

§7. Предельные теоремы в схеме испытаний Бернулли.

Нормальное распределение

При выводе распределения Пуассона мы исследовали асимптотику биномиального распределения, когда $n \rightarrow \infty$, $p \rightarrow 0$, $np = \lambda(\text{const})$. Существует, однако, широкий класс практических задач, в которых построение вероятностных моделей требует асимптотического анализа биномиального распределения при фиксированном $p \in (0; 1)$ и $n \rightarrow \infty$.

Пример 7.1 (*определение видимой звездной величины*). Наблюдения за изменением блеска небесных светил, в частности звезд, является одной из важнейших задач практической астрономии. Только с помощью анализа таких наблюдений можно обнаружить *переменные* звезды, поставляющие информацию о расстояниях до отдаленных светил (цефеиды), а также об их массах, размерах и пр. (затменные переменные и спектрально-двойные звезды).

Величина блеска звезды определяется так называемой *видимой звездной величиной* – характеристикой светимости, пропорциональной количеству квантов света, исходящих от звезды и достигших прибора (электрического фотометра, фотографической пластинки и т.п.), который регистрирует поток лучевой энергии. С точки зрения проблемы построения вероятностной модели изменчивости в повторных наблюдениях блеска, мы имеем ту же картину, что и при измерениях интенсивности радиоактивного источника: каждый квант света с определенной вероятностью p достигает регистрирующего прибора, и общее количество регистрируемых квантов определяет результат наблюдения блеска звезды. Принципиальное различие с измерениями радиоактивности состоит в достаточно большом значении вероятности „успешного исхода” p , в то время как общее количество „испытаний” n (в данном случае – количество квантов, направленных на прибор) чрезвычайно велико. Таким образом возникает проблема асимптотического анализа биномиального распределения при фиксированном p и $n \rightarrow \infty$.

Пример 7.2 (*определение общего содержания серы в дизельном топливе*). Общее содержание серы служит одной из важных характеристик экологической чистоты дизельного топлива. Речь идет не об „элементарной сере” – процентном содержании химического эле-

мента S, что с высокой степенью точности определяется с помощью спектрального анализа вещества, а способности элемента S при сгорании топлива соединяться с кислородом, образуя серный газ SO_2 . Именно этот газ через выхлопные трубы машин попадает в среду нашего обитания и соединяется с водой, образуя серную кислоту H_2SO_4 . Ну, а что такое серная кислота, и что она может натворить с нашими легкими, вы знаете из школьного курса химии.

Итак, речь идет о химической активности серы, содержащейся в дизельном топливе в связанном виде. Анализ этой активности производится следующим образом. Берется определенное количество дизельного топлива, скажем 100 грамм, и сжигается в замкнутой колбе. Продукты сгорания частично выпадают в золу или в виде дыма по трубчатому отводу попадают в другую замкнутую колбу, наполненную водой. Серный газ соединяется с водой, образуя раствор серной кислоты. Титруя этот раствор определенным количеством щелочи, мы можем определить общее количество элемента серы, которое из дизельного топлива через сжигание и последующее соединение с кислородом и водой перешло в серную кислоту. Разделив это количество серы на вес анализируемой пробы топлива (100 грамм) и умножив результат на 100%, мы получим результат x нашего статистического эксперимента по наблюдению случайной величины X .

Повторные анализы аналогичных проб той же партии топлива, в тех же условиях эксперимента и на тех же приборах указывают на значительную изменчивость результатов каждого эксперимента. Метрологический анализ испытаний указывает на то, что эта изменчивость в первую очередь обусловлена случайным характером процессов „спекания” определенного количества серы с другими продуктами сгорания и выпадения их в золу, а также неполным соединением серного газа с водой. Грубо говоря, каждая молекула серы только с некоторой достаточно высокой вероятностью p может достичь своего конечного состояния в молекуле серной кислоты и внести свой вклад в результат x наблюдения X . Понятно, что количество n молекул серы в пробе топлива чрезвычайно велико, тогда как вероятность p достаточно велика (близка к единице). Следовательно, мы имеем дело с проблемой асимптотического анализа биномиального распределения при растущем числе испытаний n и постоянной вероятности успеха p .

