу нужно выписать функцию Лагранжа

$$\mathcal{L} = \ln L + \lambda(\theta_1 + \theta_2 + \theta_3 + \theta_4 - 2\pi)$$

и решить систему уравнений

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \theta_i} = 0, & i = 1, 2, 3, 4 \\ \frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \theta_1 + \theta_2 + \theta_3 + \theta_4 - 2\pi = 0 \end{cases}$$

Найдем производные по θ_i :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_i} &= \sum_{j=1}^n \frac{x_{i,j} - \theta_i}{\sigma^2} + \lambda = \frac{n}{\sigma^2} (\bar{x}_i - \theta_i) + \lambda = 0 \\ \theta_i &= \bar{x}_i + \frac{\sigma^2}{n} \lambda \end{aligned}$$

Теперь подставим в условие на сумму углов

$$\sum_{i=1}^4 \bar{x}_i + \frac{4\sigma^2}{n} \lambda = 2\pi \Rightarrow \lambda = \frac{n}{4\sigma^2} \left(2\pi - \sum_{i=1}^4 \bar{x}_i \right)$$

Подставим обратно в выражения для θ_i :

$$\boxed{\hat{\theta}_i = \bar{x}_i + \frac{1}{4} \left(2\pi - \sum_{i=1}^4 \bar{x}_i \right)}$$

То есть добавляется усредненное отклонение суммы углов от 2π .

Теперь найдем оценку для σ^2 :

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \sigma^2} &= \frac{\partial \ln L}{\partial \sigma^2} = -\frac{2n}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^4 \sum_{j=1}^n (x_{i,j} - \theta_i)^2 = 0 \quad | \cdot \sigma^4 \\ &- 2n\sigma^2 + \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^n (x_{i,j} - \theta_i)^2 = 0 \\ \hat{\sigma}^2 &= \frac{1}{4n} \sum_{i=1}^4 \sum_{j=1}^n (x_{i,j} - \hat{\theta}_i)^2\end{aligned}$$

Теперь остается только выразить σ^2 через выборочные средние и выборочные дисперсии (обозначим $s = \frac{1}{n} \sum (x_i - \bar{x})^2$).

Для этого обозначим $\alpha = \frac{1}{4} (2\pi - \sum_{i=1}^4 \bar{x}_i)$, тогда

$$\hat{\theta}_i = \bar{x}_i + \alpha.$$

Тогда

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{4n} \sum_{i=1}^4 \sum_{j=1}^n (x_{i,j} - \bar{x}_i - \alpha)^2 = \\ &= \frac{1}{4n} \sum_{i=1}^4 \sum_{j=1}^n ((x_{i,j} - \bar{x}_i)^2 - 2\alpha(x_{i,j} - \bar{x}_i) + \alpha^2) = \\ &= \frac{1}{4} \sum_{i=1}^4 (s_i - 2\alpha \underbrace{\sum_{j=1}^n (x_{i,j} - \bar{x}_i)}_{=0} + \alpha^2) = \frac{1}{4} \sum_{i=1}^4 s_i + \alpha^2\end{aligned}$$

То есть, если суммировать, то получаем

$$\hat{\sigma}^2 = \frac{1}{4} \sum_{i=1}^4 s_i + \left(\frac{2\pi - \sum_{i=1}^4 \bar{x}_i}{4} \right)^2$$

■

Задача. Для оценки числа N рыб в пруду пометили m_1 рыб, затем отпустили, а через несколько дней выловили m_2 рыб и подсчитали число μ помеченных рыб. Найти ОММ и ОМП для N . Выборка состоит из одного наблюдения.

□ Сперва нужно понять, что здесь случайная величина, а что параметр. Итак, случайное наблюдение здесь — число μ помеченных рыб среди выловленных m_2 .

Случайной величине μ соответствует количество "успехов" при выборе m_2 элементов из N -элементного множества без возвращения, количество успешных вариантов — m_1 . Распределение такой случайной величины называется гипергеометрическим:

$$\mu \sim \text{HyperGeom}(N, m_1, m_2)$$

а, значит,

$$P(\mu = k) = \frac{C_{m_1}^k C_{N-m_1}^{m_2-k}}{C_N^{m_2}}.$$

Итак, параметры m_1, m_2 известны, нужно оценить параметр N .

Построим функцию правдоподобия по одному наблюдению:

$$L(k; N) = P(\mu = k) = \frac{C_{m_1}^k C_{N-m_1}^{m_2-k}}{C_N^{m_2}}.$$

То есть буквально переписали формулу. Теперь нужно максимизировать $L(k; N)$ по N . Для этого применяется трюк такой. Рассмотрим отношение $L(k; N)$ к $L(k; N-1)$:

$$R(k; N) = \frac{L(k; N)}{L(k; N-1)}$$

и найдем, в каких точках это отношение равняется единице. Если только в одной, то мы найдем глобальный максимум.

