STAT 4100 Lecture Note

Week Thirteen (Dec 5, 7 & 9, 2022)

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

2022/Dec/06 00:53:05

Large-sample hypothesis testing (con'd)

Wald test (CB pp. 493)

- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
  - Wald statistic: $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)}$ (if $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)} \xrightarrow{d} \mathcal{N}(0,1)$ under H_0 as $n \to \infty$)
    - Asymptotically equivalent to LRT for this two sided test if $\hat{\theta}_n = \hat{\theta}_{\mathrm{ML}}$
    - Substitute $\widehat{\mathrm{var}}(\hat{\theta}_n)$ for $\mathrm{var}(\hat{\theta}_n)$ if $\mathrm{var}(\hat{\theta}_n)$ is well approximated by $\widehat{\mathrm{var}}(\hat{\theta}_n)$ (obtained by the delta methods/bootstrap)
  - Level α Wald rejection region: $\{\boldsymbol{x} : |\hat{\theta}_n - \theta_0|/\sqrt{\mathrm{var}(\hat{\theta}_n)} \geq \Phi^{-1}_{1-\alpha/2}\}$
  - p-value $= 2\Phi\left(-|\hat{\theta}_n - \theta_0|/\sqrt{\mathrm{var}(\hat{\theta}_n)}\right)$

Score test (CB pp. 494)

- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
  - Score statistic: $S(\theta_0; \mathbf{X})/\sqrt{I_n(\theta_0)}$ ($\xrightarrow{d} \mathcal{N}(0,1)$ under H_0 as $n \to \infty$)
  - Level α score rejection region: $\{\boldsymbol{x} : |S(\theta_0; \boldsymbol{x})|/\sqrt{I_n(\theta_0)} \geq \Phi^{-1}_{1-\alpha/2}\}$.
  - p-value $= 2\Phi\left(-|S(\theta_0; \boldsymbol{x})|/\sqrt{I_n(\theta_0)}\right)$

CB Examples 10.3.5 & 10.3.6

- iid $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$, $p \in (0,1)$. Derive LRT, Wald and score tests for $H_0 : p = p_0$ versus $H_1 : p \neq p_0$.

Asymptotic confidence set

- Inverting rejection regions, e.g.,
  - $1-\alpha$ LRT confidence set for θ: $\{\theta : -2\ln\{L(\theta; \boldsymbol{x})/L(\hat{\theta}_{\mathrm{ML}}; \boldsymbol{x})\} < \chi^2_{1,1-\alpha}\}$
  - $1-\alpha$ Wald confidence set for θ: $\{\theta : |\hat{\theta}_n - \theta|/\sqrt{\mathrm{var}(\hat{\theta}_n)} < \Phi^{-1}_{1-\alpha/2}\}$
  - $1-\alpha$ score confidence set for θ: $\{\theta : |S(\theta; \boldsymbol{x})|/\sqrt{I_n(\theta)} < \Phi^{-1}_{1-\alpha/2}\}$
- Bootstrap
  1. For j in $1 : B$, do steps 2–3.
  2. Draw the jth resample $\boldsymbol{x}^*_j$ of size n from the empirical CDF (nonparametric bootstrap) OR a fitted parametric model (parametric bootstrap).
  3. Let $\hat{\theta}^*_j = \hat{\theta}(\boldsymbol{x}^*_j)$.

4. $1-\alpha$ bootstrap confidence interval for θ is $(q_{\alpha/2}, q_{1-\alpha/2})$, where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are $\alpha/2$ and $1-\alpha/2$ sample quantiles of $\{\hat\theta_1^*, \ldots, \hat\theta_B^*\}$, respectively.

- Depending on probabilistic inequalities, e.g.,
  - Constructing a $1-\alpha$ confidence set of μ by finding the smallest c such that $\Pr(|\bar X_n - \mu| \geq c) \leq \alpha$ through the Chebyshev's inequality

CB Examples 10.4.2, 10.4.3 & 10.4.5

- iid $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$, construct $1-\alpha$ confidence set for p.

```
options(digits = 4)
set.seed(1)
B = 1e4L
n = 1e3L
alpha = .05
x = rbinom(n, 1, prob = .6)
theta_ml = mean(x)
theta_star_np = numeric(B)
theta_star_p = numeric(B)
# Nonparametric bootstrap
for (j in 1:B){
  x_star = sample(x, size = n, replace = T)
  theta_star_np[j] = mean(x_star)
}
quantile(theta_star_np, probs = c(alpha/2, 1-alpha/2))
# Parametric bootstrap
for (j in 1:B){
  x_star = rbinom(n, size = 1, prob = theta_ml)
  theta_star_p[j] = mean(x_star)
}
quantile(theta_star_p, probs = c(alpha/2, 1-alpha/2))
```