Ограничимся рассмотрением этих двух примеров, из которых легко видеть, что существует обширнейший класс статистических экспериментов, связанных с наблюдением линейной функции от случайной величины с биномиальным законом распределения $B(n, p)$, в котором $p = \text{const}$, а n чрезвычайно велико. Проведем асимптотический анализ такой ситуации и начнем его с исследования асимптотического поведения X/n – частотной оценки вероятности p успешного испытания в схеме Бернулли. Тот факт, что при $n \rightarrow \infty$ относительная частота X/n стремится к p в определенном вероятностном смысле устанавливает один из основных законов теории вероятностей, открытый И. Бернулли в XVII веке.

Теорема 7.1. (Закон больших чисел Бернулли). Пусть $X \sim B(n, p)$. Тогда, каково бы ни было $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{X}{n} - p \right| > \varepsilon \right) = 0.$$

Доказательство. Воспользуемся неравенством Чебышева в форме следствия 6.1, где в случае биномиального распределения $\mathbf{E}X = np$ и $\mathbf{D}X = np(1 - p)$. Имеем

$$P \left(\left| \frac{X}{n} - p \right| > \varepsilon \right) \leq \frac{\mathbf{D}(X/n)}{\varepsilon^2} = \frac{np(1 - p)/n^2}{\varepsilon^2} = \frac{p(1 - p)}{n\varepsilon^2} \rightarrow 0,$$

когда $n \rightarrow \infty$.

Закон больших чисел разъясняет природу стабилизации относительной частоты выпадения герба около значения $p = 1/2$, которую мы наблюдали на первой лекции по теории вероятностей. Действительно, в случайных экспериментах нельзя утверждать, что $|X/n - p| \leq \varepsilon$, начиная с некоторого n . Истина в том, что, начиная с некоторого n , это неравенство выполняется с любой, наперед заданной и сколь угодно близкой к единице вероятностью. Таким образом мы должны сказать, что в данном случае наблюдается *сходимость по вероятности*, которая имеет совершенно другую природу, чем та сходимость, которую мы изучаем в курсе математического анализа.

Вывод закона больших чисел содержит также объяснение феномену, связанному с порядком $n^{-1/2}$ ошибки в приближении $p (= 1/2)$

величиной X/n . Действительно, в случае $p = 1/2$ стандартное отклонение $\sigma = \sqrt{\mathbf{D}(X/n)} = (2\sqrt{n})^{-1}$, распределение случайной величины X/n симметрично, и в силу правила „двух сигм“ интервал $0.5 \pm n^{-1/2}$ покрывает 90% центральной части области возможных значений X/n .

Естественно, если не делить X на n , то $X \rightarrow \infty$ по вероятности, когда $n \rightarrow \infty$. Но если X центрировать ее средним значением np и затем масштабировать стандартным отклонением, то построенная таким образом случайная величина $Y_n = (X - np)/\sqrt{np(1-p)}$ имеет при $n \rightarrow \infty$ невырожденное распределение. Вид этого распределения устанавливает знаменитая *предельная теорема Муавра–Лапласа* (18 век!). При доказательстве существенно используется следующий технический результат.

Лемма 7.1. Пусть $X \sim B(n, p)$, $n \rightarrow \infty$ и целое $k \rightarrow \infty$ так, что $1 > \hat{p} = k/n = O(1)$. Тогда

$$P(X = k) = f(k | n, p) = \frac{1}{\sqrt{2\pi n \hat{p}(1-\hat{p})}} \exp\{-nH(\hat{p})\} (1 + O(n^{-1})),$$