Итак,

$$\begin{aligned} R(k; N) &= \frac{C_{N-m_1}^{m_2-k}}{C_N^{m_2}} \cdot \frac{C_{N-1}^{m_2}}{C_{N-1-m_1}^{m_2-k}} = \\ &= \frac{(N-m_1)!}{(N-m_1-m_2+k)!} \frac{m_2!(N-m_2)!}{N!} \cdot \\ &\quad \cdot \frac{(N-1)!}{m_2!(N-1-m_2)!} \frac{(m_2-k)!(N-1-m_1-m_2+k)!}{(N-1-m_1)!} = \\ &= \frac{(N-m_1)(N-m_2)}{(N-m_1-m_2+k)N} \end{aligned}$$

Теперь

$$\begin{aligned} R(k; N) \leq 1 &\Leftrightarrow \frac{(N-m_1)(N-m_2)}{(N-m_1-m_2+k)N} \leq 1 \Leftrightarrow \\ &\Leftrightarrow N^2 - Nm_1 - Nm_2 + m_1m_2 \leq N^2 - Nm_1 - Nm_2 + Nk \Leftrightarrow \\ &\Leftrightarrow N \geq \frac{m_1m_2}{k} \end{aligned}$$

То есть, начиная с $N = \lfloor \frac{m_1m_2}{k} \rfloor + 1$ функция правдоподобия не возрастает. При этом вплоть до $N = \lfloor \frac{m_1m_2}{k} \rfloor$ — не убывает. Поэтому ОМП здесь будет

$$\hat{N} = \left\lfloor \frac{m_1m_2}{\mu} \right\rfloor$$

См. обсуждения по ссылке на MathStackExchange.

Теперь построим ОММ. Для этого надо найти $E\mu$. Понятно, что считать в лоб не хочется, ибо имеем выражение

$$E\mu = \sum_{k=0}^{m_2} kP(\mu = k) = \sum_{k=0}^{m_2} k \frac{C_{m_1}^k C_{N-m_1}^{m_2-k}}{C_N^{m_2}}.$$

Поэтому разобьем в сумму $\mu = X_1 + \dots + X_{m_2}$, где X_i — i -ая рыба помеченная или не помеченная. То есть $X_i \sim \text{Bern}(p_i)$. Кто-то сразу скажет, чему равны p_i , но на мой взгляд ответ не очевиден. Найдем, чему же они равны.

Итак, $p_1 = m_1/N$, это очевидно. А что с p_2 ?

$$\begin{aligned} P(X_2 = 1) &= P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) + P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) = \\ &= \frac{m_1}{N-1} \left(1 - \frac{m_1}{N}\right) + \frac{m_1-1}{N-1} \frac{m_1}{N} = \frac{m_1}{N} \end{aligned}$$

То есть $X_i \sim \text{Bern}(\frac{m_1}{N})$.

Значит, $E\mu = \sum EX_i = m_2 EX_1 = m_2 \frac{m_1}{N}$. По ЗБЧ, оценка $\bar{\mu} \equiv \mu$ является состоятельной для $E\mu = a(N) = \frac{m_1 m_2}{N}$, откуда

$$\hat{N} = \frac{m_1 m_2}{\mu}.$$

Обсуждения можно почитать тут и тут.

А в целом эта задача имеет самые разные применения и об этом методе подсчета популяции есть статья на википедии Mark and recapture. ■

Информация

Чисто задачи по теории вероятностей

Сходимости (Convergences)

Задача. Пусть $X_i \sim R[0, 1]$ — н.о.р. Найти предел по распределению $n \min_{i \leq n} X_i$.

□ Обозначим $Y_n = n \min_{i \leq n} X_i$. Найдем функцию распределения Y_n :

$$\begin{aligned} F_{Y_n}(x) &= P(n \min_{i \leq n} X_i \leq x) = P\left(\min_{i \leq n} X_i \leq \frac{x}{n}\right) = \\ &= P\left(\exists i \leq n : X_i \leq \frac{x}{n}\right) = P\left(\overline{\forall i \leq n : X_i > \frac{x}{n}}\right) = \\ &= 1 - P\left(\forall i \leq n : X_i > \frac{x}{n}\right) = 1 - P\left(X_1 > \frac{x}{n}\right)^n = \\ &= 1 - \left(1 - F_{X_1}\left(\frac{x}{n}\right)\right)^n = 1 - \left(1 - \frac{x}{n}\right)^n \xrightarrow{n \rightarrow \infty} 1 - e^{-x}. \end{aligned}$$

Таким образом, $Y_n \xrightarrow{d} \text{Exp}(1)$. ■

Задача. Показать, что для сходимости по вероятности $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ следует $X_n Y_n \xrightarrow{P} XY$. Верно ли это для сходимости по распределению?

□ Покажем, что $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY$. Для этого нужно показать, что $\forall \varepsilon > 0 \quad P(|X_n Y_n - XY| > \varepsilon) \rightarrow 0$.

Когда видим подобную разность, сразу хочется прибавить и вычесть $X_n Y$, чтобы сгруппировались более удобные для оценки члены:

$$|X_n Y_n + X_n Y - X_n Y - XY| = |X_n(Y_n - Y) + Y(X_n - X)| \leq |X_n(Y_n - Y)| + |Y(X_n - X)|$$

Тогда

$$P(|X_n Y_n - XY| > \varepsilon) \leq P(|X_n(Y_n - Y)| + |Y(X_n - X)| > \varepsilon)$$

Что делать с этим? Хочется разбить в сумму двух вероятностей, но так просто это сделать нельзя. Нужен трюк. Воспользуемся тем, что если $a + b > c$, то $a > c/2$ или $b > c/2$. Поэтому верно вложение событий

$$\{|X_n(Y_n - Y)| + |Y(X_n - X)| > \varepsilon\} \subseteq \{|X_n(Y_n - Y)| > \varepsilon/2\} \cup \{|Y(X_n - X)| > \varepsilon/2\},$$

а из него очевидно следует неравенство

$$P(|X_n(Y_n - Y)| + |Y(X_n - X)| > \varepsilon) \leq P(|X_n(Y_n - Y)| > \varepsilon/2) + P(|Y(X_n - X)| > \varepsilon/2)$$

Теперь нужно оценить каждое из двух получившихся неравенств, чтобы показать, что они оба стремятся к нулю.

