Теория вероятностей. Лекция двадцатая
Центральная предельная теорема

Дмитрий Валерьевич Хлопин
glukanat@mail.ru

Институт математики и механики им. Н.Н.Красовского

14.03.2019

Что разобрали:

- Неравенства концентрации меры
- Сходимости случайных величин
- Усиленный закон больших чисел
- Слабая сходимость
- Характеристические функции
- Центральная предельная теорема

Характеристические функции — это функанский способ (фактически преобразование Фурье) работы с вероятностными распределениями. Это не только позволяет восстановить распределение (формула обращения), но и сохраняет сходимости, переводя близкие распределения в близкие гармоники и наоборот. Это дает и фундамент статистике — ЦПТ.

Характеристические функции. Определение

Каждой вероятности μ (над $(\mathbb{R},\mathcal{F})$) сопоставим функцию по правилу

$$\tilde{\mu}(y) \triangleq \int_{\mathbb{R}} e^{ixy}\mu(dx) \qquad \forall y \in \mathbb{R}.$$

Функцию $\tilde{\mu}$ назовем характеристической функцией вероятности μ; (здесь $e^{it} = \cos t + i\sin t$).

При этом, каждой случайной величине ξ можно вслед за $\mu_\xi = \xi\#\mathbb{P}$ рассмотреть характеристическую функцию случайной величины ξ:

$$\tilde{\xi} = \widetilde{\mu_\xi} = \mathbb{E}e^{iy\xi}\Big(= \int_{\mathbb{R}} e^{ixy} f_\xi(x)dx\Big).$$

Следствие 7. Вероятности равны тогда и только тогда, когда равны их характеристические функции. Распределения случайных величин равны тогда и только тогда, когда равны их характеристические функции.

Слабая сходимость

Будем говорить, что ξ_n сходится к ξ по распределению, обозначая её $\xi_n \xrightarrow{d} \xi$, если $\mu_{\xi_n} = \xi_n \# \mathbb{P}$ сходится к $\mu_\xi = \xi \# \mathbb{P}$ слабо, то есть для всякой непрерывной ограниченной функции ϕ

$$\int_\Omega \phi(\xi_n(\omega))\mathbb{P}(d\omega) \to \int_\Omega \phi(\xi(\omega))\mathbb{P}(d\omega).$$

При этом достаточно проверить сходимость для бесконечное число раз дифференцируемых финитных функций ϕ.

Предложение 8. Последовательность случайных величин ξ_n сходится по распределению к случайной величине ξ тогда и только тогда, когда для всех точек непрерывности функции распределения F_ξ

$$F_{\xi_n}(t) \to F_\xi(t) \text{ при } n \to \infty.$$

Слабая сходимость $\mu_n \Leftrightarrow$ поточечная сходимость $\tilde{\mu}_n$

Теорема 7. Последовательность вероятностей $\{\mu_n\}$ сходится к вероятности μ слабо тогда и только тогда, когда $\{\tilde{\mu}_n\}$ поточечно сходится к $\tilde{\mu}$.

Доказательство.

$\Rightarrow$. Если $\mu_n \xrightarrow{w} \mu$, то поскольку $e^{iyx} = \cos(yx) + i\sin(yx)$, где обе функции непрерывны и ограничены, характеристические функции $\tilde{\mu}_n(y)$ сходятся к $\tilde{\mu}(y)$ для каждого $y \in \mathbb{R}$.

$\Leftarrow$. Для доказательства в обратную сторону потребуется теорема Прохорова и следующая лемма:

Шаг 0: техническая лемма

Для каждого положительного r имеет место оценка:

$$\mu([-2/r, 2/r]) \geq \left|\frac{1}{r}\int_{-r}^{r} \tilde{\mu}(y)dy\right| - 1.$$

Доказательство.

$$\frac{1}{r}\int_{-r}^{r} \tilde{\mu}(y)dy = \frac{1}{r}\int_{-\infty}^{+\infty}\int_{-r}^{r} e^{iyx}dy\mu(dx)$$
$$= \int_{-\infty}^{+\infty} \frac{e^{ixr} - e^{-ixr}}{ixr}\mu(dx) = 2\int_{-\infty}^{+\infty} \frac{\sin(rx)}{rx}\mu(dx).$$

Далее,

$$\left|\frac{1}{r}\int_{-r}^{r} \tilde{\mu}(y)dy\right| \leq 2\int_{|x|\leq 2/r}\left|\frac{\sin(rx)}{rx}\right|\mu(dx) + 2\int_{|x|>2/r}\left|\frac{\sin(rx)}{rx}\right|\mu(dx)$$
$$\leq 2\mu([-2/r, 2/r]) + \mu(\mathbb{R} \setminus [-2/r, 2/r])$$
$$= 2\mu([-2/r, 2/r]) + (1 - \mu([-2/r, 2/r])) = 1 + \mu([-2/r, 2/r]).$$

