Тема 4. Центральная предельная теорема

Вспомогательная теория

Напомним формулировку центральной предельной теоремы.

Теорема 1. *Пусть X_i - н.о.р. случайные величины, $\mathbf{E}X_i = \mu$, $0 < \mathbf{D}X_i = \sigma^2 < \infty$. Тогда*

$$\frac{X_1 + \ldots + X_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

То есть

$$\mathbf{P}\left(\frac{X_1 + \ldots + X_n - \mu n}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x),\ n \to \infty,\ \forall x,$$

где $\Phi(x)$ – функция распределения $\mathcal{N}(0,1)$.

Обратите внимание, что, вообще говоря, сходимость по распределению не гарантирует сходимости плотностей, однако, справедлив следующий замечательный результат с которым вы познакомитесь в курсе дополнительных глав теории вероятностей.

Теорема 2. *Пусть X_i - н.о.р. случайные величины, причем $\int_{\mathbb{R}} |\psi_X(t)|^a dt < +\infty$ при некотором $a > 0$, где ψ – характеристическая функция. При этом $\mathbf{E}X_i = \mu$, $0 < \mathbf{D}X_i = \sigma^2 < \infty$. Тогда*

$$f_{(S_n - \mu n)/(\sigma\sqrt{n})}(x) \xrightarrow{d} \phi(x),$$

где ϕ – плотность стандартной нормальной величины.

Помимо самого факта сходимости функций распределений центрированных нормированных сумм известны также следующие результаты.

Теорема 3 (Неравенство Берри–Эссеена.)**.** *Пусть выполнены условия ЦПТ и дополнительно $\mathbf{E}|X|^3 < \infty$. Тогда*

$$\left|\mathbf{P}\left(\frac{X_1 + \ldots + X_n - \mu n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x)\right| \leq C\frac{\mathrm{E}|X_1 - \mathrm{E}X_1|^3}{(\mathrm{D}X_1)^{3/2}\sqrt{n}},$$

где C – некоторая константа, не зависящая от распределения X_i. По последним данным $C \leqslant 0.4784$.

Теорема 4. *Пусть $X_1, \cdots, X_n$ н.о.р. $\mathbf{E}X = \mu$, $\mathbf{D}X = \sigma^2$, $\mathbf{E}(X-\mu)^3 = \rho_3$.*

Справедливо также следующее асимптотическое разложение.

Пусть $\mathbf{E}(X-\mu)^3 = \rho_3$, $a_3 := \rho_3/\sigma^3$ – коэффициент асимметрии. Тогда

$$\mathbf{P}\left(\frac{X_1 + \ldots + X_n - \mu n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x) = \frac{a_3}{6\sqrt{2\pi n}}(1 - x^2)\exp\left(-\frac{x^2}{2}\right) + o\left(\frac{1}{\sqrt{n}}\right)$$

при $n \to \infty$.

Теорема 4 дает более точные приближения чем теорема 3, зато теорема 3 не предельная, а верна при всех n.

Задачи

1. Моделировать выборки $X_{i,j}$, $i \leq 1000$, $j \leq n$, где i) n=20 ii) n=100 величин из распределений а) Bern(1/2), б) $R[0,1]$, в) exp(1), г) Коши. Найти $S_{i,n} = \sum_{j=1}^{n} X_{i,j}$ и построить на одном графике ЭФР $S_{n,i}$ и ф.р. $\mathcal{N}(n\overline{X}, nS^2)$, где $\overline{X}$, S^2 – выборочное среднее и выборочная дисперсия всех имеющихся наблюдений. Похожи ли визуально полученные графики?

2. Пусть $X \sim Gamma(n, 4)$ Построить на одном графике графики плотности распределения с.в. $(X - \mathbf{E}X)/\sqrt{\mathbf{D}X}$ и плотности $\mathcal{N}(0, 1)$ для различных n.

3. Построить гистограмму по набору значений с.в. $S_n = X_1 + \ldots + X_n - \mu n$ (генерируем k выборок $X_1, \ldots, X_n$, по каждой находим одно значение суммы). На том же графике построить плотность распределения с.в. S_n (для дискретных – дискретное распределение) и плотность $\mathcal{N}(0, \sigma^2 n)$.

   Здесь распределения X_i рассматриваются следующие:

   - 1 вариант: $Poiss(\lambda)$, 2 вариант: $Geom(p)$,
   - 1 вариант: $exp(\lambda)$, 2 вариант: $Gamma(a, b)$.
   - * Для всех вариантов $R[0, 1]$ (для поиска плотности распределения суммы можно использовать sympy.stats.UniformSum() или написать формулу самостоятельно, см. распределение Ирвина–Холла),

4. Обозначим $Y = (S_n - n\mu)/(\sigma\sqrt{n})$. Построить на одном графике: $F_Y(x) - \Phi(x)$, правую часть неравенства Берри–Эссеена, ее же, умноженную на -1, правую часть асимптотического разложения. Рассмотреть $n = 5, 10, 20, 50, 100, 500$. Соотнести полученные результаты с теоремами 3 и 4. Рассмотрите следующие распределения X_i: $Bern(p)$, $exp(\lambda)$.