Linearni statistički modeli

Bojana Miloševič

Raspodela kvadratne forme

Theorem (Kohran)

Neka su $X_1, ..., X_n$ nezavisne $\mathcal{N}(0, \sigma^2)$ slučajne velićine i neka je

$$\sum_{i=1}^{n} X_i^2 = \sum_{j=1}^{k} Q_j,$$

gde je Q_j kvadratna forma definisana sa $Q_j = X^T A_j X$, za $j = 1, 2, ..., k$, pri čemu je $R(A_j) = r_j$. Tada je $\sum_{j=1}^{k} r_j = n$ ako i samo ako

1. *$Q_1, ..., Q_k$ su nezavisne slučajne veličine i*
2. *Q_j/σ^2 ima $\chi^2_{r_j}$ raspodelu.*

Primer 1

Uzoračka sredina $\bar{X}$ i popravljena uzoračka disperzija S^2 su nezavine slučajne veličine, i $(n-1)S^2/\sigma^2$ ima χ^2_{n-1} raspodelu

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

$$X^T X = X^T \Big(I - \frac{J}{n}\Big) X + X^T \frac{J}{n} X$$

Primer 2

$$\frac{SSTO}{\sigma^2} \sim \chi^2_{n-1}$$
$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-p-1}$$
$$\frac{SSR}{\sigma^2} \sim \chi^2_{p}$$

i nezavisne su

Dokaz

Lemma

Neka su $x_1, ... x_n$ *realni brojevi. Pretpostavimo da se suma* $\sum_{i=1}^{n} x_i^2$ *mozhe predstaviti kao zbir* k *kvadratnih formi* $\sum_{j=1}^{k} Q_j$ *gde je* $Q_i = x^T A_i x$ *i* $R(A_i) = r_i$*. Ako je* $\sum_{i=1}^{k} r_i = n$ *tada postoji ortogonalna matrica* M*, takva da za* $x = My$ *vazhi*

$$
\begin{aligned}
Q_1 &= y_1^2 + \cdots y_{r_1}^2 \\
Q_2 &= y_{r_1+1}^2 + \cdots y_{r_1+r_2}^2 \\
&\cdots \\
Q_k &= y_{n-r_k+1}^2 + \cdots y_n^2
\end{aligned}
$$

Raspodela kvadratne forme

Theorem

Neka slučajni vektor X ima $\mathcal{N}(\theta, \sigma^2 I)$ raspodelu i neka je

$$\sum_{i=1}^{n} X_i^2 = \sum_{j=1}^{k} Q_j,$$

gde je Q_j kvadratna forma definisana sa $Q_j = X^T A_j X$, za $j = 1, 2, ..., k$, pri čemu je $R(A_j) = r_j$. Tada je $\sum_{j=1}^{k} r_j = n$ i $\sum_{j=1}^{k} \mu_j = \theta^T \theta$ ako i samo ako

1. *$Q_1, ..., Q_k$ su nezavisne slučajne veličine i*
2. *Q_j/σ^2 ima $\chi^2_{r_j}(\mu_j)$ raspodelu, pri čemu je $\mu_j = \theta^T A_j \theta$*

Raspodela kvadratne forme

Neka slučajni vektor X ima $N(\theta, \sigma^2 I)$ raspodelu i neka je $Q_1 = X^T A_1 X$ i $Q_2 = X^T A_2 X$, gde su A_1 i A_2 dve simetrične matrice. Tada su Q_1 i Q_2 nezavisne ako i samo ako je $A_1 A_2 = 0$.

Gausov model

Funkcija verodostojnosti

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^2}}$$

Ocena metodom maksimalne verodostojnosti za β poklapa sa ocenom metodom najmanjih kvadrata, odnosno

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Ocena za σ^2 je

$$\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

Gausov model

$$L(\hat{\beta}, \hat{\sigma^2}) = \frac{1}{(2\pi\hat{\sigma}^2)^{\frac{n}{2}}} e^{-\frac{1}{2n}} \sim \hat{\sigma}^{-n}$$

$$\hat{\beta} \sim N(\beta, (X^T X)^{-1}\sigma^2)$$

Testiranje linearne hipoteze

$$H_0 : C\beta = \gamma,$$

gde je C $m \times (p+1)$ matrica.

Test količnika verodostojnosti

Treba nam minimum $S(\beta) = \frac{(Y-X\beta)^T(Y-X\beta)}{n}$ uz uslov $C\beta - \gamma = 0$.

