STAT 5430

Lec 29, M, Apr 7

- Homework 7 posted, due, M, Apr 7 – No homework this week!
  ↑ testing

- Exam 2 is coming up (≈ 1 week away)
  on W, April 16, 6:15–8:15 PM, 3rd floor seminar room
- No class on that W.
- I'll post: study guide (sufficiency/completeness/tests)
  - practice exams
  - bring new 1 page (front/back) formula sheet on exam 2 material
    (I'll post one to use if you'd like)
  - can bring calculator & previous formula sheet for exam 1
  - I'll provide table of distributions / STAT 542 facts on test as before

Hypothesis Testing II

Likelihood Ratio Tests: Large Sample Calibrations

$\leftarrow \lambda(\underset{\sim}{x})$

The following result describes the asymptotic distribution of the likelihood ratio statistic (under appropriate regularity conditions) & may be used to calibrate a LRT in a simple fashion when the sample size n is "sufficiently large."

Theorem: Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ be iid random vectors with common pdf/pmf $f(\underset{\sim}{x}|\theta)$, $\underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^p$ (the parameter $\underset{\sim}{\theta}$ can be vector-valued). Let $\lambda_n(\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n)$ denote the likelihood ratio statistic based on $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ for testing $H_0 : \underset{\sim}{\theta} \in \Theta_0 \subset \mathbb{R}^p$ vs $H_1 : \underset{\sim}{\theta} \notin \Theta_0$, where Θ_0 has the form

$r = \#$ of parameters to be tested $(1 \le r \le p)$

$$\Theta_0 = \Big\{ \underset{\sim}{\theta} \equiv (\theta_1, \ldots, \theta_p) \in \Theta : \underbrace{\theta_1 = \theta_1^0, \ldots, \theta_r = \theta_r^0}_{\substack{\text{hypothesized values} \\ \text{for first } r \le p \text{ parameters}}} \Big\}$$

claimed values under H_0

for some $\theta_1^0, \ldots, \theta_r^0$, $r \le p$. That is, from the p parameters, we make a claim about exactly r of these parameters and the hypotheses are

$$\text{“}H_0 : \theta_1 = \theta_1^0, \ldots, \theta_r = \theta_r^0\text{” vs “}H_1 : \theta_i \ne \theta_i^0 \text{ for some } 1 \le i \le r\text{”}$$

$\leftarrow$ "not H_0"

Then, under the Cramér-Rao type regularity conditions, it holds that: $\leftarrow$ e.g. hold for exponential families

$$\textit{if } H_0 \textit{ is true,} \quad -2\log \lambda_n(\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n) \xrightarrow{d} \chi_r^2 \text{ as } n \to \infty.$$

$\leftarrow \#$ of parameters under H_0

$\uparrow$ converges in distribution to χ^2_r

Remark: The above limiting distribution suggests the following testing procedure based on the $(1-\alpha)$-quantile of a χ_r^2 distribution, denoted as $\chi^2_{1-\alpha}(r)$ for which $P\big(\chi_r^2 \le \chi^2_{1-\alpha}(r)\big) = 1 - \alpha$ and $P\big(\chi_r^2 > \chi^2_{1-\alpha}(r)\big) = \alpha$.

recall: We reject H_0 if $\lambda(\underset{\sim}{x}) \in [0,1]$ is too small $\Rightarrow$ $-2\log\lambda(\underset{\sim}{x})$ is too big

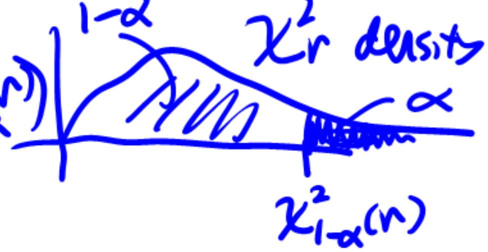
Namely, for n large (e.g., say $\cancel{n \ge 30}$ observations),

$$\varphi(\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n) = \begin{cases} 1 & \text{if} \quad -2\log \lambda_n(\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n) > \chi^2_{1-\alpha}(r) \\ 0 & \text{otherwise} \end{cases}$$

or $\lambda_n(X_1, \ldots, X_n) < e^{-\chi^2_{1-\alpha}(r)/2}$

is an approximate size α LRT for testing "$H_0 : \theta_1 = \theta_1^0, \ldots, \theta_r = \theta_r^0$" vs "$H_1 : \theta_i \ne \theta_i^0$ for some $1 \le i \le r$."