Шаг 1: μ_n плотны, то есть $\forall\varepsilon\,\exists R\,\forall n : \mu_n([-R,R]) \geq 1-\varepsilon$

Выберем $\varepsilon > 0$, в силу непрерывности $\tilde\mu$ найдется $r > 0$ такое, что $|\tilde\mu(y) - 1| < \varepsilon/4$ для всех $y \in (-r, r)$. Имеем,

$$\left|\int_{-r}^{r} \tilde\mu(y)dy\right| = \left|\int_{-r}^{r} (\tilde\mu(y) - 1)dy + 2r\right| \geq 2r - \left|\int_{-r}^{r} (\tilde\mu(y) - 1)dy\right|$$
$$\geq 2r - \int_{-r}^{r} |\tilde\mu(y) - 1|dy \geq 2r - 2r\cdot\frac{\varepsilon}{4} = 2r(1-\varepsilon/4).$$

Поскольку $\tilde\mu_n$ сходится к $\tilde\mu$ поточечно, а $|\tilde\mu_n(y)| \leq 1$, по теореме Лебега $\int_{-r}^{r}\tilde\mu_n(y)dy \to \int_{-r}^{r}\tilde\mu(y)dy$. Найдем N такое, что для $n \geq N$

$$\left|\int_{-r}^{r} \tilde\mu_n(y)dy\right| \geq 2r(1-\varepsilon/2), \text{ то есть } \left|\frac{1}{r}\int_{-r}^{r} \tilde\mu_n(y)dy\right| \geq 2-\varepsilon.$$

Теперь из леммы следует $\mu_n([-2/r, 2/r]) \geq 1 - \varepsilon$ для $n \geq N$; но для каждого μ_n найдется такое $s_n > 0$, что $\mu_n([-s_n, s_n]) \geq 1-\varepsilon$. Осталось принять $R = \max(2/r, s_1, \ldots, s_N)$.

Шаг 2: собственно доказательство

В силу плотности, по теореме Прохорова можно выделить сходящуюся (в слабом смысле) подпоследовательность μ_{n_k} к некоторой предельной вероятности ν. Используя уже показанную "$\Rightarrow$", мы получаем, что $\tilde{\mu}_{n_k}$ поточечно сходится к $\tilde{\nu}$, то есть $\tilde{\nu} = \tilde{\mu}$. Следовательно, $\nu = \mu$.

Пусть последовательность μ_n еще не сходится к μ, т.е существуют число $\varepsilon > 0$, функция $\phi \in C_b(\mathbb{R})$ и подпоследовательность μ_{n_i} такие, что

$$\left| \int_{\mathbb{R}} \phi(x)\mu_{n_i}(dx) - \int_{\mathbb{R}} \phi(x)\mu(dx) \right| \geq \varepsilon.$$

Из μ_{n_i} можно выделить сходящуюся подподпоследовательность. Это противоречит предположению.

Закон больших чисел в форме Хинчина

Теорема 8 Пусть дана последовательность независимых одинаково распределенных случайных величин ξ_n со средним m. Тогда имеет место закон больших чисел: $\frac{\xi_1+\dots+\xi_n}{n}$ сходится по вероятности к m.

Доказательство. Достаточно проверить сходимость характеристических функций у $\eta_n \triangleq \frac{\xi_1+\dots+\xi_n}{n}$: для всех t имеем

$$\ln\tilde{\eta}_n(t) = \sum_{k=1}^{n}\ln\tilde{\xi}_k(t/n) = n\ln\tilde{\xi}_1(t/n) = t\cdot\frac{\ln\tilde{\xi}_1(t/n) - \ln\tilde{\xi}_1(0)}{t/n},$$

$$\lim_{n\to\infty}\ln\tilde{\eta}_n(t) = t\frac{d\ln\tilde{\xi}_1(s)}{ds}\Big|_{s=0} = t\frac{d\tilde{\xi}_1(s)}{\tilde{\xi}_1(0)ds}\Big|_{s=0} = it\mathbb{E}\xi_1 = itm = \ln\tilde{m}(t),$$

то есть $\tilde{\eta}_n(t) \to \tilde{m}(t)$. Тогда из полученной поточечной сходимости следует, что η_n слабо сходится к m. Поскольку m — константа, имеет место и сходимость по вероятности.