Recap for final

Statistical model

- Characterizing distributions
  - cdf/pdf/pmf
  - mgf
    - Existence: if $\mathrm{E}\{\exp(tX)\} < \infty$ for all t inside a neighbourhood of 0
    - $M_Y(t) = \exp(bt)\prod_i M_{X_i}(a_i t)$ if $Y = b + \sum_i a_i X_i$, where b and a_i are constants, $X_1, \ldots, X_p$ are independent, and each $M_{X_i}(\cdot)$ exists
- Exponential family
  - The pdf/pmf is of the following form

$$f(x \mid \boldsymbol\theta) = h(x)c(\boldsymbol\theta)\exp\left\{\sum_{i=1}^k w_i(\boldsymbol\theta)t_i(x)\right\}$$

  - Special cases: normal, binomial, gamma, beta, Poisson, negative binomial
- Variable transformation
- Normal sampling theory

- $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$
- $X/\sqrt{Y/n} \sim t(n)$ if $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(n)$ are independent
- $(X/m)/(Y/n) \sim F(m,n)$ if $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ are independent
- $n^{1/2}(\bar{X} - \mu)/\sigma \sim \mathcal{N}(0,1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $\bar{X}$ and S^2 are independent of each other if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
- $n^{1/2}(\bar{X} - \mu)/S \sim t(n-1)$ if iid $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

- Checking independence
  - Separable joint cdf: $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
  - Separable joint pdf or pmf: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - conditional pdf or pmf: $f_{X|Y}(x \mid y) = f_X(x)$
  - Separable mgf: $\mathrm{E}(e^{t_1 X + t_2 Y}) = \mathrm{E}(e^{t_1 X})\mathrm{E}(e^{t_2 Y})$
  - Basu's theorem
    - Sometimes it is even more complex to find complete statisitcs than to obtain the joint pdf
  - Zero $\mathrm{cov}(X,Y)$ for joint normal (X,Y)

- Convergence of random variables
  - Definitions
    - Convergence in probability $X_n \xrightarrow{p} X \iff \forall \epsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0$
      - To be verified through the Markov's/Chebyshev's inequality
    - Almost sure convergence $X_n \xrightarrow{\text{a.s.}} X \iff \forall \epsilon > 0$, $\Pr(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$
    - Convergence in distribution $X_n \xrightarrow{d} X \iff \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for each x with $\Pr(X = x) = 0$
      - Resort to CLT/delta methods if the limiting distribution is normal
  - (CMT) $\tau(\cdot)$ is continuous and $X_n \xrightarrow{\text{a.s.}/p/d} X \Rightarrow \tau(X_n) \xrightarrow{\text{a.s.}/p/d} \tau(X)$
  - The chain of implications
$$\xrightarrow{\text{a.s.}} \quad \Rightarrow \quad \xrightarrow{p} \quad \Rightarrow \quad \xrightarrow{d}$$
    (The inverse is typically incorrect but $X_n \xrightarrow{d}$ constant $c \Rightarrow X_n \xrightarrow{p} c$.)
  - $X_n \xrightarrow{\text{a.s.}/p} X$ and $Y_n \xrightarrow{\text{a.s.}/p} Y \Rightarrow$
    - $aX_n + bY_n \xrightarrow{\text{a.s.}/p} aX + bY$
    - $X_n Y_n \xrightarrow{\text{a.s.}/p} XY$
  - (Slutsky's theorem) $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d}$ constant $c \Rightarrow$
    - $aX_n + bY_n \xrightarrow{d} aX + bc$
    - $X_n Y_n \xrightarrow{d} cX$
  - (LLN) $X_1, \ldots, X_n$ are iid with finite mean $\mu \Rightarrow \bar{X}_n \xrightarrow{p/\text{a.s.}} \mu$
  - (CLT) $X_1, \ldots, X_n$ are iid with finite mean μ and finite variance $\sigma^2 \Rightarrow \sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$

Point estimation

- MOM estimators
  - Equate raw moments to their empirical counterparts
  - Not unique but an acceptable starting point for iterative algorithms
- MLE
  - $\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}; \boldsymbol{x}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}; \boldsymbol{x})$
  - Maximizing $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
    - For discrete $\boldsymbol{\Theta}$: compare $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ over all the possible values of $\boldsymbol{\theta}$
    - For continuous $\boldsymbol{\Theta}$:
      - If $\mathbf{S}(\boldsymbol{\theta})$ has no zero point (i.e., stationary points): utilize the monotonicity of $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$