i.e. $\max_{\theta \in \Theta_0} E\phi(\underset{\sim}{X}) = \max_{\theta \in \Theta_0} P_\theta\big(-2\log\lambda(\underset{\sim}{X}) > \chi^2_{1-\alpha}(r)\big) \approx \alpha$



Hypothesis Testing II

Likelihood Ratio Tests + Large Sample Calibration: Illustration

Example: Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ (← ← data points as vectors in $\mathbb{R}^2$) be iid $N_2(\mu, A)$ random vectors, where $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$ (unknown → μ, A ← known) and A is a known 2×2 positive definite matrix. Find a size α LRT for testing $H_0 : 2\mu_1 + 3\mu_2 = 0$ vs $H_1 : 2\mu_1 + 3\mu_2 \neq 0$, using χ^2-calibration.

Solution: Let $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ where $\theta_1 = 2\mu_1 + 3\mu_2$ ←

$\theta_2 = \mu_2$

Note: $\underset{\sim}{\theta} = \underbrace{\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}}_{B^{-1}} \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}_{\underset{\sim}{\mu}}$

$\Rightarrow \underset{\sim}{\mu} = \underbrace{\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}^{-1}}_{B} \underset{\sim}{\theta} \equiv B\underset{\sim}{\theta}$

Hence, $\underset{\sim}{X}_1, \ldots \underset{\sim}{X}_n$ iid $N_2(\overset{\underset{\sim}{\mu}}{B\underset{\sim}{\theta}}, A)$ and, in terms of $\underset{\sim}{\theta}$, the testing problem is $\underbrace{H_0 : \theta_1 = 0 \text{ vs } H_1 : \theta_1 \neq 0}_{\text{this } H_0 \text{ form needed for } \chi^2\text{-calibration for } \lambda(\underset{\sim}{X})}$

So, we reject H_0 if $-2\log\lambda(\underset{\sim}{X}) > \chi^2_{1-\alpha}(1)$

or $\lambda(\underset{\sim}{X}) < e^{-\chi^2_{1-\alpha}(1)/2}$

Need to find LRS $\lambda(\underset{\sim}{X})$ as follows:

$L(\underset{\sim}{\theta}) \equiv$ joint pdf of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n = f(\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n \mid \underset{\sim}{\theta})$

$$= \prod_{i=1}^{n} \left(\left(\frac{1}{2\pi\sqrt{|A|}} \right) e^{-\frac{1}{2}(\underset{\sim}{X}_i - B\underset{\sim}{\theta})^T A^{-1} (\underset{\sim}{X}_i - B\underset{\sim}{\theta})} \right)$$

← $|A| \equiv \det(A)$

$$= \left(\frac{1}{2\pi}\frac{1}{\sqrt{|A|}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (\underset{\sim}{X}_i - B\underset{\sim}{\phi})^T A^{-1} (\underset{\sim}{X}_i - B\underset{\sim}{\phi})}$$

$$= \left(\frac{1}{2\pi}\frac{1}{\sqrt{|A|}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (\underset{\sim}{Y}_i - \underset{\sim}{\phi})^T B^T A^{-1} B (\underset{\sim}{Y}_i - \underset{\sim}{\phi})}$$

where $\underset{\sim}{Y}_i = B^{-1}\underset{\sim}{X}_i = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}\begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix} = \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix}$

So, the MLE over Θ is $\hat{\underset{\sim}{\phi}} = \underset{\sim}{\bar{Y}}_n = \frac{1}{n}\sum_{i=1}^n \underset{\sim}{Y}_i$

(just sample means)

Under $H_0: \phi_1 = 0$,

$f(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n \mid \underset{\sim}{\phi} = (\phi_1, \phi_2), \phi_1 = 0)$ ← likelihood under H_0

$$= \left(\frac{1}{2\pi\sqrt{|A|}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n \{\sigma_{22}(Y_{i2}-\phi_2)^2 + 2\sigma_{12}Y_{i1}(Y_{i2}-\phi_2) + \sigma_{11}Y_{i1}^2\}}$$

where $B^T A^{-1} B = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$

& maximizer for $\underset{\sim}{\phi}$ over Θ_0 is

$\underset{\sim}{\tilde{\phi}} = (0, \tilde{\phi}_2)$, $\tilde{\phi}_2 = \bar{Y}_{2n} + \frac{\sigma_{12}}{\sigma_{22}}\bar{Y}_{1n}$ where $\bar{Y}_{jn} = \frac{1}{n}\sum_{i=1}^n Y_{ij}$, $j = 1, 2$