Sistem normalnih jednačina

$$\mathcal{L}(\beta, a) = (Y - X\beta)^T(Y - X\beta) - a^T(C\beta - \gamma).$$

$$\frac{\partial}{\partial\beta}\mathcal{L}(\beta, a) = -2X^TY + 2X^TX\beta - C^Ta = 0$$
$$\frac{\partial}{\partial a}\mathcal{L}(\beta, a) = \gamma - C\beta = 0.$$

Rešenje

$$\hat{\beta}_0 = \hat{\beta} + (XX^T)^{-1}C^T[C(XX^T)^{-1}C^T]^{-1}(\gamma - C\hat{\beta})$$

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{(Y-\hat{\beta}_0X)^T(Y-\hat{\beta}_0X)}{n} \\ &= \frac{1}{n}\Big((Y-\hat{\beta}X)^T(Y-\hat{\beta}X) + (C\hat{\beta}-\gamma)^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta}-\gamma)\Big)\end{aligned}$$

$$L(\hat{\beta}_0, \hat{\sigma^2}_0) \sim \hat{\sigma}_0^{-n}$$

Test količnika verodostojnosti

Pretpostavimo da je $H_1 : \beta X \neq \gamma$. Tada je količnik verodostojnosti

$$\lambda = \frac{\max\limits_{H_1 \cup H_0} L(\beta, \sigma^2)}{\max\limits_{H_0} L(\beta, \sigma^2)} = \frac{L(\hat{\beta}, \hat{\sigma}^2)}{L(\hat{\beta}_0, \hat{\sigma}_0^2)} = \Big(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\Big)^{\frac{n}{2}}$$

$$= \Big(1 + \frac{(C\hat{\beta} - \gamma)^T [C(X^T X)^{-1} C^T]^{-1} (C\hat{\beta} - \gamma)}{e^T e}\Big)^{\frac{n}{2}}$$

$W = \{\lambda > c\}$.

Ako važi H_0 :

$$C\hat{\beta} - \gamma \sim N_m(C\beta - \gamma, \sigma^2 C(X^T X)^{-1} C^T)$$

$$Q = (C\hat{\beta} - \gamma)^T [C(X^T X)^{-1} C^T]^{-1} (C\hat{\beta} - \gamma) \sim \sigma^2 \chi^2_m$$

$$\begin{aligned} Q &= (C\hat{\beta} - C\beta)^T [C(X^T X)^{-1} C^T]^{-1} (C\hat{\beta} - C\beta) \\ &= (C(X^T X)^{-1} X^T \varepsilon)^T [C(X^T X)^{-1} C^T]^{-1} C(X^T X)^{-1} X^T \varepsilon \\ &= \varepsilon^T X(X^T X)^{-1} C^T [C(X^T X)^{-1} C^T]^{-1} C(X^T X)^{-1} X^T \varepsilon \\ &= \varepsilon^T P \varepsilon, \end{aligned}$$

gde je $P = X(X^T X)^{-1} C^T [C(X^T X)^{-1} C^T]^{-1} C(X^T X)^{-1} X^T$ #

$$e^T e = SSE = \varepsilon^T M \varepsilon \sim \sigma^2 \chi^2_{n-p-1}$$

Raspodela test statistike

$$PM = 0$$

$$\frac{\frac{Q}{m}}{\frac{e^T e}{n-p-1}} \sim F_{m,n-p-1}$$

Primer

Pretpostavimo da testiramo nultu hipotezu $H_0 : \beta_1 = \beta_2 = \cdots \beta_p = 0$ protiv komplementarne hipoteze. Tada je

$$L(\hat{\beta}_0, \hat{\sigma}_0^2) \sim \Big(\frac{SSTO}{n}\Big)^{-\frac{n}{2}}.$$

Odavde, je

$$\lambda = \Big(\frac{SSTO}{SSE}\Big)^{\frac{n}{2}} = \Big(1 + \frac{SSR}{SSE}\Big)^{\frac{n}{2}}.$$

Primenom Kohranove teoreme, ili teoreme o nezavisnosti kvadratnih formi dobija se da

$$\frac{\frac{SSR}{p}}{\frac{SSE}{n-p-1}} \sim F_{p,n-p-1}.$$

Primer

U slučaju da želimo da testiramo hipotezu da je $\beta_k = 0$

$$\frac{Q}{\frac{e^T e^T}{n-p-1}} \sim F_{1,n-p-1}.$$