Matematisk Statistik

15. Forelæsning 23.03.2019

Ute Hahn

Hypotesetests i klassisk statistik

- Styrkefunktion
- Forsøgsplanlægning
- Interpretation af testresultater;
  p-værdier, konfidensintervaller og signifikanstests
- Likelihood ratio tests, og Neyman-Pearson lemma

- p-værdi er sandsynligheden for et mindst så ekstremt resultat, givet H_0 er sand.
- statistisk signifikans vs praktisk betydning: er alt som er signifikant også betydningsfuld? og eller omvendt?
- Et „ikke signifikant" resultat betyder *ikke* at „H_0 er sandt".
- Mange „fisker" efter signifikante sammenhæng – hvad er konsekvenserne?

Likelihood ratio test: Motivation

Situation: To konkurrerende modeller, beskrevet ved en nul- og alternativhypotese i en parametrisk model $X_i \sim F(\cdot;\theta)$:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_A : \theta = \theta_A.$$

Formål: Test skal afgøre, hvilken model der passer bedst til en given stikprøve.

Test til simple hypoteser i en parametrisk model $X_i \sim F(\cdot;\theta)$:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_A : \theta = \theta_A. \tag{8.3}$$

Husk likelihood:

$$L(\theta \mid x_1, \ldots, x_n) = f(x_1, \ldots, x_n \mid \theta)$$

Neyman og Pearson foreslog, at bruge teststørrelse Q givet ved

$$T(x_1, \ldots, x_n) = \frac{L(\theta_0 \mid x_1, \ldots, x_n)}{L(\theta_A \mid x_1, \ldots, x_n)} = \frac{\prod_{i=1}^n f(x_i;\theta_0)}{\prod_{i=1}^n f(x_i;\theta_A)}. \tag{8.4}$$

Typisk bliver

$$\frac{f(X;\theta_0)}{f(X;\theta_A)} = \begin{cases} \text{stor, under } H_0\text{:} & X \sim F(.;\theta_0) \\ \text{lille, under } H_A\text{:} & X \sim F(.;\theta_A) \end{cases}$$

☞ forkast H_0 for når Q er lille.

Model: $X_i \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$, dvs. tæthedsfunktion $f(x;\lambda) = \lambda e^{-\lambda x}$.

Hypoteser:

$$H_0 : \lambda = \lambda_0 \quad \text{versus} \quad H_A : \lambda = \lambda_A.$$

Stikprøve: $x_1, \dots, x_n$.

(i bogen: $\lambda_0 = 8$, $\lambda_A = 10$, $n = 9$)

Likelihood ratio:

$$\begin{aligned} Q(x_1,\dots,x_n) = \frac{\prod_{i=1}^n f(x_i;\lambda_0)}{\prod_{i=1}^n f(x_i;\lambda_A)} &= \frac{\prod_{i=1}^n \lambda_0 \exp(-x_i\lambda_0)}{\prod_{i=1}^n \lambda_A \exp(-x_i\lambda_A)} \\ &= \frac{\lambda_0^n \exp(-\lambda_0 \sum_{i=0}^n x_i)}{\lambda_A^n \exp(-\lambda_A \sum_{i=0}^n x_i)} \\ &= \left(\frac{\lambda_0}{\lambda_A}\right)^n \exp\Big(-(\lambda_0 - \lambda_A)\sum_{i=0}^n x_i\Big). \end{aligned}$$

Eksempel 8.20 fortsat: Fra likelihood ratio til teststørrelse til test

Husk: vi forkaster ved små værdier af Q.

☞ Find et kritisk område $(-\infty, C]$ til Q således, at

$$P(Q(X_1,\dots,X_n) < C_Q \mid X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(\lambda_0)) \overset{!}{\leq} \alpha$$

Likelihood ratio Q har her formen

$$Q(x_1,\dots,x_n) = \left(\frac{\lambda_0}{\lambda_A}\right)^n \exp\Big(-(\lambda_0 - \lambda_A)\sum_{i=0}^{n} x_i\Big).$$

☞ om Q bliver lille, afhænger af summen $x. = \sum_{i=1}^n x_i$ og fortegnet af $(\lambda_0 - \lambda_A)$.

$$Q(\boldsymbol{x}) \text{ bliver lille, når } x. \text{ er } \begin{cases} \text{lille}, & \text{hvis } \lambda_0 > \lambda_A \\ \text{stor}, & \text{hvis } \lambda_0 < \lambda_A \end{cases}$$

☞ Brug $X. = \sum_{i=1}^n x_i$ som teststørrelse.

Eksempel 8.20 fortsat: Fra teststørrelse til test, næsten i mål

Fra $X_i \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda_0) = \Gamma(1, \lambda_0)$ følger $X. = \sum_{i=0}^{n} X_i \sim \Gamma(n, \lambda_0)$, se Theorem B.9.4, side 499.

Ved eksemplet i bogen, $\lambda_0 = 8, \lambda_A = 10, n = 9$:

Da $\lambda_A > \lambda_0$ forkastes H_0 når $x.$ er mindre end α-fraktilen af $\Gamma(9, 8)$-fordelingen.

Theorem 8.6.1 (The Neyman-Pearson Lemma)

Blandt alle mulige tests med samme signifikansniveau α i en fordelingsfamilie $\{F(.;\theta)\}$, af to simple hypoteser

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_A : \theta = \theta_A$$

er LR-testen givet ved

$$Q = \frac{L(\theta_0)}{L(\theta_A)}; \quad \text{forkast } H_0 \text{ når } Q < c$$

den med den mindste sandsynlighed for type 2 fejl.