Закон больших чисел vs центральная предельная теорема

Из только что показанного закона больших чисел следует, что если у независимых одинаково распределенных случайных величин ξ_n есть матожидание $m = m_n$, то $\frac{\xi_1+\dots+\xi_n}{n}$ сходится по вероятности к m, то есть

$$\frac{\xi_1 + \dots + \xi_n - nm}{n} \xrightarrow{P} 0.$$

Как говорит центральная предельная теорема (следующий слайд), если у независимых одинаково распределенных случайных величин ξ_n есть дисперсия $\sigma^2 > 0$, то по распределению сходится $\frac{\xi_1+\dots+\xi_n-nm}{\sqrt{n}}$, точнее

$$\frac{\xi_1 + \dots + \xi_n - nm}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1).$$

Подумать: как из теоремы на следующем слайде следует формулировка выше.

Центральная предельная теорема. Формулировка

Пусть дана последовательность независимых в совокупности случайных величин ξ_i, причем для

$$m_n \triangleq \mathbb{E}\xi_n, \qquad \sigma_n^2 \triangleq D\xi_n, \qquad \mathcal{D}_n^2 \triangleq D(\xi_1+\dots+\xi_n) = \sum_{i=1}^{n} \sigma_i^2, \qquad \mu_n \triangleq \xi_n \# \mathbb{P}$$

выполнено условие Линдеберга:

$$\lim_{n\to\infty} \frac{1}{\mathcal{D}_n^2} \sum_{i=1}^{n} \int_{\{x:|x-m_i|\geq\varepsilon\mathcal{D}_n\}} (x-m_i)^2 \mu_i(dx) = 0 \qquad \forall \varepsilon > 0.$$

Тогда случайные величины

$$\frac{(\xi_1 - m_1) + \ldots + (\xi_n - m_n)}{\sqrt{\sigma_1^2 + \ldots + \sigma_n^2}}$$

сходятся по распределению к нормальному распределению с нулевым средним и единичной дисперсией.

Шаг 1: вклад каждого эксперимента идет к нулю

При доказательстве можно считать, что все $m_n = 0$, потому введем $\tau_n \overset{\triangle}{=} \frac{\xi_1+\ldots+\xi_n}{\sqrt{\sigma_1^2+\ldots+\sigma_n^2}}$. Достаточно доказать, что

$$\tilde{\tau}_n(y) = \mathbb{E}e^{i\tau_n y} \to e^{-y^2/2}.$$

для всякого y. Зафиксируем y.
Для каждого $\varepsilon > 0$

$$\max_{k=1,\ldots,n} \frac{\sigma_k^2}{\mathcal{D}_n^2} \leq \max_{k=1,\ldots,n} \frac{\int_{\{x:|x|\geq\varepsilon\mathcal{D}_n\}} x^2\mu_k(dx)}{\mathcal{D}_n^2} + \max_{k=1,\ldots,n} \frac{\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} x^2\mu_k(dx)}{\mathcal{D}_n^2}.$$

Первый член идет к нулю при $n\uparrow\infty$ по условию Линдеберга, а второй может быть оценен числом ε^2 в силу неравенства $|x| < \varepsilon\mathcal{D}_n$. Итак,

$$\lim_{n\uparrow\infty} \max_{k=1,\ldots,n} \frac{\sigma_k^2}{\mathcal{D}_n^2} = 0.$$

Шаг 2: $\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = 1 - \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n$.

Из $e^{it} = 1 + it + \frac{\theta_1 t^2}{2}$, $e^{it} = 1 + it - \frac{t^2}{2} + \frac{\theta_2 t^3}{6}$ для некоторых $\theta_1, \theta_2 \in [-1, 1]$, зависящих от t, следует

$$\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = \int_{-\infty}^{+\infty} e^{ix\frac{y}{\mathcal{D}_n}}\mu_k(dx)$$

$$= \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} e^{ix\frac{y}{\mathcal{D}_n}}\mu_k(dx) + \int_{\{x:|x|\geq\varepsilon\mathcal{D}_n\}} \left(e^{ix\frac{y}{\mathcal{D}_n}} \pm \frac{(yx)^2}{2\mathcal{D}_n^2}\right)\mu_k(dx)$$

$$= \int_{-\infty}^{+\infty} \left(1 + i\frac{y}{\mathcal{D}_n}x - \frac{(yx)^2}{2\mathcal{D}_n^2}\right)\mu_k(dx)$$

$$+ \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} \frac{\theta_2(yx)^3}{6\mathcal{D}_n^3}\mu_k(dx) + \int_{\{x:|x|\geq\varepsilon\mathcal{D}_n\}} (1+\theta_1)\frac{(yx)^2}{2\mathcal{D}_n^2}\mu_k(dx)$$

$$= 1 - \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n.$$

Шаг 3: $\sum_{k=1}^n |a_k^n| \to 0$.