$$-2\log\lambda(\underset{\sim}{x}) = -2\log\frac{L(\tilde{\phi})}{L(\hat{\phi})} = \sum_{i=1}^n (\underset{\sim}{Y}_i - \hat{\underset{\sim}{\phi}})^T B^T A^{-1} B (\underset{\sim}{Y}_i - \hat{\underset{\sim}{\phi}}) - \sum_{i=1}^n (\underset{\sim}{Y}_i - \tilde{\underset{\sim}{\phi}}) B^T A^{-1} B (\underset{\sim}{Y}_i - \tilde{\underset{\sim}{\phi}})$$

End of Material on Exam 2

Hypothesis Testing II

Bayes Tests

⇐ general test like LRT (and is based on likelihoods too)

Let $X_1, \ldots, X_n$ have joint pdf/pmf $f(\underset{\sim}{x}|\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$, and we want to test $H_0 : \theta \in \Theta_0 \subset \mathbb{R}^p$ vs $H_1 : \theta \notin \Theta_0$. Let

- $\pi(\theta)$ be a prior pdf $\Rightarrow$ posterior $f_{\theta|\underset{\sim}{x}}(\theta) \propto \pi(\theta) \underbrace{f(\underset{\sim}{x}|\theta)}_{\text{likelihood } L(\theta)}$
- $P(\theta \in \Theta_0|\underset{\sim}{x}) = \int_{\Theta_0} f_{\theta|\underset{\sim}{x}}(\theta)d\theta \quad \Leftarrow$ posterior prob that $\theta \in \Theta_0$
- $P(\theta \notin \Theta_0|\underset{\sim}{x}) = \int_{\Theta\setminus\Theta_0} f_{\theta|\underset{\sim}{x}}(\theta)d\theta \quad \Leftarrow$ posterior prob that $\theta \notin \Theta_0$
- Note that $P(\theta \in \Theta_0|\underset{\sim}{x}) + P(\theta \notin \Theta_0|\underset{\sim}{x}) = 1$

Then, *a Bayes test* for testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$ is given by

$$
\begin{aligned}
\varphi(\underset{\sim}{x}) &= \begin{cases} 1 & \text{if } \; P(\theta \notin \Theta_0|\underset{\sim}{x}) \geq P(\theta \in \Theta_0|\underset{\sim}{x}) \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 1 & \text{if } \; P(\theta \notin \Theta_0|\underset{\sim}{x}) \geq 1/2 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 1 & \text{if } \; P(\theta \in \Theta_0|\underset{\sim}{x}) < 1/2 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

⇐ reject H_0 if posterior prob. of $\theta \notin \Theta_0$ exceeds the posterior prob of $\theta \in \Theta_0$

need only compute 1 posterior prob.

Hypothesis Testing II

Bayes Test: Illustration

Example: Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$, $\theta \in \mathbb{R}$. Find the Bayes test for $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ under the $N(\mu, \tau^2)$ prior for θ, where μ, τ^2, θ_0 are fixed.

↑ fixed/given of θ_0 (eg $\theta_0 = 0$)

Solution: Check that the posterior distribution of θ ($f_{\theta|\underset{\sim}{x}}(\theta) \propto L(\theta)\pi(\theta)$) given $\underset{\sim}{x}$ is

$$N(\mu_{\theta|\underset{\sim}{x}}, \sigma^2_{\theta|\underset{\sim}{x}}), \text{ where}$$

$$\mu_{\theta|\underset{\sim}{x}} = \frac{n\tau^2 \bar{X}_n + \mu}{n\tau^2 + 1}, \quad \sigma^2_{\theta|\underset{\sim}{x}} = \frac{\tau^2}{n\tau^2 + 1}.$$

So, reject H_0 if $P(\theta \leq \theta_0 | \underset{\sim}{x}) < 1/2$

$$\iff P\Big(\underbrace{\frac{\theta - \mu_{\theta|\underset{\sim}{x}}}{\sigma_{\theta|\underset{\sim}{x}}}}_{Z \sim N(0,1)} \leq \frac{\theta_0 - \mu_{\theta|\underset{\sim}{x}}}{\sigma_{\theta|\underset{\sim}{x}}}\Big) < 1/2$$

$$\iff \Phi\Big(\frac{\theta_0 - \mu_{\theta|\underset{\sim}{x}}}{\sigma_{\theta|\underset{\sim}{x}}}\Big) < 1/2$$

where $\Phi(z) = P(Z \leq z)$ for $z \in \mathbb{R}$.