где

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}, \quad 0 < x < 1.$$

Доказательство. Воспользуемся асимптотической формулой Стирлинга

$$n! = \sqrt{2\pi n} n^n e^{-n} (1 + O(n^{-1}))$$

для факториалов $n!$, $k!$ и $(n-k)!$ в биномиальном коэффициенте C_n^k и представим функцию плотности биномиального распределения в асимптотическом виде:

$$\begin{aligned} f(k | n, p) &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\ &= \frac{\sqrt{2\pi n} n^n e^{-n} p^k (1-p)^{n-k}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi(n-k)} (n-k)^{n-k} e^{-n+k}} \left(1 + O\left(\frac{1}{n}\right)\right) = \\ &= \frac{\exp\{n \ln n - k \ln k - (n-k) \ln(n-k) + k \ln p + (n-k) \ln(1-p)\}}{\sqrt{2\pi n \frac{k}{n} \left(1 - \frac{k}{n}\right)}}. \end{aligned}$$

$$(1 + O(n^{-1})).$$

Доказательство завершается очевидными преобразованиями выражения, стоящего в фигурных скобках под экспонентой, к виду $\{-nH(\hat{p})\}$.

Лекция 12

Теорема 7.2. (Локальная предельная теорема Муавра–Лапласа). Пусть при $n \rightarrow \infty$ целое $k = np + O(\sqrt{n})$. Тогда

$$f(k | n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp \left\{ -\frac{(k - np)^2}{2np(1-p)} \right\} (1 + O(n^{-1/2})).$$

Доказательство. Так как по условию теоремы $\hat{p} = k/n = p + O(n^{-1/2})$, то естественно воспользоваться асимптотической формулой леммы 7.1, разлагая функции $(\hat{p}(1-\hat{p}))^{-1/2}$ и $H(\hat{p})$ в ряд Тейлора по степеням $\hat{p} - p = O(n^{-1/2})$.

Имеем

$$\begin{aligned} (\hat{p}(1-\hat{p}))^{-1/2} &= ((p + O(n^{-1/2}))(1-p + O(n^{-1/2})))^{-1/2} = \\ &= (p(1-p))^{-1/2} (1 + O(n^{-1/2})), \end{aligned}$$

и для доказательства теоремы остается показать, что

$$nH(\hat{p}) = \frac{(k - np)^2}{2np(1-p)} + O\left(\frac{1}{\sqrt{n}}\right). \quad (1)$$

Разложим

$$H(\hat{p}) = \hat{p} \ln \frac{\hat{p}}{p} + (1-\hat{p}) \ln \frac{1-\hat{p}}{1-p}$$

в ряд Тейлора в окрестности точки $\hat{p} = p$:

$$H(\hat{p}) = H(p) + (\hat{p} - p)H'(p) + \frac{(\hat{p} - p)^2}{2!}H''(p) + \frac{(\hat{p} - p)^3}{3!}H'''(p + \lambda(\hat{p} - p)),$$

где, как и в любом разложении Тейлора, $0 < \lambda < 1$.

Имеем $H(p) = 0$, и так как $H'(x) = \ln(x/p) - \ln((1-x)/(1-p))$, то $H'(p) = 0$. Далее, $H''(x) = 1/x + 1/(1-x)$, откуда $H''(p) = (p(1-p))^{-1}$. Наконец, $H'''(x) = -x^{-2} + (1-x)^{-2}$, что влечет ограниченность

$H'''(p + \lambda(\hat{p} - p))$ при больших n , поскольку p отграничено от 0 и 1. Таким образом,

$$H(\hat{p}) = \frac{(\hat{p} - p)^2}{2p(1-p)} + O((\hat{p} - p)^3),$$

что, очевидно, эквивалентно (1).

Теорема 7.3. (Интегральная предельная теорема Муавра–Лапласа). Для любых постоянных a и b и случайной величины $X \sim B(n, p)$ справедливо асимптотическое представление

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx. \quad (2)$$

Доказательство. Используя теорему 7.2, представим вероятность $P(a \leq Y_n < b)$ с $Y_n = (X - np)/\sqrt{np(1-p)}$ в виде

$$P(a \leq Y_n < b) = \sum_{k \in A} \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{(k - np)^2}{2np(1-p)}\right\} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \quad (3)$$

где множество целых чисел

$$A = \left\{k : a \leq \frac{k - np}{\sqrt{np(1-p)}} < b\right\}.$$

Применение локальной предельной теоремы в данном случае оправдано: если $k \in A$, то при $n \rightarrow \infty$ справедливо асимптотическое представление $k = np + O(\sqrt{n})$.

