

PH 712 Probability and Statistical Inference

Recap for Final

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/Dec/11 12:32:15

Hypothesis testing (Part IX)

- Hypotheses
 - $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$.
 - * $\Theta = \Theta_0 \cup \Theta_1$
 - * $\emptyset = \Theta_0 \cap \Theta_1$
 - * Θ_0 must be a closed set, i.e., include “=” in H_0
- Rejection region R : reject H_0 if $\{x_1, \dots, x_n\} \in R$
- Power function: $\beta(\theta) = \Pr(\{X_1, \dots, X_n\} \in R \mid \theta)$, $\theta \in \Theta$
 - $\Pr(\text{type I error}) = \beta_\phi(\theta)$ if H_0 is correct (i.e., $\theta \in \Theta_0$)
 - $\Pr(\text{type II error}) = 1 - \beta_\phi(\theta)$ if H_1 is correct (i.e., $\theta \in \Theta_1$)
- Level α : $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$
 - Size α : $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$
- Uniformly most powerful (UMP) level α test \iff its power function \geq power functions of any level α tests at each $\theta \in \Theta_1$
- Likelihood ratio test (LRT, equivalent to the UMP test when the UMP test exists)
 - Test statistic

$$\lambda(X_1, \dots, X_n) = \frac{L(\hat{\theta}_{\text{ML},0})}{L(\hat{\theta}_{\text{ML}})}$$

- * $\hat{\theta}_{\text{ML}}$: MLE with constraint $\theta \in \Theta$
- * $\hat{\theta}_{\text{ML},0}$: MLE with constraint $\theta \in \Theta_0$
- Level α LRT rejection region: $\{\{x_1, \dots, x_n\} : \lambda(x_1, \dots, x_n) \leq c_\alpha\}$ where c_α is the solution to

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} \Pr\{\lambda(X_1, \dots, X_n) \leq c_\alpha \mid \theta\} = \alpha$$

- Special cases: Z -test, t -test and F -test
- Asymptotic LRT
 - More feasible than LRT
 - Test statistic
$$-2 \ln \lambda(X_1, \dots, X_n) = -2\{\ell(\hat{\theta}_{\text{ML},0}) - \ell(\hat{\theta}_{\text{ML}})\}$$
 - Level α asymptotic LRT rejection region: $\{\{x_1, \dots, x_n\} : -2 \ln \lambda(x_1, \dots, x_n) \geq \chi_{\nu, 1-\alpha}^2\}$
 - * ν : the difference of numbers of free parameters between Θ_0 and Θ

- Wald test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
 - Test statistic $(\hat{\theta}_{\text{ML}} - \theta_0) / \sqrt{\widehat{\text{var}}(\hat{\theta}_{\text{ML}})}$
 - * Obtain $\widehat{\text{var}}(\hat{\theta}_{\text{ML}})$ via the Fisher information and/or delta methods
 - Level α Wald rejection region: $\{\{x_1, \dots, x_n\} : |\hat{\theta}_{\text{ML}} - \theta_0| / \sqrt{\widehat{\text{var}}(\hat{\theta}_{\text{ML}})} \geq \Phi_{1-\alpha/2}^{-1}\}$
- Score test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
 - Test statistic: $\ell'(\theta_0) / \sqrt{I_n(\theta_0)}$
 - Level α score rejection region: $\{\{x_1, \dots, x_n\} : |\ell'(\theta_0)| / \sqrt{I_n(\theta_0)} \geq \Phi_{1-\alpha/2}^{-1}\}$
- p -value
 - A test statistic facilitating the report of testing result because its critical point is by default α , i.e., the rejection region in terms of $p(x_1, \dots, x_n)$ is always

$$\{\{x_1, \dots, x_n\} : p(x_1, \dots, x_n) \leq \alpha\}$$

- Not always well-defined
- Special cases
 - * Asymptotic LRT: $p(x_1, \dots, x_n) = 1 - F_{\chi^2(\nu)}(-2\lambda(x_1, \dots, x_n))$
 - $F_{\chi^2(\nu)}(\cdot)$: the cdf of $\chi^2(\nu)$
 - * Wald test: $p(x_1, \dots, x_n) = 2\Phi\left(-|\hat{\theta}_{\text{ML}} - \theta_0| / \sqrt{\widehat{\text{var}}(\hat{\theta}_{\text{ML}})}\right)$
 - * Score test: $p(x_1, \dots, x_n) = 2\Phi\left(-|\ell'(\theta_0)| / \sqrt{I_n(\theta_0)}\right)$

$(1 - \alpha) \times 100\%$ **confidence set of θ (Part X)**

- A set covering true θ with probability AT LEAST $(1 - \alpha) \times 100\%$
- Constructed by inverting a level α rejection region for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
- The narrower the better
- Special cases:
 - $(1 - \alpha) \times 100\%$ asymptotic LRT confidence set for θ : $\{\theta : -2(\ell(\theta) - \ell(\hat{\theta}_{\text{ML}})) < \chi_{1,1-\alpha}^2\}$
 - $(1 - \alpha) \times 100\%$ Wald confidence set for θ : $\{\theta : |\hat{\theta}_{\text{ML}} - \theta| / \sqrt{\widehat{\text{var}}(\hat{\theta}_{\text{ML}})} < \Phi_{1-\alpha/2}^{-1}\}$
 - * Why is it preferable to use $\hat{\theta}_{\text{ML}}$ rather than other estimators? Hint: Bridge the width of Wald confidence set to the efficiency (Part VI) and asymptotic efficiency (Part VIII).
 - $(1 - \alpha) \times 100\%$ score confidence set for θ : $\{\theta : |\ell'(\theta)| / \sqrt{I_n(\theta)} < \Phi_{1-\alpha/2}^{-1}\}$

Techniques involved in inference

- MLE (Part V)
 - Its consistency and asymptotic efficiency (Part VIII)
- Approximating the variance of $h(\hat{\theta}_{\text{ML}})$ (Part VIII)
 - Fisher information
 - Delta methods