

Summary of Lecture 3

- Data X_1, X_2, \dots iid real random variables.
- Nonparametric model: cdf F can be any cdf.
- Estimate F by

$$\hat{F}_n(x) = \frac{\# X_i \leq x}{\text{sample size}} = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x)$$

- Properties of F :

$$n\hat{F}_n(x) \sim \text{Binomial}(n, F(x))$$

$$E(\hat{F}_n(x)) = F(x) \text{ so Bias}_{\hat{F}_n} = 0$$

$$\text{Var}(\hat{F}_n(x)) =$$

$$\text{MSE}(\hat{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}$$

$$\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \frac{F(\min\{x, y\}) - F(x)F(y)}{n}$$



Summary of Lecture 3

- Needed properties of Covariance and Variance:
- Cov is bilinear

$$\text{Cov}\left(\sum a_i U_i, \sum b_j V_j\right) = \sum_i \sum_j \text{Cov}(U_i, V_j)$$

- Var is Cov with two equal arguments:

$$\text{Var}(W) = \text{Cov}(W, W)$$

- So

$$\begin{aligned}\text{Var}\left(\sum a_i U_i\right) &= \text{Cov}\left(\sum a_i U_i, \sum a_j U_j\right) \\ &= \sum_i \sum_j a_i a_j \text{Cov}(U_i, U_j)\end{aligned}$$

- If the U_i are independent then all covariances with $i \neq j$ are 0 so

$$\text{Var}\left(\sum a_i U_i\right) = \sum a_i^2 \text{Var}(U_i).$$



A biased but reasonable estimator

- X_1, \dots, X_n iid mean μ , SD σ .
- \bar{X} is an unbiased est of μ ; $\overline{X^2} = \sum X_i^2/n$ is an unbiased est of $E(X^2)$.
- Suggests the estimate

$$\hat{\sigma}^2 = \overline{X^2} - \bar{X}^2.$$

- This estimate has expected value

$$\begin{aligned} E(\hat{\sigma}^2) &= E(\overline{X^2}) - E(\bar{X}^2) \\ &= E(X^2) - (\text{Var}(\bar{X}) + E^2(\bar{X})) \\ &= \sigma^2 - \text{Var}(\bar{X}) = \frac{n-1}{n}\sigma^2 \end{aligned}$$

- So an unbiased estimate of σ^2 is

$$s^2 = \frac{n}{n-1}\hat{\sigma}^2$$

- But $\hat{\sigma}^2$ has smaller MSE and
- s is a biased estimate of σ because

$$\sigma^2 = E(s^2) = \text{Var}(s) + E^2(s).$$



Confidence intervals

- Confidence intervals defined.
- Level β confidence interval for parameter ϕ is a pair of random variables $L(X)$, $U(X)$ such that

$$P_{\theta}(L \leq \phi \leq U) \geq \beta$$

- If \geq is $=$ the interval is *exact*; otherwise *conservative*.
- ϕ is a function of θ – like 1 co-ordinate of θ .
- Many intervals based on pivot: $T(X, \phi)$ whose distribution does not depend on θ .
- Find a, b so that

$$P_{\theta}(a \leq T(X, \phi) \leq b) = \beta$$

- Then solve the inequalities $a \leq T \leq b$ to get $L \leq \phi \leq U$.



Confidence intervals based on normal approximation

- Most common kind based on application of central limit theory.
- Suppose

$$\frac{\hat{\phi} - \phi}{\sigma_{\hat{\phi}}} \implies N(0, 1).$$

- So $\sigma_{\hat{\phi}}$ is a parameter – the **Standard Error** of $\hat{\phi}$.
- So if ϕ is the standard normal density then

$$P\left(a < \frac{\hat{\phi} - \phi}{\sigma_{\hat{\phi}}} < b\right) \rightarrow \int_a^b \phi(u) du$$

- Solve to get the interval

$$\hat{\phi} - b\sigma_{\hat{\phi}}, \hat{\phi} - a\sigma_{\hat{\phi}}$$

- Usually replace the SE $\sigma_{\hat{\phi}}$ by the Estimated SE $\hat{\sigma}_{\hat{\phi}}$.
- Usually $a = -z_{\alpha/2} = -b$ to get a level $\beta = 1 - \alpha$ interval.
- If ϕ occurs in $\sigma_{\hat{\phi}}$ then the interval will not be more complex; see homework for $F(x)$ case.



Course coverage

- Chapter 3.3.
- Chapter 7.1.