Покажем теперь, что правая часть (3) представляет собой сумму Дарбу для интеграла в правой части равенства (2). Для этого положим

$$x_k = \frac{k - np}{\sqrt{np(1-p)}}, \quad \Delta x_k = x_k - x_{k-1} = \frac{1}{\sqrt{np(1-p)}}, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

и разобьем отрезок $[a; b]$ точками x_k , $k \in A$. Поскольку $\Delta x_k \rightarrow 0$ при $n \rightarrow \infty$, а суммарная длина отрезков разбиения

$$\sum_{k \in A} \Delta x_k \approx b - a,$$

то число отрезков разбиения растет с ростом n , в то время как их длина стремится к нулю. Следовательно,

$$\sum_{k \in A} \varphi(x_k) \Delta x_k \longrightarrow \int_a^b \varphi(x) dx.$$

Для завершения доказательства остается только заметить, что $0 < \varphi(x) < 1$, и поэтому при $n \rightarrow \infty$

$$0 \leq \sum_{k \in A} \varphi(x_k) \Delta x_k \cdot O\left(\frac{1}{\sqrt{n}}\right) \leq \frac{b-a}{\sqrt{n}} \rightarrow 0.$$

З а м е ч а н и е. Интегральная теорема Муавра–Лапласа иногда формулируется в терминах следующего приближенного равенства для распределения биномиальной случайной величины X :

$$P(a \leq X < b) \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{np(1-p)}}}^{\frac{b-np}{\sqrt{np(1-p)}}} \exp\left\{-\frac{x^2}{2}\right\} dx, \quad n \gg 1. \quad (4)$$

В такой записи теоремы знак \approx означает асимптотическую эквивалентность правой и левой частей (4) (их отношение стремится к единице при $n \rightarrow \infty$) лишь в случае незначительной удаленности a и b от центра np биномиального распределения. Для этого достаточно сравнить запись одного и того же утверждения с помощью формул (2) и (4), чтобы убедиться в справедливости формулы (4) лишь при значениях a и b порядка $np + O(\sqrt{np})$. В противном случае, как левая, так и правая части (4) с ростом n стремятся к нулю, но с разной скоростью. Асимптотический анализ биномиальных вероятностей в областях, удаленных от np на порядок больший, чем $O(\sqrt{np})$, составляет содержание *теорем о больших отклонениях биномиального распределения*, которые в нашем курсе теории вероятностей рассматриваться не будут.

Как известно из общего курса анализа, интеграл Эйлера–Пуассона

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

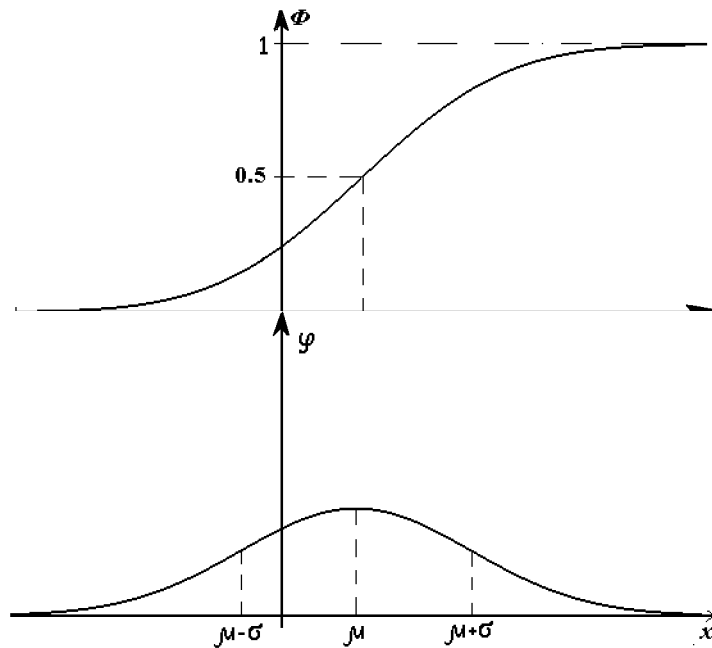
поэтому при любых $\mu \in \mathbb{R}$ и $\sigma \in \mathbb{R}_+$