Рассмотрим $P(|X_n(Y_n - Y)| > \varepsilon/2)$. Заметим, что тут мы по сути хотим доказать лемму, что если $X_n \xrightarrow{P} X$, $Z_n \xrightarrow{P} 0$, то $X_n Z_n \xrightarrow{P} 0$. Но останемся в наших терминах. Итак, нужно воспользоваться исходными сходимостями, применив формулу полной вероятности.

$$\begin{aligned} P(|X_n(Y_n - Y)| > \varepsilon/2) &= P(|X_n(Y_n - Y)| > \varepsilon/2 \mid |Y_n - Y| < \delta)P(|Y_n - Y| < \delta) + \\ &\quad + P(|X_n(Y_n - Y)| > \varepsilon/2 \mid |Y_n - Y| > \delta)P(|Y_n - Y| > \delta) \end{aligned}$$

Второе слагаемое стремится к нулю, а $P(|Y_n - Y| < \delta) \rightarrow 1$. Поэтому нужно оценить

$$P(|X_n(Y_n - Y)| > \varepsilon/2 \mid |Y_n - Y| < \delta) \leq P(\delta |X_n| > \varepsilon/2) = P\left(|X_n| > \frac{\varepsilon}{2\delta}\right)$$

Тут с одной стороны не понятно, что дальше делать. А с другой, ничего не остается как воспользоваться сходимостью X_n к X . Как воспользоваться опять не понятно, но опять вариант только один — формула полной вероятности.

$$P\left(|X_n| > \frac{\varepsilon}{2\delta}\right) = P\left(|X_n| > \frac{\varepsilon}{2\delta} \mid |X - X_n| > \gamma\right) P(|X_n - X| > \gamma) + \\ + P\left(|X_n| > \frac{\varepsilon}{2\delta} \mid |X - X_n| \leq \gamma\right) P(|X_n - X| \leq \gamma)$$

Первое слагаемое стремится к нулю по условию. Во втором $P(|X_n - X| \leq \gamma) \rightarrow 1$. Таким образом, остается посмотреть на

$$P\left(|X_n| > \frac{\varepsilon}{2\delta} \mid |X - X_n| \leq \gamma\right) = P\left(|X_n - X + X| > \frac{\varepsilon}{2\delta} \mid |X - X_n| > \gamma\right) \leq \\ \leq P\left(|X_n - X| + |X| > \frac{\varepsilon}{2\delta} \mid |X - X_n| > \gamma\right) \leq \\ \leq P\left(\gamma + |X| > \frac{\varepsilon}{2\delta}\right) = P\left(|X| > \frac{\varepsilon}{2\delta} - \gamma\right)$$

Так как γ и δ — любые положительные числа, то число можно сделать эту вероятность сколько угодно малой, увеличивая число $\frac{\varepsilon}{2\delta} - \gamma$. Таким образом,

$$P(|X_n(Y_n - Y)| > \varepsilon/2) \rightarrow 0.$$

Осталось показать, что $P(|Y(X_n - X)| > \varepsilon/2) \rightarrow 0$. Это делается аналогично, но чуть легче.

Идея решения взята отсюда.

При этом для сходимости по распределению утверждение неверно. Возьмем $X_n = Y_n = X \sim \mathcal{N}(0, 1)$ и $Y \sim \mathcal{N}(0, 1)$. Тогда $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, но $X_n Y_n = X^2$ не сходится по распределению к XY , так как

$$P(X^2 < 0) = 0, \quad P(XY < 0) = \frac{1}{2}.$$

Пример взят отсюда. ■

ЦПТ и ЗБЧ (CLT and LLN)

Таблица со значениями функции распределения нормальной случайной величины $\mathcal{N}(0, 1)$ — здесь.

Задача. Вероятность выпуска бракованной скрепки 0.02. Скрепки продаются упаковками по 100 штук. Какая (приблизительно) вероятность того, что в двух упаковках будет хотя бы одна бракованная скрепка? Что 102 упаковок хватит для набора 10000 целых скрепок?

□ Пусть X_i — состояние i -ой скрепки, т.е. $X_i \sim \text{Bern}(0.98)$. Тогда

$$EX_i = 0.98, \quad DX_i = p(1 - p) = 0.98 \cdot 0.02 = 0.0196.$$

Нам надо найти

$$P\left(\sum_{i=1}^{200} X_i \leq 199\right) = P\left(\frac{\sum_{i=1}^{200} X_i - 200 \cdot 0.98}{\sqrt{0.0196 \cdot 200}} \leq \frac{199 - 200 \cdot 0.98}{\sqrt{0.0196 \cdot 200}}\right) \approx \Phi(1.515) \approx 0.93$$

Во втором вопросе нужно найти

$$\begin{aligned} P\left(\sum_{i=1}^{10200} X_i \geq 10000\right) &= 1 - P\left(\sum_{i=1}^{10200} X_i \leq 9999\right) = \\ &= 1 - P\left(\frac{\sum_{i=1}^{10200} X_i - 10200 \cdot 0.98}{\sqrt{0.0196 \cdot 10200}} \leq \frac{9999 - 10200 \cdot 0.98}{\sqrt{0.0196 \cdot 10200}}\right) \approx \\ &\approx 1 - \Phi(0.212) \approx 1 - 0.58 = 0.42. \end{aligned}$$

■

Задача. Пусть $\xi_i \sim R[1, 3]$ — н.о.р. Найти предел п.н. а) среднего арифметического ξ_i , б) среднего гармонического ξ_i , в) среднего геометрического ξ_i .

