STAT 4100 Lecture Note

Week Seven (Oct 17, 19, & 21, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Oct/15 22:12:38

Hypothesis Testing

Binary classification

- Assume $\mathbf{X} = [X_1, \ldots, X_n]^\top \sim f(\boldsymbol{x} \mid \boldsymbol{\theta}^*) \in \{f(\boldsymbol{x} \mid \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$
  - Fixed unknown $\boldsymbol{\theta}^*$ to be inferred
- Make a decision on $\boldsymbol{\theta}^*$ between two hypotheses $H_0 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$ and $H_1 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$
  - $\boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$
  - $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \emptyset$

- Decision and correctness
  - True positive (TP) = H_0 correctly rejected
  - False positive (FP, type I error) = H_0 incorrectly rejected
  - True negative (TN) = H_0 is correctly accepted
  - False negative (FN, type II error) = H_0 incorrectly accepted
- E.g., H_0 : healthy vs H_1 : sick
  - TP: sick people identified as sick
  - FP: healthy people identified as sick
  - TN: healthy people identified as healthy
  - FN: sick people identified as healthy

	Accept H_0	Reject H_0
H_0 is true	True negative (TN)	False positive (FP, type I error)
H_0 is false	False negative (FN, type II error)	True positive (TP)

- Misclassification rate = Pr(FP) + Pr(FN)
- False discovery rate (FDR) = Pr(FP)/{Pr(FP) + Pr(TP)}
  - FDR controlling for sequential/simultaneous testing
- Receiver operating characteristic curve (ROC curve): plot of TPR vs FPR
  - True positive rate (TPR, sensitivity) = Pr(TP)/{Pr(TP) + Pr(FN)}
  - False positive rate (FPR) = Pr(FP)/{Pr(FP) + Pr(FN)}

– Area under the ROC curve (AUC)

- True negative rate (TNR, specificity) = Pr(TN)/{Pr(TN) + Pr(FP)}
- The optimal hypothesis testing seeking to minimize Pr(FN) subject to capped Pr(FP), i.e.,

$$\min \Pr(\text{type II error}) \text{ subject to } \Pr(\text{type I error}) \leq \alpha$$

Power function

- Rejection/critical region: $R = \{\boldsymbol{x} : \text{data } \boldsymbol{x} \text{ corresponding to the rejection of } H_0\}$
  - Typically specified in terms of a function of the sample (called the *test statistic*); e.g., if $R = \{\boldsymbol{x} : \bar{x} \geq 3\}$, then $\bar{X}$ is the test statistic.
- Test function $\phi : \text{supp}(\mathbf{X}) \to \{0, 1\}$ defined as $\phi(\boldsymbol{x}) = \mathbf{1}_R(\boldsymbol{x})$
  - $\phi(\boldsymbol{x}) = 1$ implying the rejection of H_0
- Each test function ϕ corresponds to a unique rejection region $R_\phi = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}$
  - Two tests considered to be equivalent if they correpond to the same rejection region/test function
- Power function (for ϕ): $\beta_\phi(\boldsymbol{\theta}) = \Pr(\mathbf{X} \in R_\phi \mid \boldsymbol{\theta}) = \mathrm{E}\{\phi(\mathbf{X}) \mid \boldsymbol{\theta}\}$
  - Pr(type I error) $= \beta_\phi(\boldsymbol{\theta}^*)$ if H_0 is correct ($\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$)
  - Pr(type II error) $= 1 - \beta_\phi(\boldsymbol{\theta}^*)$ if H_1 is correct ($\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$)
- Prefer larger $\beta_\phi(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ and smaller $\beta_\phi(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ (because $\boldsymbol{\theta}^*$ is unknown)

Example Lec14.2

- iid $X_1, \ldots, X_n \sim N(\theta, \sigma_0^2)$ with known σ_0. Consider a test for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ with rejection region $\{\boldsymbol{x} : \sqrt{n}|\bar{\boldsymbol{x}} - \theta_0|/\sigma_0 > c\}$.
  a. Elaborate the power function.
  b. Find sample size n and threshold c if one desires that the type I error rate is 5% and the maximal type II error rate is 25% and attained at $\theta = \theta_0 + \sigma_0$.

