Спецкурс "Большие уклонения"

Шкляев А.В.

5 октября 2018 г.

Задача о больших уклонениях

1 Введение

1.1 Мотивация

Классические предельные теоремы, которые мы изучаем, позволяют нам описывать "типичное" поведение случайных последовательностей и процессов. Так, например, классическая центральная предельная теорема утверждает, что

$$\mathbf{P}\left(\frac{S_n - na}{\sqrt{n}\sigma} \in [a,b]\right) \to \mathbf{P}(\xi \in [a,b]),$$

где $\xi \sim \mathcal{N}(0,1)$. Эта теорема описывает поведение только в "типичной области" $na + O(\sqrt{n})$. Мы не можем сказать насколько вероятно, что S_n окажется, скажем, правее $na + n^{2/3}$ — лишь можем сказать, что эта вероятность мала.
С другой стороны, может возникнуть вопрос — а зачем нам оценивать "малость" вероятностей?

Пример 1.1. Рассмотрим факт появления жизни на Земле. Само по себе появление жизни на случайно взятой планете очень маловероятно. С другой стороны, количество планет во вселенной огромное, скажем, 10^{20}. Тогда если вероятность появления жизни окажется 10^{-10}, то число обитаемых планет, по видимости, огромно. Если вероятность появления жизни, скажем, 10^{-20}, то вполне возможно, что мы единственная обитаемая планета. Если же она 10^{-30}, то шанс зарождения жизни крайне мал и это либо чудо, либо мы неправильно произвели наши оценки и неправильно восстановили процесс появления жизни.

Пример 1.2. Представим себе появление, что дамба выстроена так, что вероятность того, что река ее превысит равна 0.001. Тогда за 100 лет (если мы считаем независимыми уровни в разные дни) вероятность наводнения уже совсем не маленькая — порядка 10%.
Следовательно, для нас жизненно важно не только представлять, что то или иное событие редко, но и уметь оценивать его вероятность.

1.2 Неприменимость центральной предельной теоремы к редким событиям

Как выглядит асимптотика $\mathbf{P}(S_n \geq \theta n)$ при $\theta > \mu$, где X_i н.о.р., $\mathbf{E}X_i = \mu$?

Пример 1.3. Попробуем применить ЦПТ

$$\mathbf{P}(S_n \geq \theta n) = \mathbf{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \geq \frac{(\theta-\mu)\sqrt{n}}{\sigma}\right) \approx 1 - \Phi\left(\frac{(\theta-\mu)\sqrt{n}}{\sigma}\right)?$$

Это, вообще говоря, нелегально, но вдруг формула окажется верной? Давайте поймем как ведет себя $1-\Phi(x)$ при $x \to \infty$. Угадаем ответ и воспользуемся правилом Лопиталя:

$$\frac{1-\Phi(x)}{\frac{1}{\sqrt{2\pi}x}e^{-x^2/2}} = \frac{-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}}{-\frac{1}{\sqrt{2\pi}x^2}e^{-x^2/2} - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \to 1, \ x \to \infty.$$

Итак, если бы ЦПТ была верна не только при фиксированных x, но и при $x = O(n)$, то мы бы получили

$$\mathbf{P}(S_n \geq \theta n) \sim \frac{1}{\sqrt{2\pi n}\sigma(\theta-\mu)} \exp\left(-\frac{(\theta-\mu)^2 n}{2\sigma^2}\right).$$

Пример 1.4. Если $X_i \sim \mathcal{N}(\mu, \sigma^2)$, то $S_n \sim \mathcal{N}(\mu n, \sigma^2 n)$, то ЦПТ не предельная, а точная теорема, поэтому асимптотика верна.

Пример 1.5. Рассмотрим X_i с пуассоновским распределением с параметром λ, $S_n = X_1 + ... + X_n$ — пуассоновская с параметром λn. Тогда при $m = [\theta n] + 1$

$$\mathbf{P}(S_n \geq \theta n) = \sum_{k=m}^{n} \mathbf{P}(S_n = k) = \sum_{k=m+1}^{n} \frac{(\lambda n)^k}{k!}e^{-\lambda n} = (\lambda n) = \frac{(\lambda n)^m}{m!}e^{-\lambda n} \sum_{k=0}^{\infty} \frac{(\lambda n)^k}{(m+k)(m+k-1)...(m+1)}$$

Ряд в правой части не превосходит геометрической прогрессии с шагом λ/θ, а значит суммируется к некоторой величине порядка константы. Тем самым порядок нашей суммы есть

$$\frac{(\lambda n)^m}{m!}e^{-\lambda n} \sim \frac{(\lambda n)^m}{\sqrt{2\pi m} m^m} e^{m-\lambda n} = \frac{1}{\sqrt{2\pi\theta n}} e^{-m(\ln(\lambda n) - \ln m) + m - \lambda n}.$$

При $\theta n \in \mathbb{Z}$ имеем

$$\mathbf{P}(S_n \geq \theta n) \sim \frac{1}{\sqrt{2\pi\theta n}} \exp\left(-n\left(\theta \ln\frac{\lambda}{\theta} - \theta + \lambda\right)\right)$$

Общий вид асимптотики похож на тот, что мы получили выше, но величина в экспоненте другая.