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \exp\left\{-\frac{t^2}{2}\right\} dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$$

есть функция распределения, а

$$\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} -$$

функция плотности. Эти функции определяют двухпараметрическое семейство *нормальных* или *гауссовских* распределений с носителем $\mathcal{X} = \bar{\mathbb{R}} = [-\infty, +\infty]$ и параметрическим пространством $\Theta = \mathbb{R} \times \mathbb{R}_+$. Мы будем обозначать это распределение $\mathcal{N}(\mu, \sigma^2)$.



Если $\mu = 0$, а $\sigma = 1$, то $\mathcal{N}(0, 1)$ называется *стандартным нормальным* распределением; ему соответствуют функция распределения $\Phi(x)$ и функция плотности $\varphi(x)$. Поскольку параметры нормального распределения являются параметрами сдвига (μ) и масштаба (σ), то семейство нормальных распределений замкнуто относительно линейных преобразований случайных величин: если $X \sim \mathcal{N}(0, 1)$, то $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

Так как $\exp\{-x^2/2\}$ – четная функция, то нормальное распределение симметрично относительно точки $x = \mu$, которая, как легко видеть, является модой распределения. Симметричность функции

плотности влечет также очевидные равенства: $\Phi(-x) = 1 - \Phi(x)$ и $\Phi(0) = 1/2$. Графики функции распределения и функции плотности нормального закона $\mathcal{N}(\mu, \sigma^2)$ представлены на рисунке.

Так как среднее значение стандартного нормального распределения

$$\mathbf{E}X = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx = 0,$$

(как интеграл от нечетной функции по всему \mathbf{R}), то $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma)$ имеет среднее значение μ . В силу той же нечетности подынтегральных функций все центральные моменты нечетного порядка $\mu_{2k+1} = \mathbf{E}(X - \mu)^{2k+1} = 0$.

Четные моменты вычисляются с помощью гамма-функции Эйлера:

$$\begin{aligned} \mu_{2k} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^{2k} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} dx = \frac{2\sigma^{2k}}{\sqrt{2\pi}} \int_0^{\infty} t^{2k} \exp\left\{-\frac{t^2}{2}\right\} dt = \\ &= \frac{\sigma^{2k} 2^k \sqrt{2}}{\sqrt{2\pi}} \int_0^{\infty} x^{k-1/2} e^{-x} dx = \frac{\sigma^{2k} 2^k \sqrt{2}}{\sqrt{2\pi}} \Gamma\left(k + \frac{1}{2}\right) = \sigma^{2k} (2k - 1)!!. \end{aligned}$$

В частности, $\mathbf{D}X = \sigma^2$, что оправдывает обозначения параметров μ и σ^2 нормального распределения. Так как $\mu_4 = 3\sigma^4$, то коэффициент эксцесса $\gamma_2 = 0$. В силу этого пикообразность или сплюсченность вершины функции плотности любого распределения соотносится с кривой нормальной плотности, которая часто называется в честь Ф. Гаусса *гауссиадой*.

Итак, возвращаясь к нашим примерам с определениями видимой звездной величины и общего содержания серы в дизельном топливе, мы должны прийти к заключению о нормальности распределения наблюдаемой случайной величины (заметим, что это предположение блестяще подтверждается статистическим анализом реальных данных). В этом распределении μ играет роль параметра, неизвестное значение которого составляет предмет проводимого исследования (эксперимента), в то время как значение σ характеризует ошибку наблюдений.