Uniformly most powerful (UMP) level α test (CB Sec 8.3.2)

- ϕ is of level α iff $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta}) \leq \alpha$
  - ϕ is of size α iff $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta}) = \alpha$
- Let ϕ is a level α test for $H_0 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$ vs $H_1 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$. If $\beta_\phi(\boldsymbol{\theta}) \geq \beta_{\phi'}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ and all ϕ' of level α, then ϕ is a UMP level α test.
- If ϕ is a UMP level α test, then $\beta_\phi(\boldsymbol{\theta}) \geq \alpha \geq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ (unbiasedness for testing, CB Def 8.3.9)

UMP level α test for simple hypotheses ($H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ vs $H_1 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_1$)

- To maximize $\beta_\phi(\boldsymbol{\theta}_1)$ with respect to ϕ subject to $\beta_\phi(\boldsymbol{\theta}_0) \leq \alpha$
- Neymann-Pearson (NP) Lemma (CB Thm 8.3.12): ϕ is the UMP test of level α for simple hypotheses $\iff \exists k > 0$ such that $\beta_\phi(\boldsymbol{\theta}_0) = \mathrm{E}\{\phi(\mathbf{X}) \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} = \alpha$, where

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & f(\boldsymbol{x} \mid \boldsymbol{\theta}_1) > k f(\boldsymbol{x} \mid \boldsymbol{\theta}_0), \\ 0, & f(\boldsymbol{x} \mid \boldsymbol{\theta}_1) < k f(\boldsymbol{x} \mid \boldsymbol{\theta}_0). \end{cases}$$

  In practice (especially for discrete distributions), k is the largest real number such that

$$\Pr\{f(\mathbf{X} \mid \boldsymbol{\theta}_1)/f(\mathbf{X} \mid \boldsymbol{\theta}_0) \geq k \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} \geq \alpha$$

  and

$$\Pr\{f(\mathbf{X} \mid \boldsymbol{\theta}_1)/f(\mathbf{X} \mid \boldsymbol{\theta}_0) \leq k \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} \geq 1 - \alpha.$$

- What shall we do if $\Pr\{f(\mathbf{X} \mid \boldsymbol{\theta}_1) = kf(\mathbf{X} \mid \boldsymbol{\theta}_0) \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} \neq 0$? Take a randomized test! Specifically, if $f_{\theta_1}(\mathbf{X})/f_{\theta_0}(\mathbf{X}) = k$, let $\phi(\boldsymbol{x}) = \gamma \in [0, 1]$ such that

$$\Pr\{f(\mathbf{X} \mid \boldsymbol{\theta}_1)/f(\mathbf{X} \mid \boldsymbol{\theta}_0) > k \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} + \gamma \Pr\{f(\mathbf{X} \mid \boldsymbol{\theta}_1)/f(\mathbf{X} \mid \boldsymbol{\theta}_0) = k \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\} = \alpha.$$

  That is, we reject H_0 with probability γ.

- For simple hypotheses, UMP test at level $\alpha \iff$ UMP test at size α.

- UMP test and sufficiency (CB Coro 8.3.13): sufficient statistics can be taken as test statistics for UMP ϕ.

UMP level α test for one-sided hypotheses ($H_0 : \theta^* \leq \theta_0$ (or $\theta^* = \theta_0$) vs $H_1 : \theta^* > \theta_0$)

- Consider cases with only one unknown parameter
- Monotone likelihood ratio (MLR, CB Def 8.3.16): fixing $\theta_1 > \theta_2$, $g(t \mid \theta_1)/g(t \mid \theta_2)$ is monotonic with respect to t for $\{g(t \mid \theta) : \theta \in \Theta \subset \mathbb{R}\}$
  - E.g., one-parameter exponential family bears MLR

- Karlin-Rubin (CB Thm 8.3.17): Suppose T is sufficient for θ and $T \sim g(t \mid \theta)$ bearing MLR. A UMP level α test for $H_0 : \theta^* \leq \theta_0$ (or $\theta^* = \theta_0$) vs $H_1 : \theta^* > \theta_0$ is

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & T(\boldsymbol{x}) > t_\alpha, \\ 0, & T(\boldsymbol{x}) < t_\alpha, \end{cases}$$

  where t_α is a real number such that $\beta_\phi(\theta_0) = \mathrm{E}\{\phi(\mathbf{X}) \mid \theta = \theta_0\} = \Pr(T > t_\alpha \mid \theta = \theta_0) = \alpha$.

Example Lec14.1

- iid $X_1, \ldots, X_n \sim N(\mu, 1)$. Construct UMP level α test for following hypotheses.
  a. $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$ with $\mu_0 < \mu_1$;
  b. $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$;
  c. $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$;
  d. $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.