□ Обозначим

$$\begin{aligned} A(x_1, \dots, x_n) &= \frac{1}{n} \sum_{i=1}^n x_i \quad \text{среднее арифметическое} \\ H(x_1, \dots, x_n) &= \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad \text{среднее гармоническое} \\ G(x_1, \dots, x_n) &= \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \quad \text{среднее геометрическое} \end{aligned}$$

Пункт (а). Из УЗБЧ знаем, что $A(\xi_1, \dots, \xi_n) \xrightarrow{\text{п.н.}} E\xi_1 = 2$.

Пункт (б). Нужно сводить к УЗБЧ, т.е. выразить среднее гармоническое через среднее арифметическое. Понятно как:

$$H(x_1, \dots, x_n) = \frac{1}{A\left(\frac{1}{x_1}, \dots, \frac{1}{x_n}\right)},$$

поэтому

$$H(\xi_1, \dots, \xi_n) \xrightarrow{\text{п.н.}} \frac{1}{E\left(\frac{1}{\xi_1}\right)}.$$

Вычислим

$$E\left(\frac{1}{\xi_1}\right) = \frac{1}{3-1} \int_1^3 \frac{1}{x} dx = \frac{1}{2} \ln 3$$

Таким образом,

$$H(\xi_1, \dots, \xi_n) \xrightarrow{\text{п.н.}} \frac{2}{\ln 3}$$

Пункт (в). Заметим, что $\ln G(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \ln x_i = A(\ln x_1, \dots, \ln x_n)$, т.е.

$$G(\xi_1, \dots, \xi_n) = e^{A(\ln \xi_1, \dots, \ln \xi_n)} \xrightarrow{\text{п.н.}} e^{E \ln \xi_1}.$$

Вычислим

$$E(\ln \xi_1) = \frac{1}{2} \int_1^3 \ln x dx = \frac{1}{2} \left(x \ln x \Big|_1^3 - \int_1^3 dx \right) = \frac{3 \ln 3 - 2}{2}.$$

Таким образом,

$$G(\xi_1, \dots, \xi_n) \xrightarrow{\text{п.н.}} e^{\frac{3 \ln 3 - 2}{2}}$$



Задача. В подъезде горит лампочка. Если лампочка сгорает, то ее тут же меняют на новую. Лампочки работают н.о.р. времена T_i , $ET_i = a$. Пусть N_t — число сгоревших за время t лампочек. Доказать, что

$$\frac{N_t}{t} \xrightarrow{P} \frac{1}{a}, \quad t \rightarrow +\infty.$$

(В теории случайных процессов N_t называется Пуассоновским процессом, а это утверждение является аналогом ЗБЧ для Пуассоновского процесса).

□ Здесь нужно применить какую-то из предельных теорем. ЦПТ здесь явно не помогает, поэтому попробуем применить ЗБЧ. Для этого нам нужно перейти к рассмотрению сумм случайных величин.

У нас из условия есть времена T_i . Обозначим $S_n = \sum_{i=0}^n T_i$. Легко понять, что

$$S_{N_t} \leq t < S_{N_t+1}$$

Поделим на N_t :

$$\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t}$$

Тут мы уже видим, что нужно применить лемму о двух полицейских, показав сходимость к одному числу левого и правого выражений. По УЗБЧ

$$\frac{S_{N_t}}{N_t} \xrightarrow{\text{п.н.}} a$$

Теперь рассмотрим

$$\frac{S_{N_t+1}}{N_t} = \frac{S_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t} = \frac{S_{N_t+1}}{N_t+1} \cdot \left(1 + \frac{1}{N_t}\right) \xrightarrow{\text{п.н.}} a.$$

И по лемме о двух полицейских

$$\frac{t}{N_t} \xrightarrow{\text{п.н.}} a \Rightarrow \frac{N_t}{t} \xrightarrow{\text{п.н.}} \frac{1}{a}.$$

Мы доказали более сильный результат, используя УЗБЧ. Если бы использовали ЗБЧ, то получили бы ровно то, что просили. ■

▷ Приведем другой вариант доказательства предыдущего утверждения. Это доказательство взято из лекций по случайным процессам В.И. Афанасьева. Оно выглядит сложнее в плане значков, но зато не подразумевает знания леммы о двух полицейских (и вообще идеи о том, что к ней нужно сводить задачу), утверждения о том, что $X_n \xrightarrow{P} c \Rightarrow \frac{1}{X_n} \xrightarrow{P} \frac{1}{c}$, и о том, что $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY$.

Итак, мы хотим доказать, что $\frac{N_t}{t} \xrightarrow{P} \frac{1}{a}$. По определению это означает, что

$$P\left(\left|\frac{N_t}{t} - \frac{1}{a}\right| > \varepsilon\right) \rightarrow 0, \quad t \rightarrow +\infty.$$

Ясно, что

$$P\left(\left|\frac{N_t}{t} - \frac{1}{a}\right| > \varepsilon\right) = P\left(\frac{N_t}{t} - \frac{1}{a} > \varepsilon\right) + P\left(\frac{N_t}{t} - \frac{1}{a} < -\varepsilon\right)$$

Нужно показать, что оба слагаемых стремятся к нулю. Рассмотрим первое.

$$P\left(\frac{N_t}{t} - \frac{1}{a} > \varepsilon\right) = P\left(N_t > t\left(\varepsilon + \frac{1}{a}\right)\right) = P(N_t > k(t)),$$