При суммировании по k первых слагаемых в

$$a_k^n = \frac{y^2}{2\mathcal{D}_n^2}\int_{\{x:|x|\geq\varepsilon\mathcal{D}_n\}}(1+\theta_1)x^2\mu_k(dx) + \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\frac{\theta_2(yx)^3}{6\mathcal{D}_n^3}\mu_k(dx),$$

сумма идет в ноль по условию Линдеберга, а сумма вторых слагаемых не превосходит по модулю

$$\sum_{k=1}^n \frac{|y|^3}{6\mathcal{D}_n^3}\left|\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\theta_2 x^3\mu_k(dx)\right| \\ \leq \sum_{i=1}^n \frac{|y^3|\varepsilon}{6\mathcal{D}_n^3}\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} x^2\mathcal{D}_n\mu_k(dx) \leq \sum_{k=1}^n \frac{|y^3|\varepsilon\sigma_k^2}{6\mathcal{D}_n^2} = \frac{\varepsilon|y|^3}{6},$$

то есть может быть сделана сколь угодно близкой к нулю за счет произвольности выбора ε.

Шаг 4: $\tilde{\tau}_n(y) \to -y^2/2$

$$\tilde{\tau}_n(y) = \mathbb{E}e^{i\tau_n y} = \mathbb{E}e^{i(\xi_1+\ldots+\xi_n)\frac{y}{\mathcal{D}_n}} = \prod_{k=1}^{n} \tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right);$$

$$\ln\tilde{\tau}_n(y) = \sum_{k=1}^{n} \ln\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = \sum_{k=1}^{n} \ln\left(1 - \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right)$$

$$= -\sum_{k=1}^{n} \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + \sum_{k=1}^{n} a_k^n + \sum_{k=1}^{n} \theta_3 \left|-\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right|^2$$

в силу $\ln(1+t) = t + \theta_3|t|^2$ при $|t| < 1/2$ для некоторого $\theta_3 \in [-1, 1]$, зависящего от t. Осталось оценить последнее слагаемое.

$$\sum_{k=1}^{n} |\theta_3| \left|-\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right|^2 \leq \max_{k=1,\ldots,n}\left[\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + |a_k^n|\right] \sum_{k=1}^{n} \left|\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + |a_k^n|\right|$$

$$\leq \left(\frac{y^2}{2} + \sum_{k=1}^{n} |a_k^n|\right) \max_{k=1,\ldots,n}\left[\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + |a_k^n|\right] \to 0.$$

Как следствие: простейшая формулировка ЦПТ

Пусть $\{\xi_n\}$ независимы и одинаково распределены. Пусть $\mathbb{E}\xi_1 = m$, $D\xi_1 = \sigma^2 > 0$. Тогда распределения случайных величин $\frac{\xi_1+\ldots+\xi_n-nm}{\sigma\sqrt{n}}$ слабо сходятся к нормальному распределению с нулевым средним и единичной дисперсией, в частности, для всех $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}, (a < b)$

$$\mathbb{P}\Big(a < \frac{\xi_1 + \ldots + \xi_n - nm}{\sigma\sqrt{n}} < b\Big) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Неравенство Берри-Эссена.(без д-ва) При этом, для всех $b > 0$

$$\Big| F_{\frac{\xi_1+\ldots+\xi_n-nm}{\sigma\sqrt{n}}}(b) - \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \Big| < \frac{1}{2} \cdot \frac{\mathbb{E}|\xi_1 - m|^3}{\sigma^3\sqrt{n}}.$$

Центральная предельная теорема. Условие Ляпунова

Пусть $\{\xi_n\}$ независимы, их матожидания m_n и дисперсии σ_n^2 конечны. Пусть для $\mathcal{D}_n^2 \triangleq \sum_{i=1}^n \sigma_i^2$ при некотором $\delta > 0$ также выполнено условие Ляпунова:

$$\lim_{n\to\infty} \frac{1}{\mathcal{D}_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}(\xi_i - m_i)^{2+\delta} = 0.$$

Тогда распределения случайных величин

$$\frac{(\xi_1 - m_1) + \ldots + (\xi_n - m_n)}{D_n}$$

по распределению сходятся к нормальному распределению с нулевым средним и единичной дисперсией.

