Index disperze pro diskrétní rozdělení

Valerij Semjonov

31. Března 2020

Osnova

Úvodní motivace

- Index disperze udává poměr rozptylu vzhledem ke střední hodnotě a jako charakteristika rozdělení bývá uvažován především pro diskrétní rozdělení.
- Jeho výběrový ekvivalent se pak využívá např. pro testování shody s Poissonovým rozdělením.
- Jak je známo, pro Poissonovo rozdělení platí, že střední hodnota a rozptyl se rovnají.
- Na základě toho, jak moc se liší poměr výběrového rozptylu a výběrového průměru od 1, jsme schopni usuzovat, zda analyzovaná data pochází z Poissonova rozdělení.

Základní pojmy

Definice (Poissonovo rozdělení)

Řekneme, že náhodná veličina X má *Poissonovo rozdělení*, jestliže

$$P[X = x] = \frac{\lambda^x}{x!} e^{-\lambda}$$

pro všechna x = 0, 1, 2... a $\lambda \geq 0$.

Základní pojmy

Definice (Výběrový průměr a výběrový rozptyl)

Nechť $\mathbf{X} = (X_1, X_2, ..., X_n)$ je náhodný výběr o rozsahu n. Potom

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

nazýváme *výběrovým průměrem* a

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

výběrovým rozptylem náhodného výběru $\mathbf{X}$.

Základní pojmy

Definice (Index disperze)

Nechť **X** je náhodná veličina, μ její střední hodnota a σ^2 její rozptyl. Potom číslo

$$\frac{\sigma^2}{\mu}$$

nazýváme index disperze náhodné veličiny **X**. Jejím výběrovým ekvivalentem pak rozumíme

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X}_n)^2}{\bar{X}_n} = (n-1)\frac{S_n^2}{\bar{X}_n}.$$

Základní vlastnosti

- Pochází-li náhodný výběr **X** z Poissonova rozdělení, pak $(n-1)\frac{S_n^2}{\bar{X}_n}$ má asymptotické rozdělení χ^2 o n - 1 stupních volnosti [B.Selby, Biometrika, Vol. 52, No. 3/4 (Dec., 1965), pp. 627-629].
- Odvodíme netriviální výsledek, že v případě Poissonova rozdělení je střední hodnota $\frac{S_n^2}{\bar{X}_n}$ rovna jedné, tedy

$$E(\frac{S_n^2}{\bar{X}_n}) = 1.$$

[John J. Bartko, Samuel W. Greenhouse and Clifford S. Patlak, Biometrics, Vol. 24, No. 1 (Mar., 1968), pp. 97-102]

Základní vlastnosti

- Využijeme toho, že náhodné veličiny $X_1, X_2, ..., X_n$ jsou nezávislé a stejně rozdělené, neboť se jedná o náhodný výběr.
- Sdružené rozdělení $X_1, X_2, ..., X_n$ je tedy součinem jednorozměrných rozdělení a platí

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = \frac{e^{-n\lambda}}{\prod_{i=1}^{N} x_i!} \lambda^{\sum_{i=1}^{n} X_i}$$

- Tedy

$$E(\frac{S_n^2}{\bar{X}_n}) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} ... \sum_{x_n=0}^{\infty} \frac{S_n^2}{\bar{X}_n} \frac{e^{-n\lambda}}{\prod_{i=1}^{N} x_i!} \lambda^{\sum_{i=1}^{n} X_i}$$

Základní vlastnosti

- Derivováním posledního výrazu podle λ dostane po úpravách

$$\frac{\mathrm{d}E(\frac{S_n^2}{\bar{X}_n})}{\mathrm{d}\lambda} = \frac{n}{\lambda}E(S_n^2) - nE(\frac{S_n^2}{\bar{X}_n})$$

- Víme, že $E(S_n^2) = \sigma^2$, neboť výběrový rozptyl je nestranným odhadem skutečného roztpylu.
- V našem případě je $E(S_n^2) = \lambda$. Díky tomu lze rovnici výše upravit na diferenciální rovnici

$$\frac{\mathrm{d}E(\frac{S_n^2}{\bar{X}_n})}{\mathrm{d}\lambda} = 1 - Ke^{-n\lambda},$$

kde K je konstanta.

Základní vlastnosti

- Řešením této rovnice a úvahou dojdeme k závěru, že $E(\frac{S_n^2}{\bar{X}_n}) = 1$ pro všechna λ.
- Lze studovat index disperze i u jiných diskrétních rozdělení. Například u binomického rozdělení je index disperze menší než 1.

Statistický test

- Chceme testvat hypotézu H_0: Náhodný výběr pochází z Poissonova rozdělení oproti alternativě H_1: Náhodný výběr nepochází z Poissonova rozdělení
- Testová statistika
$$VT = (n-1)\frac{S_n^2}{\bar{X}_n}$$
(variance test)
- Ve své bakalářské práci mám v plánu podrobněji popsat statistické testy výše uvedené hypotézy a provést praktickou studii.

Díky za pozornost!