$$\iff \frac{\theta_0 - \mu_{\theta|\underset{\sim}{x}}}{\sigma_{\theta|\underset{\sim}{x}}} < 0$$

$$\iff \mu_{\theta|\underset{\sim}{x}} > \theta_0$$

Hence, the Bayes test is $\phi(\underset{\sim}{x}) = \begin{cases} 1 & \mu_{\theta|\underset{\sim}{x}} > \theta_0 \\ 0 & \text{o.w} \end{cases}$

ASIDE

Hypothesis Testing II

Bayes Tests: Discussion

The Bayes test here follows from minimizing the Bayes Risk BR_{φ_1} of a simple test $\varphi_1(\underset{\sim}{x})$ (a test where $\varphi_1(\underset{\sim}{x}) \in \{0,1\}$ for any $\underset{\sim}{x}$)

- Two possible actions depending on the data $\underset{\sim}{x}$: reject H_0 if $\varphi_1(\underset{\sim}{x}) = 1$ and don't reject H_0 if $\varphi_1(\underset{\sim}{x}) = 0$

- W.r.t. data $\underset{\sim}{x}$, the so-called "0-1" loss function is given by

$$L(\theta, \varphi_1(\underset{\sim}{x})) = I_{\{\theta \in \Theta_0\}} I_{\{\varphi_1(\underset{\sim}{x})=1\}} + I_{\{\theta \notin \Theta_0\}} I_{\{\varphi_1(\underset{\sim}{x})=0\}}.$$

  That is, the loss $L(\theta, \varphi_1(\underset{\sim}{x})) = 0$ for a correct decision and $L(\theta, \varphi_1(\underset{\sim}{x})) = 1$ for an incorrect decision:

$$L(\theta, \varphi_1(\underset{\sim}{x})) = \begin{cases} 0 & \text{if } \theta \in \Theta_0 \ \& \ \varphi_1(\underset{\sim}{x}) = 0 \text{ or if } \theta \notin \Theta_0 \ \& \ \varphi_1(\underset{\sim}{x}) = 1 \\ 1 & \text{otherwise} \end{cases}$$

- We can find the Bayes test by minimizing the posterior risk of a simple test $\varphi_1(\underset{\sim}{x})$ for each fixed $\underset{\sim}{x}$, where the posterior risk is

$$\begin{aligned} \mathrm{E}_{\theta|\underset{\sim}{x}} L(\varphi_1(\underset{\sim}{x}), \theta) &= \int_\Theta \left(I_{\{\theta \in \Theta_0\}} I_{\{\varphi_1(\underset{\sim}{x})=1\}} + I_{\{\theta \notin \Theta_0\}} I_{\{\varphi_1(\underset{\sim}{x})=0\}} \right) f_{\theta|\underset{\sim}{x}}(\theta) d\theta \\ &= I_{\{\varphi_1(\underset{\sim}{x})=1\}} \int_{\Theta_0} f_{\theta|\underset{\sim}{x}}(\theta) d\theta + I_{\{\varphi_1(\underset{\sim}{x})=0\}} \int_{\Theta \setminus \Theta_0} f_{\theta|\underset{\sim}{x}}(\theta) d\theta \\ &= I_{\{\varphi_1(\underset{\sim}{x})=1\}} P(\theta \in \Theta_0 | \underset{\sim}{x}) + I_{\{\varphi_1(\underset{\sim}{x})=0\}} P(\theta \notin \Theta_0 | \underset{\sim}{x}) \end{aligned}$$

  you get to minimize this by choosing action $\varphi_1(\underset{\sim}{x})=1$ or $\varphi_1(\underset{\sim}{x})=0$

  For each fixed $\underset{\sim}{x}$, we choose the values $\varphi_1(\underset{\sim}{x}) = 1$ or 0 of the test to minimize the posterior risk; that is, for each fixed $\underset{\sim}{x}$, we should pick $\varphi_1(\underset{\sim}{x}) = 1$ if $P(\theta \notin \Theta_0 | \underset{\sim}{x}) \geq P(\theta \in \Theta_0 | \underset{\sim}{x})$ and pick $\varphi_1(\underset{\sim}{x}) = 0$ if $P(\theta \notin \Theta_0 | \underset{\sim}{x}) < P(\theta \in \Theta_0 | \underset{\sim}{x})$. Note this is the same decision rule as the Bayes test $\varphi(\underset{\sim}{x})$ above.