1.3 Примеры

Давайте попробуем рассмотреть нашу задачу на случайном блуждании. Пусть X_i — н.о.р. с $EX_i = \mu$, $S_n = X_1 + ... + X_n$. Рассмотрим $\theta > \mu$ и поглядим на вероятность $P(S_n \geq \theta n)$. Давайте посмотрим несколько примеров:

Пример 1.6. Пусть X_i принимают значения 0 и 1 с вероятностями $1-p$, p, $\theta > p$. Тогда

$$P(S_n \geq \theta n) = \sum_{k=(1+\theta)n/2}^{n} C_n^k p^k (1-p)^{n-k}.$$

С помощью формулы Стирлинга можно оценить слагаемые этой суммы, откуда можно вывести, что

$$P(S_n \geq \theta n) \sim \frac{1}{\sqrt{2\pi(1-\theta)\theta n}} \exp\left(-\left(\theta \ln\left(\frac{\theta}{p}\right) + (1-\theta)\ln\left(\frac{1-\theta}{1-p}\right)\right) n\right).$$

Как мы видим, в обоих случаях вероятность имеет вид $C(\theta)n^{-1/2}\exp(-\Lambda(\theta)n)$ при каких-то функциях $C(\theta)$, $\Lambda(\theta)$. Отсюда возникает гипотеза, что такая ситуация будет и в других случаях. Однако, так будет заведомо не всегда:

Пример 1.7. Пусть $P(X > x) \sim x^{-a}$. Тогда

$$P(S_n \geq \theta n) \geq P(X_1 > (\theta - \mu + \varepsilon)n) P(X_2 + ... + X_n > (\mu - \varepsilon)n) \sim (\theta - \mu + \varepsilon)^{-a} n^{-a}.$$

Таким образом, если X имеет степенные хвосты, то и S_n будет иметь не более чем степенные хвосты.

2 Явные оценки

Попробуем оценить $\mathbf{P}(S_n \geq \theta n)$ сверху. Воспользуемся неравенством Крамера:

$$\mathbf{P}(S_n \geq \theta n) = \mathbf{P}(\exp(S_n h) \geq \exp(\theta n h)) \leq \frac{\mathbf{E}\exp(S_n h)}{\exp(\theta n h)} = \left(\frac{\mathbf{E}e^{hX_1}}{e^{\theta h}}\right)^n,$$

где $h > 0$. Выберем из приведенных оценок наилучшую. Прежде всего потребуем конечность математического ожидания в правой части: $R(h) = \mathbf{E}e^{hX} < \infty$ при $0 \leq h < h^+$, $h^+ > 0$. Это условие будем называть правосторонним условием Крамера.

Наилучшая оценка соответствует такому $h \in [0, h^+)$, что $f(h) = \theta h - \ln R(h)$ минимально. Для исследования нам понадобится дифференцируемость $R(h)$ и явные значения ее производной.

По определению

$$R'(h) = \lim_{\Delta h \to 0} \frac{R(h + \Delta h) - R(h)}{\Delta} = \lim_{\Delta h \to 0} \mathbf{E}e^{hX}\left(\frac{e^{\Delta h X} - 1}{\Delta h}\right).$$

В силу теоремы Лагранжа о среднем значении

$$(\Delta h)^{-1}\left(e^{\Delta h X} - 1\right) = e^{\xi} e^{|\Delta h X|} \leq e^{\xi} \leq e^{|\Delta h| X} + e^{-|\Delta h| X},$$

где $\xi \in [0, \Delta h X]$. Таким образом при любом $h \in (0, h^+)$

$$e^{hX}\left(\frac{e^{\Delta h} - 1}{\Delta h}\right) \leq e^{(h + \Delta h)X} + e^{(h - \Delta h)X},$$

откуда по теореме о мажорируемой сходимости

$$R'(h) = \mathbf{E}e^{hX} \lim_{\Delta h \to 0}\left(\frac{e^{\Delta h X} - 1}{\Delta h}\right) = \mathbf{E}e^{hX} X.$$

Аналогично

$$R''(h) = \mathbf{E}e^{hX} X^2.$$

Отметим, что

$$m(h) := (\ln R(h))' = \frac{\mathbf{E}e^{hX} X}{R(h)} = \int_{-\infty}^{\infty} x e^{hx} \frac{\mathbf{P}(X \in dx)}{R(h)} = \mathbf{E}X^{(h)},$$

