

## 12 ЭФР, критерий Колмогорова

**Определение 1.** Эмпирической функцией распределения называют функцию

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x}.$$

**Теорема 1** (Колмогорова). Пусть  $X_i \sim F$ , где  $F$  непрерывна, тогда

$$\mathbf{P}(\sqrt{n} \sup_x |\widehat{F}_n(x) - F(x)| \leq y) \rightarrow K(y),$$

где

$$K(y) = \sum_{k=-\infty}^{\infty} e^{-2k^2 y^2}$$

называют функцией распределения Колмогорова.

Отметим, что если распределение не непрерывно, то

$$\mathbf{P}(\sqrt{n} \sup_x |\widehat{F}_n(x) - F(x)| \leq y) \rightarrow K'(y),$$

где  $K'(y)$  зависит от  $F$ , но  $K'(y) \geq K(y)$  при всех  $y$ .

**Теорема 2** (Неравенство Дворецкого-Кифера-Вольфовица). Пусть  $X_i \sim F$ , тогда

$$\mathbf{P}(\sup_x |\widehat{F}_n(x) - F(x)| \geq y) \leq 2e^{-2ny^2}.$$

Отсюда мы можем узнать погрешность оценки  $F(x)$  функцией  $\widehat{F}_n(x)$ , построить доверительную полосу для  $F(x)$ . Отметим, что непосредственно критерий не слишком мощный и на практике, как правило, используют более мощный критерий Андерсона-Дарлинга (об этом в следующем семестре).

Критерий Колмогорова реализован в пакете `scipy`.

1. Пусть  $X_i \sim \mathcal{N}(0, 1)$ .

1) Построить эмпирическую функцию распределения (ЭФР) на одном графике с теоретической функцией распределения при разных  $n$  (проиллюстрировать сходимость ЭФР к ф.р.). То же для  $X_i \sim R[0, 1]$ ,  $X_i \sim \text{Bin}(m, p)$ , где  $m = 3$ ,  $m = 10$ .

2) Построить 95% доверительную полосу для ф.р., используя а) критерий Колмогорова. б) неравенство Дворецкого-Кифера-Вольфовица.

2. С помощью критерия Колмогорова проверить гипотезу  $H_0 : X_i \sim \mathcal{N}(0, 1)$ , если

$X_i = (Y_i - \mathbf{E}Y_i)/\sqrt{\mathbf{D}Y_i}$ , где

(а)  $Y_i \sim \mathcal{N}(-3, 8)$ ,

(б)  $Y_i \sim \text{Bin}(m, 1/2)$ ,

(с)  $Y_i$  имеют распределение Ирвина-Холла (сумма  $m$  независимых равномерных с.в.).

Построить графики  $p$ -value для  $m = 1, 2, 5, 20$ . Рассмотреть выборки длины  $n = 50, 100, 500$ .

3. Если выборка имеет нормальное распределение с неизвестными параметрами, то можно попробовать перейти от выборки  $X_i$  к  $(X_i - \bar{X})/S$  и применить к ним критерий Колмогорова нормальности  $\mathcal{N}(0, 1)$ . Построить график  $p$ -value при верной гипотезе и посмотреть корректно ли работает "модифицированный" критерий.