Bemærk:

- En test opfattes som en regel, hvorefter udfald af stikprøven $X_1, \ldots, X_n$ udmøntes i e beslutning for enten H_0 eller H_A.
- Der findes flere ækvivalente tests, fx fører teststørrelsen Q i eksempel 8.20 til de samme resultater som teststørrelsen $X.$.

 (se bogen: 8.6.1)

Definition 8.5
En simpel hypotese fastlægger fordelingen under H_0 entydigt. En hypotese der ikke er simpel kaldes for sammensat.

I parametriske modeller tager simple hypoteser formen $H_0 : \theta = \theta_0$.

Eksempler:

- i Poisson modellen Poisson(λ) er $H_0 : \lambda = 2$ en simpel hypotese, og både $H_1 : \lambda \in \{2, 2.5\}$ eller $H_2 : \lambda > 2$ eller $H_3 : \lambda \neq 2$ er sammensat.
- i normalmodellen med kendt varians $N(\mu, \sigma_0^2)$ er $H_0 : \mu = 4$ en simpel hypotese.
- i normalmodellen med ukendt varians $N(\mu, \sigma^2)$, er $H_0 : \mu = 4$ en sammensat hypotese, fordi fordelingen ikke er helt fastlagt ved μ.

Betragt en model med parameter $\theta \in \Theta$ og likelihoodfunktion L.

Definition 8.6: LR-test for hypoteserne

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_A : \theta \in \Theta_A$$

er givet ved

$$Q(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\theta \in \Theta_A} L(\theta \mid \mathbf{x})}, \quad \text{forkast } H_0 \text{ når } Q(\mathbf{x}) < c,$$

hvor $c > 0$ vælges således at sandsynligheden for en type 1 fejl er maksimalt α.

Bemærk:

- $\sup_{\theta \in \Theta} L(\theta \mid \mathbf{x}) = L(\hat{\theta})$, hvor $\hat{\theta}$ er maksimum-likelihood estimatoren til θ i det fulde model.
- $\sup_{\theta \in \Theta_0} L(\theta \mid \mathbf{x}) = L(\hat{\theta}_0)$, hvor $\hat{\theta}_0$ er maksimum-likelihood estimatoren til θ i det reducerede model.

Model

$$X_i \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma_0^2), \quad i = 1, \ldots, n, \quad \mu \in \Theta = \mathbb{R}.$$

Likelihood funktion

$$L(\mu) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

Log likelihood funktion

$$\ln L(\mu) = -\frac{n}{2}\ln(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Likelihood ligningen

$$0 = \frac{\partial \ln L(\mu)}{\partial \mu} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu).$$

Maksimum likelihood estimat

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

LR-test i eksempel 8.23

Model med $\sigma_0^2 = 1$

$$X_i \overset{\text{i.i.d.}}{\sim} N(\mu, 1), \quad i = 1, \ldots, n, \quad \mu \in \Theta = \mathbb{R}$$

Hypoteser

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu < \mu_0.$$

Teststørrelse

$$Q(\mathbf{x}) = \frac{\max_{\mu \in \Theta_0} L(\mu \mid \mathbf{x})}{\max_{\mu \in \Theta} L(\mu \mid \mathbf{x})} = \frac{L(\mu_0 \mid \mathbf{x})}{\max_{\mu < \mu_0} L(\mu \mid \mathbf{x})} = \begin{cases} \frac{L(\mu_0)}{L(\bar{x})}, & \bar{x} < \mu_0 \\ 1, & \bar{x} \geq \mu_0. \end{cases}$$

$$\begin{aligned} \frac{L(\mu_0)}{L(\bar{X})} &= \frac{\exp(-1/2 \sum_{i=1}^n (X_i - \mu_0)^2)}{\exp(-1/2 \sum_{i=1}^n (X_i - \bar{X})^2)} \\ &= \exp\left[\frac{1}{2}(-\sum_{i=1}^n (X_i - \mu_0)^2 + \sum_{i=1}^n (X_i - \bar{X})^2)\right] = \exp(-\frac{n}{2}(\bar{X} - \mu_0)^2). \end{aligned}$$

Teststørrelsen bliver lille = ekstrem, når $(\bar{X} - \mu_0)^2$ bliver stor.

Tosided LR-test i normalfordeling med kendt varians

Model

$$X_i \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma_0^2), \quad i = 1, \ldots, n, \quad \mu \in \Theta = \mathbb{R}.$$

Hypoteser

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu \neq \mu_0.$$

Teststørrelse

$$\begin{aligned}
Q(\mathbf{X}) &= \frac{\exp(-1/(2\sigma_0^2)\sum_{i=1}^n (X_i - \mu_0)^2)}{\exp(-1/(2\sigma_0^2)\sum_{i=1}^n (X_i - \bar{X})^2)} \\
&= \exp\left[\frac{1}{2\sigma_0^2}(-\sum_{i=1}^n (X_i - \mu_0)^2 + \sum_{i=1}^n (X_i - \bar{X})^2)\right] \\
&= \exp(-\frac{n}{2\sigma_0^2}(\bar{X} - \mu_0)^2).
\end{aligned}$$

„-2 log LR-Teststørrelse“ eller „-2 ln Q“:

$$-2\ln Q(\mathbf{X}) = \frac{(\bar{X} - \mu_0)^2}{\sigma_0^2/n} \sim \chi_1^2, \quad \text{fordi} \quad \frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2/n}} \sim N(0, 1)$$