где $k(t) = \lceil t(\varepsilon + \frac{1}{a}) \rceil$. Далее

$$P(N_t > k(t)) = P(S_{k(t)} < t) = P\left(\frac{S_{k(t)}}{k(t)} - a < \frac{t}{k(t)} - a\right)$$

Распишем теперь

$$\frac{t}{k(t)} - a \leq \frac{t}{t(\varepsilon + \frac{1}{a})} - a = -\frac{a^2\varepsilon}{a\varepsilon + 1} =: -\varepsilon_1.$$

Поэтому

$$P\left(\frac{S_{k(t)}}{k(t)} - a < \frac{t}{k(t)} - a\right) \leq P\left(\frac{S_{k(t)}}{k(t)} - a < -\varepsilon_1\right) \rightarrow 0, \quad t \rightarrow +\infty,$$

что следует из ЗБЧ.

То есть мы доказали, что

$$P\left(\frac{N_t}{t} - \frac{1}{a} > \varepsilon\right) \rightarrow 0, \quad t \rightarrow +\infty,$$

аналогично и второе слагаемое стремится к нулю. Откуда следует требуемый результат. \triangleleft

Задача. Не производя вычислений, доказать, что

$$e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \rightarrow \frac{1}{2}, \quad \sum_{k=0}^n C_{n+k-1}^k 2^{-n-k} \rightarrow \frac{1}{2}$$

при $n \rightarrow \infty$.

□ Вспомним распределение Пуассона с параметром λ :

$$X \sim Poiss(\lambda) \Rightarrow P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

Поэтому можно заметить, что если $X \sim Poiss(n)$, то

$$e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \sum_{k=0}^n P(X = k) = P(X \leq n).$$

Но если взять $X_1, \dots, X_n \sim Poiss(1)$, то $X \stackrel{d}{=} X_1 + \dots + X_n$, а значит, используя ЦПТ,

$$P(X \leq n) = P(X_1 + \dots + X_n \leq n) = P\left(\left(\frac{S_n}{n} - 1\right) \frac{\sqrt{n}}{\sqrt{1}} \leq 0\right) \rightarrow \Phi(0) = \frac{1}{2}.$$

Теперь докажем вторую формулу. Здесь нужно понять, какой эксперимент соответствует величине X с распределением

$$P(X = k) = C_{n+k-1}^k 2^{n+k}.$$

Можно увидеть, что некий бернуллиевский эксперимент с параметром $p = 1/2$ производится $n + k$ раз, из них $n + k - 1$ результатов могут варьироваться так, что k из них выпадает один из двух вариантов.

Такое распределение называется *отрицательным биномиальным распределением* или *распределением Паскаля*, обозначается $NB(n, p)$. Случайная величина с таким распределением показывает количество неуспехов до n -го успеха.

Тогда мы можем понять, что исходную величину можно представить в виде суммы $X_i \sim \text{Geom}(\frac{1}{2})$:

$$X = X_1 + \dots + X_n.$$

Вспомним, как для $X_i \sim \text{Geom}(p)$ выражаются математическое ожидание и дисперсия:

$$EX_i = \frac{1-p}{p} \Big|_{p=\frac{1}{2}} = 1, \quad DX_i = \frac{1-p}{p^2} \Big|_{p=\frac{1}{2}} = 2.$$

Теперь можем применять ЦПТ:

$$P(X \leq n) = P(S_n \leq n) = P\left(\left(\frac{S_n}{n} - 1\right) \frac{\sqrt{n}}{\sqrt{2}} \leq 0\right) \rightarrow \Phi(0) = \frac{1}{2}.$$

■

Задача. Показать, что в ЦПТ не может выполняться сходимость по вероятности.

□ В ЦПТ мы имеем утверждение о том, что если X_i — н.о.р. величины с конечной дисперсией, то

$$\left(\frac{S_n}{n} - EX_1\right) \frac{\sqrt{n}}{\sqrt{DX_1}} \xrightarrow{d} U \sim \mathcal{N}(0, 1).$$

Вопрос в том, можно ли сходимость по распределению заменить на более сильную сходимость по вероятности. Утверждается, что нельзя. Для доказательства этого достаточно привести пример, когда сходимость по вероятности не выполняется.

Рассмотрим $X_i \sim \mathcal{N}(0, 1)$ (можно и другие, главное, чтобы $EX_i = 0$, $DX_i = 1$). Тогда утверждение ЦПТ упростится:

$$Y_n := \frac{S_n}{\sqrt{n}} \xrightarrow{d} U \sim \mathcal{N}(0, 1).$$

Покажем, используя критерий Коши, что данная последовательность по вероятности расходится. Для этого нужно показать, что

$$\forall \varepsilon > 0 \quad P(|Y_n - Y_m| > \varepsilon) \not\rightarrow 0, \quad n, m \rightarrow \infty.$$