Условие Ляпунова ⇒ условие Линдеберга

Доказательство.

При всех $\varepsilon > 0$ имеем

$$\lim_{n\to\infty} \frac{1}{\mathcal{D}_n^2} \sum_{i=1}^{n} \int_{\{x:|x-m_i|\geq \varepsilon \mathcal{D}_n\}} (x-m_i)^2 \mu_i(dx)$$

$$\leq \lim_{n\to\infty} \frac{1}{\mathcal{D}_n^2 (\varepsilon \mathcal{D}_n)^\delta} \sum_{i=1}^{n} \int_{\{x:|x-m_i|\geq \varepsilon \mathcal{D}_n\}} |x-m_i|^{2+\delta} \mu_i(dx)$$

$$\leq \varepsilon^{-\delta} \frac{\sum_{i=1}^n \mathbb{E}|\xi_i - m_i|^{2+\delta}}{\mathcal{D}_n^{2+\delta}} \to 0.$$

Локальная предельная теорема (без д-ва)

Теорема 12. Пусть $\{\xi_n\}$ независимы и одинаково распределены, $\mathbb{E}\xi_1 = m$, $D\xi_1 = \sigma^2$, а ξ_i принимают целочисленные значения, причем НОД$(\operatorname{supp}\xi_1) = 1$. Тогда

$$\sigma\sqrt{n}\mathbb{P}\{\xi_1 + \ldots + \xi_n - nm = k\} - \frac{1}{\sqrt{2\pi}}e^{-(k-nm)^2/2\sigma\sqrt{n}}$$

стремится при $n \uparrow \infty$ к нулю равномерно по всем целым k.

Следствие 8. (Локальная предельная теорема Муавра–Лапласа) Пусть $\xi_1, \ldots, \xi_n$ — испытания в схеме Бернулли. Тогда

$$\mathbb{P}(\xi_1 + \ldots + \xi_n = k) = \frac{1}{\sqrt{2\pi np(1-p)}}e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2}(1 + \delta_n(k)),$$

где $\delta_n(k)$ стремятся к нулю равномерно по k при $n \to \infty$.

На пять минут... РАЗ

1. Пусть для независимых одинаково распределенных случайных величин ξ_k и для некоторых последовательностей чисел a_k и b_k случайные величины

$$\frac{\xi_1 + \xi_2 + \cdots + \xi_n - a_n}{b_n}$$

сходятся по распределению к невырожденной случайной величине η. Можно ли утверждать, что распределение у η нормально?
"Нет", "нельзя" оценивались в $0,3$, "незя" — в $0,2$.
Таких пределов не так много (четырехпараметрическое семейство), такие распределения называются устойчивыми распределениями (см. о подробностях Wiki, лучше англоязычную). К ним относится помимо нормального, например распределение Коши. Более того, про то свойство распределения Коши даже на паре было сказано. Помогло немногим...
Вторая задача дальше...

На пять минут... ДВА

2. В каком-то из 1000 мешков, возможно, все монеты фальшивые, зато в остальных — точно настоящие. Все фальшивые весят одинаково, все настоящие тоже, но любая фальшивая по весу отличается от настоящей. За какое минимальное число взвешиваний можно или найти мешок с фальшивыми монетами, или убедиться, что фальшивых монет нет? Предполагалось также, что монет в каждом мешке сколь угодно много, их можно вытаскивать, но разность весов фальшивой и настоящей неизвестна.
Задача имеет две трактовки: весы со стрелкой (можно засечь разность масс) или весы без стрелки (знаем только знак $>, <, =$).
Пусть весы без стрелки. За 8 взвешиваний можно: нужно бить на группы по степеням тройки, , первым взвешивать две группы по 243, вторым-третьим взвешиванием определяя какая монета больше весит.
За 6 точно нельзя $3^6 = 729 < 1000$. Если знаете как гарантировать за 7, высылайте (стоимость $0,7$)
Полное решение для второй трактовки еще дальше...

На пять минут... ОПЯТЬ ДВА

Пусть весы со стрелкой. В первое взвешивание на одну чашку кладем по 1 монете из k-го мешка для k от 1 до 1000 кроме 500-го, из 500-го мешка берем то же число монет в сумме (999). Засекаем полученную разность X. Во второе взвешивание делаем также, но на одну чашку кладем по k монет из k-го мешка для k от 1 до 1000 кроме 500-го, из 500-го мешка берем суммарное число монет (500000). Засекаем полученную разность Y. Теперь если X — ноль, фальшивых нет, целое Y/X равно номеру фальшивого мешка, нецелое Y/X показывает на 500-й мешок.