        - If $\mathbf{S}(\boldsymbol{\theta})$ has zero point: solve $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$ (to obtain stationary points) and then compare $L(\boldsymbol{\theta}; \boldsymbol{x})$ or $\ell(\boldsymbol{\theta}; \boldsymbol{x})$ over all the stationary points and boundary points
  - Properties
    - Invariance: $\widehat{g(\boldsymbol{\theta})}_{\text{ML}} = g(\hat{\boldsymbol{\theta}}_{\text{ML}})$
    - Consistency: $\tau(\hat{\theta}_{\text{ML}}) \xrightarrow{p} \tau(\theta)$
    - Asymptotic distribution (by delta methods)
      - $\tau'(\theta) \neq 0 \Rightarrow \sqrt{n}\{\tau(\hat{\theta}_{\text{ML}}) - \tau(\theta)\} \xrightarrow{d} N(0, \{\tau'(\theta)\}^2/I_1(\theta))$.
      - $\tau'(\theta) = 0$ and $\tau''(\theta) \neq 0 \Rightarrow n\{\tau(\hat{\theta}_{\text{ML}}) - \tau(\theta)\} \xrightarrow{d} [\tau''(\theta)/\{2I_1(\theta)\}]\chi^2(1)$.
- Evaluating estimators (univariate)
  - UMVUE/MVUE/Best unbiased estimator
    - Minimizing the MSE subject to unbiasedness
    - If T is unbiased for $\tau(\theta)$ and attains the CRLB (i.e., $\text{E}(T) = \tau(\theta)$ and $\text{var}(T) = \{\tau'(\theta)\}^2/I_n(\theta)$), then T is the UMVUE for $\tau(\theta)$. (The inverse is NOT correct!)
      - (Expected) Fisher information (number) for iid sample of size n

$$I_n(\theta) = \text{var}\{S(\theta; \mathbf{X}) \mid \theta\} = \text{E}[\{S(\theta; \mathbf{X}) \mid \theta\}^2] = -\text{E}\{H(\theta; \mathbf{X}) \mid \theta\}$$

    - (Lehmann-Scheffe) debias or Rao-Blackwellize a function of sufficient complete statistics, starting with, e.g.,
      - $\sum_{i=1}^n t(X_i)$ (sufficient and complete for an exponential family)
      - MLE (often a function of sufficient complete statistics)
  - Consistency: $\hat{\theta}_n \xrightarrow{p} \theta$
  - Asymptotic efficiency: $\sqrt{n}\{\tau(\hat{\theta}_n) - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \{\tau'(\theta)\}^2/I_1(\theta))$
  - ARE of T_n with respect to W_n, say $\text{ARE}(T_n, W_n) = \sigma_W^2(\theta)/\sigma_T^2(\theta)$, if
    - $\sqrt{n}\{T_n - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_T^2(\theta))$ and $\sqrt{n}\{W_n - \tau(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_W^2(\theta))$

Hypothesis testing

- $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ vs. $H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$.
  - $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$
  - $\emptyset = \boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1$
- Characterization
  - Rejection region R: H_0 is rejected once $\boldsymbol{x} \in R$
  - Test function $\phi = \phi(\boldsymbol{x}) = \mathbf{1}_R(\boldsymbol{x})$, $\boldsymbol{x} \in \text{supp}(\mathbf{X})$
- Power function of ϕ: $\beta_\phi(\boldsymbol{\theta}) = \Pr(\mathbf{X} \in R_\phi \mid \boldsymbol{\theta}) = \text{E}\{\phi(\mathbf{X}) \mid \boldsymbol{\theta}\}$
  - $\Pr(\text{type I error}) = \beta_\phi(\boldsymbol{\theta}^*)$ if H_0 is correct
  - $\Pr(\text{type II error}) = 1 - \beta_\phi(\boldsymbol{\theta}^*)$ if H_1 is correct
  - Size α: $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta}) = \alpha$
  - Level α: $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta}) \leq \alpha$
- UMP level α test
  - ϕ is the UMP level α test $\iff$ $\beta_\phi(\boldsymbol{\theta}) \geq \beta_{\phi'}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ and for each ϕ' of level α
  - (NP Lemma) for simple hypotheses ($H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$),

$$\phi_c(\boldsymbol{x}) = \begin{cases} 1, & f(\boldsymbol{x} \mid \boldsymbol{\theta}_1)/f(\boldsymbol{x} \mid \boldsymbol{\theta}_0) > c, \\ 0, & f(\boldsymbol{x} \mid \boldsymbol{\theta}_1)/f(\boldsymbol{x} \mid \boldsymbol{\theta}_0) < c \end{cases}$$

    is the UMP test of level α, where $c > 0$ is determined so that $\beta_\phi(\theta_0) = \text{E}\{\phi_c(\mathbf{X}) \mid \theta = \theta_0\} = \alpha$
  - (Karlin-Rubin theorem)
    - Prerequisite
      - T sufficient for θ
      - $T \sim f_T(t \mid \theta)$ bearing the MLR, i.e., fixing $\theta_2 > \theta_1$, $f_T(t \mid \theta_2)/f_T(t \mid \theta_1)$ is nondecreasing with respect to t