Будем брать $m = 2n$, тогда

$$\begin{aligned}
P(|Y_{2n} - Y_n| > \varepsilon) &= P\left(\left|\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}}\right| > \varepsilon\right) = \\
&= P\left(\left|\frac{1}{\sqrt{2n}} \sum_{i=1}^{2n} X_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right| > \varepsilon\right) = \\
&= P\left(\left|\frac{1}{\sqrt{2n}} \sum_{i=n+1}^{2n} X_i - \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{2n}}\right) \sum_{i=1}^n X_i\right| > \varepsilon\right) \geq \\
&\geq [\text{воспользуемся неравенством } |a - b| \geq |a| - |b|] \geq \\
&\geq P\left(\frac{1}{\sqrt{2n}} \left|\sum_{i=n+1}^{2n} X_i\right| - \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{2n}}\right) \left|\sum_{i=1}^n X_i\right| > \varepsilon\right) \geq [\text{трюк}] \geq \\
&\geq P\left(\frac{1}{\sqrt{2n}} \left|\sum_{i=n+1}^{2n} X_i\right| > 2\varepsilon, \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{2n}}\right) \left|\sum_{i=1}^n X_i\right| < \varepsilon\right) \\
&= [\text{исп. независимость}] = \\
&= P\left(\frac{1}{\sqrt{2n}} \left|\sum_{i=n+1}^{2n} X_i\right| > 2\varepsilon\right) P\left(\left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{2n}}\right) \left|\sum_{i=1}^n X_i\right| < \varepsilon\right) = \\
&= P\left(\frac{1}{\sqrt{2n}} \left|\sum_{i=1}^n X_i\right| > 2\varepsilon\right) P\left(\left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{2n}}\right) \left|\sum_{i=1}^n X_i\right| < \varepsilon\right) = \\
&= P\left(\frac{1}{\sqrt{n}} |S_n| > 2\sqrt{2}\varepsilon\right) P\left(\frac{1}{\sqrt{n}} |S_n| > \frac{\varepsilon}{1 - \frac{1}{\sqrt{2}}}\right) \rightarrow \\
&\rightarrow 2\Phi(-2\sqrt{2}\varepsilon) \left(1 - 2\Phi\left(-\frac{\varepsilon}{1 - \frac{1}{\sqrt{2}}}\right)\right) > 0.
\end{aligned}$$

Таким образом, предела по вероятности у последовательности $\frac{S_n}{\sqrt{n}}$ не существует. Доказательство взято отсюда. ■

Задача. Верен ли ЗБЧ для Y_i , где а) $Y_i = iX_i$,

$$\frac{1 - P(X_i = 0)}{2} = P(X_i) = P(X_i = -1) = \frac{1}{(2i)^2},$$

$$\text{б) } P(Y_i = 2^{2^i}) = P(Y_i = -2^{2^i}) = \frac{1}{2}.$$

□ Пункт (а). Вычислим EY_i :

$$EY_i = iEX_i = i\left(-1 \cdot \frac{1}{(2i)^2} + 1 \cdot \frac{1}{(2i)^2}\right) = 0$$

Далее, $DY_i = EY_i^2 - (EY_i)^2 = i^2 EX_i^2 = i^2 \cdot 2 \cdot \frac{1}{(2i)^2} = \frac{1}{2} < \infty$. Таким образом, применим ЗБЧ. Это легко показать, применив неравенство Чебышева:

$$P\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{D(S_n/n)}{\varepsilon} = \frac{n/2}{n^2\varepsilon} = \frac{1}{2n\varepsilon} \rightarrow 0.$$

Пункт (б). Здесь тоже $EY_i = 0$, но дисперсия уже неограниченная: $DY_i = EY_i^2 = \left(2^{2^i}\right)^2 = 2^{2^{i+1}}$. Предположим, что тем не менее сходимость $\frac{S_n}{n} \xrightarrow{P} 0$ выполняется.

Тогда выполняется и сходимост $\frac{S_n}{n} \xrightarrow{d} 0$. Но это очевидно не так (см. графики функций распределения). А значит имеем противоречие. ■

Задача. Доказать, что если $\frac{N_n}{n} \xrightarrow{P} a$ и N_n целочисленна и не зависит от X_i — н.о.р. с $EX_i = 0$, $DX_i = 1$, то

$$\frac{S_{N_n}}{\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0, a).$$

□ **Доказательство 1.** Представим

$$\frac{S_{N_n}}{\sqrt{n}} = \frac{S_{N_n}}{\sqrt{N_n}} \cdot \frac{\sqrt{N_n}}{\sqrt{n}}$$

Покажем, что первый множитель сходится по распределению к $\mathcal{N}(0, 1)$. При этом второй множитель по условию сходится к \sqrt{a} по вероятности. Отсюда, используя лемму Слуцкого получим требуемый результат.

Итак, пусть $\varepsilon > 0$. Тогда из ЦПТ следует, что найдется такой номер N_ε , что

$$\forall n > N_\varepsilon \quad \left| P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x \mid N_n > N_\varepsilon\right) - \Phi(x) \right| < \varepsilon/2.$$

Но нам нужно избавиться от этого условия, чтобы получить требуемое, а именно, что найдется N_ε :

$$\forall n > N_\varepsilon \quad \left| P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x\right) - \Phi(x) \right| < \varepsilon.$$

Для этого распишем, используя формулу полной вероятности,

$$\begin{aligned} P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x\right) - \Phi(x) &= P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x \mid N_n > N_\varepsilon\right) P(N_n > N_\varepsilon) - \Phi(x)P(N_n > N_\varepsilon) + \\ &+ P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x \mid N_n \leq N_\varepsilon\right) P(N_n \leq N_\varepsilon) - \Phi(x)P(N_n \leq N_\varepsilon) < \\ &< \frac{\varepsilon}{2} + \left(P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x \mid N_n \leq N_\varepsilon\right) - \Phi(x) \right) P(N_n \leq N_\varepsilon) = (*) \end{aligned}$$

Здесь мы использовали оценку, которую получили выше. Теперь осталось оценить второе слагаемое из полученного выражения. Для начала заметим, что

$$P\left(\frac{S_{N_n}}{\sqrt{N_n}} \leq x \mid N_n \leq N_\varepsilon\right) - \Phi(x) \leq 1.$$

Поэтому нужно лишь показать, что $P(N_n \leq N_\varepsilon) < \varepsilon/2$.

