STAT 3690 Lecture 09

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Feb 11, 2022

Sampling distributions of $\bar{\mathbf{X}}$ and $\mathbf{S}$ (J&W Sec 4.4)

- Recall the univariate case
  - $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$
  - $S^2 \perp\!\!\!\perp \bar{X}$
    - Sample variance $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$
  - $\sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1)$
  - $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
  - $\sqrt{n}(\bar{X} - \mu)/S \sim t(n-1)$
- The multivariate case
  - $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
  - $\mathbf{S} \perp\!\!\!\perp \bar{\mathbf{X}}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \perp\!\!\!\perp \hat{\boldsymbol{\mu}}_{\text{ML}}$
  - $\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim MVN_p(\mathbf{0}, \mathbf{I})$
  - $(n-1)\mathbf{S} = n\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \sim W_p(n-1, \boldsymbol{\Sigma})$
  - $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim$ Hotelling's $T^2(p, n-1)$

- Wishart distribution
  - Def: $W_p(\boldsymbol{\Sigma}, n)$ is the distribution of $\sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top$ with $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{\text{iid}}{\sim} MVN_p(\mathbf{0}, \boldsymbol{\Sigma})$
    - A generalization of χ^2-distribution: $W_p(\boldsymbol{\Sigma}, n) = \chi^2(n)$ if $p = \boldsymbol{\Sigma} = 1$
  - Propoties
    - $\mathbf{A}\mathbf{A}^\top > 0$ and $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow \mathbf{A}\mathbf{W}\mathbf{A}^\top \sim W_p(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, n)$
    - $\mathbf{W}_i \overset{\text{iid}}{\sim} W_p(\boldsymbol{\Sigma}, n_i) \Rightarrow \mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n_1 + n_2)$
    - $\mathbf{W}_1 \perp\!\!\!\perp \mathbf{W}_2$, $\mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n)$ and $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1) \Rightarrow \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n - n_1)$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$ and $\boldsymbol{a} \in \mathbb{R}^p \Rightarrow$
$$\frac{\boldsymbol{a}^\top \mathbf{W} \boldsymbol{a}}{\boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}} \sim \chi^2(n)$$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$, $\boldsymbol{a} \in \mathbb{R}^p$ and $n \geq p \Rightarrow$
$$\frac{\boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{W}^{-1} \boldsymbol{a}} \sim \chi^2(n - p + 1)$$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow$
$$\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{W}) \sim \chi^2(np)$$

- Hotelling's T^2 distribution
  - A generalization of (Student's) t-distribution
  - If $\mathbf{X} \sim MVN_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W} \sim W_p(\mathbf{I}, n)$, then
$$\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \sim T^2(p, n)$$

- $Y \sim T^2(p, n) \Leftrightarrow \frac{n-p+1}{np} Y \sim F(p, n-p+1)$

- Wilk's lambda distribution
  - Wilks's lambda is to Hotelling's T^2 as F distribution is to Student's t in univariate statistics.
  - Given independent $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1)$ and $\mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n_2)$ with $n_1 \geq p$,

$$\Lambda = \frac{\det(\mathbf{W}_1)}{\det(\mathbf{W}_1 + \mathbf{W}_2)} = \frac{1}{\det(\mathbf{I} + \mathbf{W}_1^{-1}\mathbf{W}_2)} \sim \Lambda(p, n_1, n_2)$$

  - Resort to approximations for computation: $\{(p - n_2 + 1)/2 - n_1\} \ln \Lambda(p, n_1, n_2) \approx \chi^2(n_2 p)$

Hypothesis testing

- Model: $\mathbf{X} \sim f_{\boldsymbol{\theta}^*} \in \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$
  - $\boldsymbol{\theta}^*$: parameters of interest, fixed and unknown
  - $\boldsymbol{\Theta}$: the parameter space
- Hypotheses $H_0 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$ v.s. $H_1 : \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$
  - $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \emptyset$
  - $\boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$
- Rejection/critical region R
  - Reject H_0 if $\mathbf{X} \in R$
- Level α: $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta(\boldsymbol{\theta}) \leq \alpha$
  - Power function: $\beta(\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}(\mathbf{X} \in R)$
  - When $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$, $\Pr(\text{type I error}) = \beta(\boldsymbol{\theta}^*) \leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta(\boldsymbol{\theta}) \leq \alpha$
    - Type I error: H_0 is incorrectly rejected
  - When $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$, $\Pr(\text{type II error}) = 1 - \beta(\boldsymbol{\theta}^*)$
    - Type II error: H_0 is incorrectly accepted
- p-value: alternative to rejection region
  - Impossible to be well-defined in some cases
  - $p = p(\boldsymbol{x})$ is defined such that $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}\{p(\boldsymbol{x}) \in [0, \alpha]\} \leq \alpha$ for all $\alpha \in [0, 1]$
    - $R = \{\boldsymbol{x} : p(\boldsymbol{x}) \in [0, \alpha]\}$
- Necessary components in reporting a testing result
  1. Hypotheses
  2. Name of approach
  3. Value of test statistic
  4. Rejection region/p-value
  5. Conclusion: e.g., at the α level, we reject/do not reject H_0, i.e., we believe...

Likehood ratio test (LRT)

- Minimize the type II error rate subject to a capped type I error rate (under certain classical circumstances)
- Test statistic

$$\lambda(\mathbf{X}) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}; \mathbf{X})}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \mathbf{X})} = \frac{L(\hat{\boldsymbol{\theta}}_0; \mathbf{X})}{L(\hat{\boldsymbol{\theta}}; \mathbf{X})}$$

  - $\hat{\boldsymbol{\theta}}_0$: ML estimator for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$
  - $\hat{\boldsymbol{\theta}}$: ML estimator for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- Rejection region $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$
  - $\boldsymbol{x}$ is the realization of $\mathbf{X}$
  - $c \in \mathbb{R}$ is chosen such that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}(\lambda(\mathbf{X}) \leq c) = \alpha.$$

    - Have to know the null distribution of $\lambda(\mathbf{X})$, i.e., the distribution of $\lambda(\mathbf{X})$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$

- p-value

$$p(\boldsymbol{x}) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}\{\lambda(\mathbf{X}) \leq \lambda(\boldsymbol{x})\}$$

- Null distribution of $\lambda(\mathbf{X})$
  - Use the accurate distribution of $\lambda(\mathbf{X})$ if it is known; otherwise see below for an approximation.
  - As $n \to \infty$,

$$-2 \ln \lambda(\mathbf{X}) \sim \chi^2(\nu)$$

    * ν: the difference in numbers of free parameters between H_0 and H_1
    * Leading to an (asymptotic) rejection region $\{\boldsymbol{x} : -2 \ln \lambda(\boldsymbol{x}) \geq \chi^2_{\nu,1-\alpha}\}$
      · $\chi^2_{\nu,1-\alpha}$ is the $(1-\alpha)$- quantile of $\chi^2(\nu)$.