  - ∗ for $H_0 : \theta = \theta_0$ OR $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_1$,

$$\phi_c(\boldsymbol{x}) = \begin{cases} 1, & T(\boldsymbol{x}) > c, \\ 0, & T(\boldsymbol{x}) < c \end{cases}$$

    is the UMP test of level α, where c satisfies that $\Pr\{T(\mathbf{X}) > c \mid \theta = \theta_0\} = \alpha$
  - ∗ for $H_0 : \theta = \theta_0$ OR $H_0 : \theta \ge \theta_0$ vs. $H_1 : \theta < \theta_1$,

$$\phi_c(\boldsymbol{x}) = \begin{cases} 1, & T(\boldsymbol{x}) < c, \\ 0, & T(\boldsymbol{x}) > c \end{cases}$$

    is the UMP test of level α, where c satisfies that $\Pr\{T(\mathbf{X}) < c \mid \theta = \theta_0\} = \alpha$
  - UMP test at level $\alpha \iff$ UMP test at size α
- LRT (equivalent to the UMP test when the UMP test exists)
  - Test statistic

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} L(\boldsymbol{\theta} \mid \boldsymbol{x})}{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta} \mid \boldsymbol{x})} = \frac{L(\hat{\boldsymbol{\theta}}_{0,\mathrm{ML}} \mid \boldsymbol{x})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}} \mid \boldsymbol{x})}$$

  - Level α rejection region: $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c\}$ where

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \beta_\phi(\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \Pr\{\lambda(\mathbf{X}) \le c \mid \boldsymbol{\theta}\} = \alpha$$

  - Asymptotic level α rejection region: $R \approx \{\boldsymbol{x} : -2\ln\lambda(\boldsymbol{x}) \ge \chi^2_{\nu,1-\alpha}\}$ since $-2\ln\lambda(\mathbf{X}) \xrightarrow{d} \chi^2(\nu)$ under H_0
- Wald test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \ne \theta_0$
  - Test statistic $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)}$ (if $(\hat{\theta}_n - \theta_0)/\sqrt{\mathrm{var}(\hat{\theta}_n)} \xrightarrow{d} \mathcal{N}(0,1)$ under H_0 as $n \to \infty$)
    - ∗ Substitute $\widehat{\mathrm{var}}(\hat{\theta}_n)$ for $\mathrm{var}(\hat{\theta}_n)$ if $\mathrm{var}(\hat{\theta}_n)$ is well approximated by $\widehat{\mathrm{var}}(\hat{\theta}_n)$ (obtained by the delta methods/bootstrap)
  - Level α Wald rejection region: $\{\boldsymbol{x} : |\hat{\theta}_n - \theta_0|/\sqrt{\mathrm{var}(\hat{\theta}_n)} \ge \Phi^{-1}_{1-\alpha/2}\}$
  - Asymptotically equivalent to LRT for this two sided test if $\hat{\theta}_n = \hat{\theta}_{\mathrm{ML}}$
- Score test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \ne \theta_0$
  - Test statistic: $S(\theta_0; \mathbf{X})/\sqrt{I_n(\theta_0)}$ ($\xrightarrow{d} \mathcal{N}(0,1)$ under H_0 as $n \to \infty$)
  - Level α score rejection region: $\{\boldsymbol{x} : |S(\theta_0; \boldsymbol{x})|/\sqrt{I_n(\theta_0)} \ge \Phi^{-1}_{1-\alpha/2}\}$
- p-value
  - $p(\mathbf{X})$ is valid (to be taken as a test statistic) $\iff$ $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \Pr\{p(\mathbf{X}) \le \alpha \mid \boldsymbol{\theta}\} \le \alpha$ for each $\alpha \in [0,1]$.
    - ∗ i.e., it is possible to define "level" and "size" if we take $p(\mathbf{X})$ as a test statistic.
    - ∗ Level α rejection region (depending on $p(\boldsymbol{x})$): $R = \{\boldsymbol{x} : p(\boldsymbol{x}) \le \alpha\}$.
  - Specifically, if the rejection region is of the form that $R = \{\boldsymbol{x} : T(\boldsymbol{x}) \ge c\}$, then $p(\boldsymbol{x}) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \Pr\{T(\mathbf{X}) \ge T(\boldsymbol{x}) \mid \boldsymbol{\theta}\}$

$1 - \alpha$ confidence set of θ

- Inverting a level α rejection region for two-sided hypotheses
- Depending on probabilistic inequalities, e.g., the Markov's/Chebyshev's inequality
- Bootstrap