По условию $\frac{N_n}{n} \xrightarrow{P} a$, то есть

$$\forall \varepsilon > 0 \quad P\left(\left|\frac{N_n}{n} - a\right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty.$$

Но заметим, что

$$P\left(\left|\frac{N_n}{n} - a\right| > \varepsilon\right) \geq P\left(\frac{N_n}{n} - a < -\varepsilon\right) = P(N_n < n(a - \varepsilon)) \geq P(N_n < N_\varepsilon)$$

Последнее неравенство верно для больших n (но именно они нас и интересуют). Отсюда $P(N_n \leq N_\varepsilon) \rightarrow 0$ при $n \rightarrow \infty$, т.е. начиная с некоторого номера $P(N_n \leq N_\varepsilon) < \varepsilon/2$. Таким образом, $(*) < \varepsilon$, а значит мы показали, что

$$\frac{S_{N_n}}{\sqrt{N_n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Наконец, по лемме Слуцкого получаем, что

$$\frac{S_{N_n}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, a).$$

Доказательство 2 (с использованием неравенства Колмогорова). Сведем к ЦПТ, а дальше посмотрим что делать. Для этого представим $S_{N_n} = S_n + (S_{N_n} - S_n)$. Тогда

$$\frac{S_{N_n}}{\sqrt{n}} = \frac{S_n}{\sqrt{n}} + \frac{S_{N_n} - S_n}{\sqrt{n}}$$

Но это не совсем то, что нам нужно, потому что первое слагаемое стремится по распределению к $\mathcal{N}(0, 1)$, поэтому даже если мы найдем предел второго слагаемого, то непонятно как получить дальше $\mathcal{N}(0, a)$. Судя по утверждению, в этом слагаемом должен крыться дополнительный разброс. Поэтому хотелось бы иметь сумму из двух слагаемых, одно из которых стремится к $\mathcal{N}(0, a)$, а второе стремится к нулю по вероятности. Тогда по лемме Слуцкого мы получим требуемый результат.

Тогда попробуем так:

$$\begin{aligned} \frac{S_{N_n}}{\sqrt{n}} &= \sqrt{a} \frac{S_{N_n}}{\sqrt{na}} = \sqrt{a} \frac{S_{N_n}}{\sqrt{[na] + \{na\}}} = \\ &= \sqrt{a} \frac{S_{N_n}}{\sqrt{[na]}} \frac{\sqrt{[na]}}{\sqrt{[na] + \{na\}}} = \sqrt{a} \frac{S_{N_n}}{\sqrt{[na]}} \left(1 + O\left(\frac{1}{n}\right)\right) \end{aligned}$$

А проще так:

$$\frac{S_{N_n}}{\sqrt{n}} = \frac{S_{N_n}}{\sqrt{[na]}} \cdot \frac{\sqrt{[na]}}{\sqrt{n}}$$

Здесь второй множитель стремится к \sqrt{a} . Тогда нужно показать, что первый множитель сходится по распределению к $\mathcal{N}(0, 1)$. То есть задача сводится к случаю $a = 1$: надо доказать, что $\frac{S_{N_n}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$, если $\frac{N_n}{n} \xrightarrow{P} 1$ (эта переформулировка нужна для упрощения дальнейших рассуждений).

Тогда можем вернуться к представлению

$$\frac{S_{N_n}}{\sqrt{n}} = \frac{S_n}{\sqrt{n}} + \frac{S_{N_n} - S_n}{\sqrt{n}}$$

и тут уже остается показать, что $\frac{S_{N_n} - S_n}{\sqrt{n}} \xrightarrow{d} 0$ или, что равносильно, $\frac{S_{N_n} - S_n}{\sqrt{n}} \xrightarrow{P} 0$.

$$\begin{aligned} P\left(\left|\frac{S_{N_n} - S_n}{\sqrt{n}}\right| \geq \varepsilon\right) &= P(|S_{N_n} - S_n| \geq \varepsilon\sqrt{n}) = \\ &= P(|S_{N_n} - S_n| \geq \varepsilon\sqrt{n}, |N_n - n| \geq \varepsilon^3 n) + \\ &+ P(|S_{N_n} - S_n| \geq \varepsilon\sqrt{n}, |N_n - n| < \varepsilon^3 n) \leq \\ &\leq P(|N_n - n| \geq \varepsilon^3 n) + P\left(\max_{|k-n| < \varepsilon^3 n} |S_k - S_n| \geq \varepsilon\sqrt{n}\right) \end{aligned}$$

Первое слагаемое стремится к нулю, потому что $\frac{N_n}{n} \xrightarrow{P} 1$ по условию. Чтобы оценить второе слагаемое, воспользуемся неравенством Колмогорова:

$$P\left(\max_{|k-n| < \varepsilon^3 n} |S_k - S_n| \geq \varepsilon\sqrt{n}\right) \leq \frac{1}{\varepsilon^2 n} \sum_{k: |k-n| < \varepsilon^3 n} DX_k \leq \frac{2n\varepsilon^3}{n\varepsilon^2} = 2\varepsilon.$$

■

Теория меры

Начало. Неизмеримое множество

Понятно, что с древности нужно было измерять длины и площади. Но тогда также понятно и то, откуда взялась теория меры — раз мы умеем измерять длины прямых отрезков и площади прямоугольников, хочется как-то *продолжить* это умение и на кривые линии, и на площади фигур с кривыми границами, и вообще на все подмножества прямой и плоскости (ну и далее объема и т.п.).

