41. Проверка гипотез о совпадении дисперсий двух нормальных совокупностей.

Status	Completed

Теорема Фишера

n, m - размеры независимых нормальных выборок X, Y соответственно, тогда (если дисперсии совпадают)

$$\frac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} \sim F_{n-1,m-1}$$

Доказательство. По лемме Фишера, независимые случайные величины

$$\xi^2_{n-1} = \frac{(n-1)\, S_0^2(\mathbf{X})}{\sigma_1^2} \quad \text{и} \quad \xi^2_{m-1} = \frac{(m-1)\, S_0^2(\mathbf{Y})}{\sigma_2^2}$$

имеют распределения H_{m-1} и H_{n-1} соответственно. При $\sigma_1^2 = \sigma_2^2$ отношение

$$\frac{\xi^2_{n-1}/(n-1)}{\xi^2_{m-1}/(m-1)} = \frac{S_0^2(\mathbf{X})}{\cancel{\sigma_1^2}} \cdot \frac{\cancel{\sigma_2^2}}{S_0^2(\mathbf{Y})} = \rho(\mathbf{X}, \mathbf{Y})$$

имеет распределение Фишера с $n-1$, $m-1$ степенями свободы по определению 18 и совпадает с $\rho(\mathbf{X}, \mathbf{Y})$. □

Критерий Фишера (о совпадении дисперсий)

f - квантиль данного распределения.

$$d = \frac{S_0^2(X)}{S_0^2(Y)} \sim F_{n-1,m-1}$$

$$\delta = \begin{cases} 0, & f_{\epsilon/2} < d < f_{1-\epsilon/2} \\ 1, & \text{иначе} \end{cases}$$

☝️ Опять же, думай про квантили! Тут распределение Фишера, оно не симметричное, так что берем разные квантили!

Свойства

1. Критерий состоятельный.

Доказательство состоятельности критерия Фишера.

Покажем, что последовательность квантилей $f_\delta = f_\delta(n,m)$ любого уровня $0<\delta<1$ распределения $F_{n,m}$ сходится к 1 при $n,m \to \infty$. Возьмем величину $f_{n,m}$ с этим распределением. По определению, $P(f_{n,m}<f_\delta) = \delta$, $P(f_{n,m}>f_\delta) = 1-\delta$ при всех n,m. По свойству 2 распределения Фишера, $f_{n,m} \xrightarrow{p} 1$. Поэтому для любого $\epsilon>0$ обе вероятности $P(f_{n,m}<1-\epsilon)$ и $P(f_{n,m}>1+\epsilon)$ стремятся к нулю при $n,m\to\infty$, становясь рано или поздно меньше как δ, так и $1-\delta$. Следовательно, при достаточно больших n,m выполнено $1-\epsilon<f_\delta<1+\epsilon$.

Для доказательства состоятельности осталось предположить, что гипотеза H_1 не верна, взять ϵ равное, например, половине расстояния от 1 до σ_1^2/σ_2^2 и использовать сходимость (27). Пусть, скажем, при достаточно больших n и m

$$\frac{\sigma_1^2}{\sigma_2^2} < \frac{\sigma_1^2}{\sigma_2^2} + \epsilon = 1-\epsilon < f_{\varepsilon/2}.$$

Тогда вероятность ошибки второго рода удовлетворяет неравенствам

$$\alpha_2(\delta) = P_{H_2}(f_{\varepsilon/2} \leqslant \rho \leqslant f_{1-\varepsilon/2}) \leqslant P_{H_2}(1-\epsilon<\rho) = P_{H_2}\left(\rho > \frac{\sigma_1^2}{\sigma_2^2}+\epsilon\right) \to 0.$$

Аналогично рассматривается случай, когда (при достаточно больших n и m)

$$f_{1-\varepsilon/2} < 1+\epsilon = \frac{\sigma_1^2}{\sigma_2^2} + \epsilon < \frac{\sigma_1^2}{\sigma_2^2}.$$

2. Критерий имеет размер $1-\epsilon$.
3. Реально допустимый уровень значимости: $\epsilon^* = P(\eta \geq d) = 1 - F_{n-1,m-1}(d)$.

☝️ Если вы знаете среднее, то можно воспользоваться свойством $\sum\limits_{i=1}^{n}(\frac{X_i-a}{\sigma})^2 = \frac{nS_1^2}{\sigma^2} \sim \chi_n^2$, и сделать такой же критерий (только распределение будет $F_{n,m}$.