Домашнее задание 4

Юрасов Никита Андреевич

Обновлено 22 ноября 2019 г.

1 Проверка гипотез о виде распределения

1.1 Критерий согласия Колмогорова-Смирнова

Введем статистику, которая представляет собой максимальной отклонение эмпирической функции распределения $\hat{F}(x)$, построенной по выборке X, от гипотетической функции распределения $F(x)$:

$$D_n = D_n(X) = \sup_{-\infty<x<\infty} |\hat{F}(x) - F(x)|$$

Пусть существует $X = (X_1, ..., X_n)$ – выборка из $\mathcal{L}(\xi)$ с неизвестной функцией распределения $F_\xi(x)$, и пусть выдвинута гипотеза $H_0 : F_\xi(x) = F(x)$, где функция $F(x)$ полностью задана.

Для принятия или отвержения гипотезы H_0 необходимо по критерию Колмогорова сравнить $\sqrt{n}D_n$ с λ_α, которая определяется следующим равенством:

$$K(\lambda_\alpha) = 1 - \alpha,$$

где $K(x)$ – распределение Колмогорова.

На практике статистику D_n удобнее вычислять в следующем виде $D_n = max(D_n^+, D_n^-)$, где

$$D_n^+ = \max_{1\leqslant k\leqslant n} \left(\frac{k}{n} - F(X_{(k)})\right), \quad D_n^- = \max_{1\leqslant k\leqslant n} \left(F(X_{(k)}) - \frac{k-1}{n}\right)$$

Ответ на вопрос о виде распределения дает следующее сравнение:

- Если $\sqrt{n}D_n \geqslant \lambda_\alpha$, то гипотеза H_0 отвергается;
- Если $\sqrt{n}D_n \leqslant \lambda_\alpha$, то гипотеза H_0 принимается.

В неравенстве можно воспользоваться поправкой Большева о статистике $S(D_n)$, которая быстрее сходится к распределению Колмогорова:

$$S = \frac{6nD_n + 1}{6\sqrt{n}}$$

Преимущества
Критерий согласия Колмогорова начинает эффективно работать при выборке объемом $n \geqslant 20$, что допускает использование его при достаточно малых выборках данных.

Недостатки
Критерий Колмогорова-Смирнова применяется только для непрерывных распределениях. Также вычисление статистики D_n предполагает достаточно большие аналитические вычисления, что затрудняет проверку.