Другими словами, мы хотим определить меру — начнем с малого — на прямой, т.е. функцию:

$$\mu : 2^{\mathbb{R}} \rightarrow \mathbb{R}_{\geq 0}$$

причем такую, которая обладает свойствами

- 1) $\mu([a, b]) = b - a$
- 2) $\mu(A + \alpha) = \mu(A)$ — сдвиг
- 3) $\mu\left(\bigsqcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$ — σ -аддитивность

Утверждение. Меры μ со свойствами 1)–3) не существует.

□ Предположим, что существует такая мера μ .

Пусть $a \sim b$ (эквивалентны), если $a - b \in \mathbb{Q}$. Тогда \mathbb{R} разбивается на классы эквивалентности:

$$\mathbb{R} = \bigsqcup_{\alpha} K_{\alpha}.$$

Очевидно, что $K_{\alpha} \cap [0, 1] \neq \emptyset$, поэтому, используя аксиому выбора, мы можем взять из каждого класса элемент

$$x_{\alpha} \in K_{\alpha} \cap [0, 1].$$

И тогда мы получим множество "хитрое-прехитрое, которое Лебег придумал"

$$A := \{x_{\alpha}\}_{\alpha}$$

Теперь занумеруем все рациональные числа на $[-1, 1]$:

$$\{r_n\} = \mathbb{Q} \cap [-1, 1].$$

Заметим, что

$$[0, 1] \subseteq \bigsqcup_{n=1}^{\infty} (A + r_n) \subseteq [-1, 2]$$

Второе вложение очевидно. Посмотрим на первое. Пусть $a \in [0, 1]$. Тогда если $a \in \mathbb{Q}$, то можно взять $x_a \in A \cap \mathbb{Q}$ и $r_{n_a} = x_a - a$. Пусть теперь $a \notin \mathbb{Q}$. Тогда если бы $\forall r_n \ a - r_n \notin A$, то это бы противоречило построению A .

Таким образом, по свойству 3) получаем, что

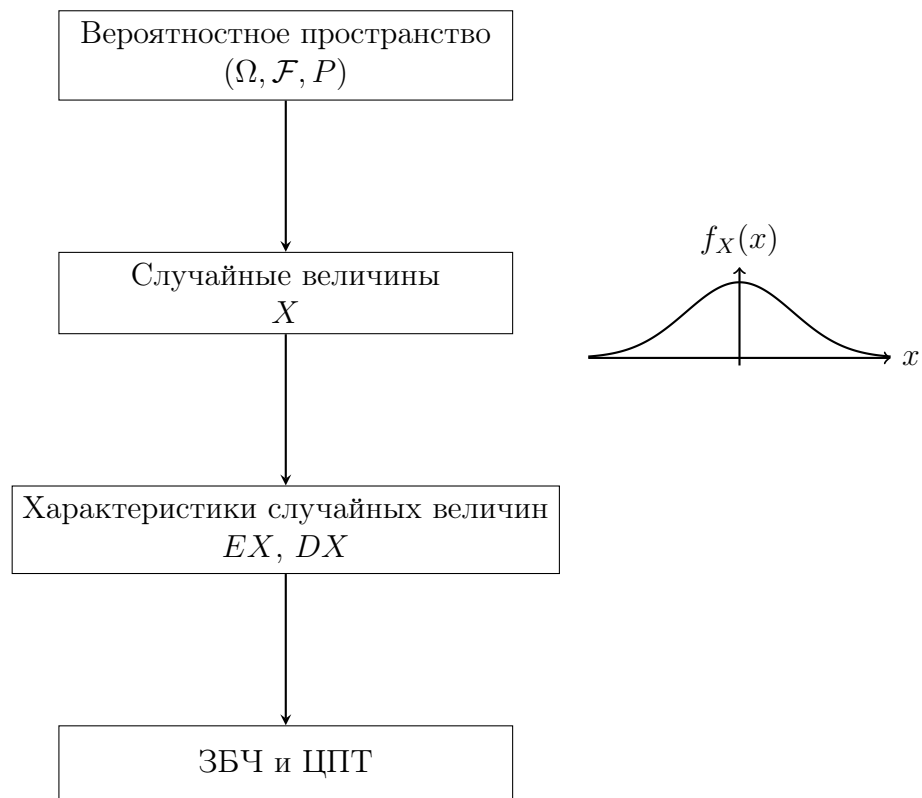
$$1 = \mu([0, 1]) \leq \mu\left(\bigsqcup_{n=1}^{\infty} (A + r_n)\right) = \sum_{n=1}^{\infty} \mu(A) \leq \mu([-1, 2]) = 3.$$

Если $\mu(A) = 0$, то получаем $1 \leq 0$. Если $\mu(A) > 0$, то ряд $\sum_{n=1}^{\infty} \mu(A)$ расходится. Получили противоречие. ■

Поэтому Лебег рассматривал меру на измеримых множествах

$$\mu : \Lambda_n \rightarrow \mathbb{R}_{\geq 0}.$$

Вопрос только в том, что из себя представляет Λ_n .



Список литературы

- [1] Шкляев А.В. *Материалы к семинарам по математической статистике на мехмате МГУ*. 2024.
- [2] Шкляев А.В. *Материалы к семинарам по теории вероятностей на мехмате МГУ*. 2024.