где величина $X^{(h)}$ называется сопряженной к X и задается распределением

$$\mathbf{P}(X^{(h)} \in A) = \int_A e^{hx} \frac{\mathbf{P}(X \in dx)}{R(h)}.$$

Это действительно вероятностное распределение, поскольку при $A = \mathbb{R}$ мы получим 1 в силу определения.
Аналогично

$$\sigma^(h) := m'(h) = \frac{R''(h)}{R(h)} - \left(\frac{R'(h)}{R(h)}\right)^2 = \int_{-\infty}^{\infty} x^2 e^{hx} \frac{\mathbf{P}(X \in dx)}{R(h)} - \left(\mathbf{E}X^{(h)}\right)^2 = \mathbf{D}X^{(h)}.$$

Величина $X^{(h)}$ невырождена (то есть не является константой), если невырождена величина X, поскольку если $\mathbf{P}(X \in A) > 0$, то и $\mathbf{P}(X^{(h)} \in A) > 0$. Значит $\mathbf{D}X^{(h)} > 0$, то есть функция $\ln R(h)$ имеет положительную вторую производную и выпукла вниз. Значит $m(h)$ монотонно возрастает. Пусть $m(h) \to m^+$, $h \to h^+$. Тогда поскольку $m(0) = \mu = \mathbf{E}X$, то $m(h)$ принимает все значения из $[\mu, m^+)$ по одному разу.

Следовательно, при всех $\theta \in (m^-, m^+)$ найдется единственное h_θ, т.ч. $m(h_\theta) = \theta$.

Поскольку производная $h\theta - \ln R(h)$ есть $\theta - m(h)$, то h_θ — единственный ноль производной $f(h)$, причем $f''(h) = -\sigma^2(h) < 0$, откуда h_θ есть единственный максимум $f(h)$,

Положим $\Lambda(\theta) := f(h_\theta) = h_\theta\theta - \ln R(h_\theta)$. Тем самым мы доказали, что

$$\mathbf{P}(S_n \geq \theta n) \leq \exp(-\Lambda(\theta)n)$$

при $\theta \in [\mu, m^+)$. Как мы увидим в следующем разделе — это неплохая оценка.

3 Теорема Петрова

3.1 Теорема Петрова

Назовем левосторонним условием Крамера $\mathbf{E}e^{hX} < \infty$ при $h \in (h^-, 0]$, положим $m^- = \lim_{h \to h^- + 0} m(h)$. Оказывается, что справедлива следующая теорема Петрова:

Теорема 1.1. 1) Пусть величины X_i нерешетчаты (то есть $\not\exists a, d : \quad \mathbf{P}(X \in \{a + kd, \ k \in \mathbb{Z}\})$) и удовлетворяют правостороннему условию Крамера. Тогда

$$\mathbf{P}(S_n \geq \theta n) \sim \frac{1}{\sqrt{2\pi n}\sigma(h_\theta)h_\theta} \exp(-\Lambda(\theta)n), \ n \to \infty,$$

равномерно по любому отрезку $[\theta_1, \theta_2] \subset (\mu, m^+)$. Аналогичный результат верен при выполнении левостороннего условия Крамера

$$\mathbf{P}(S_n \leq \theta n) \sim \frac{1}{\sqrt{2\pi n}\sigma(h_\theta)h_\theta} \exp(-\Lambda(\theta)n), \ n \to \infty,$$

равномерно по любому отрезку $[\theta_1, \theta_2] \subset (m^-, \mu)$.

2) Пусть величины X_i решетчаты с шагом d (d наибольший такой шаг) и удовлетворяют правостороннему условию Крамера. Тогда

$$\mathbf{P}(S_n \geq \theta n) \sim \frac{d}{\sqrt{2\pi n}\sigma(h_\theta)\,(1 - e^{dh_\theta})} \exp(-\Lambda(\theta)n), \ n \to \infty,$$

при $\theta n \in \{an + kd, \ k \in \mathbb{Z}\}$. Аналогично при выполнении левостороннего условия

$$\mathbf{P}(S_n \leq \theta n) \sim \frac{d}{\sqrt{2\pi n}\sigma(h_\theta)\,(1 - e^{dh_\theta})} \exp(-\Lambda(\theta)n), \ n \to \infty,$$

равномерно по любому отрезку $[\theta_1, \theta_2] \subset (m^-, \mu)$.

Пример 1.8. Найдем параметры, фигурирующие в теореме, для нормального $\mathcal{N}(0,1)$ распределения:

$$R(h) = \mathbf{E}e^{hX} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{hx} e^{-x^2/2} dx = e^{h^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-h)^2/2} dx = e^{h^2/2},$$
$$m(h) = (\ln R(h))' = (h^2/2)' = h, \ h_\theta : m(h_\theta) = h_\theta = \theta, \ \sigma^2(h) = m'(h) = 1,$$
$$\Lambda(\theta) = \theta h_\theta - \ln R(h_\theta) = \theta^2 - \frac{1}{2}\theta^2 = \frac{1}{2}\theta^2.$$

Задача 1.1. Сформулируйте теорему Петрова для пуассоновского, геометрического, экспоненциального распределений.