Summary of Lecture 3

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

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

- Properties of F:

$$\begin{aligned}
n\hat{F}_n(x) &\sim \text{Binomial}(n, F(x)) \\
\mathrm{E}(\hat{F}_n(x)) &= F(x) \text{ so } \text{Bias}_{\hat{F}_n} = 0 \\
\mathrm{Var}(\hat{F}_n(x)) &= \\
\mathrm{MSE}(\hat{F}_n(x)) &= \frac{F(x)(1-F(x))}{n} \\
\mathrm{Cov}(\hat{F}_n(x), \hat{F}_n(y)) &= \frac{F(\min\{x,y\}) - F(x)F(y)}{n}
\end{aligned}$$

Summary of Lecture 3

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

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

- Var is Cov with two equal arguments:

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

- So

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

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

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

A biased but reasonable estimator

- $X_1, \ldots, X_n$ 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 $\mathrm{E}(X^2)$.
- Suggests the estimate

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

- This estimate has expected value

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

- 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

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

Course coverage

- Chapter 3.3.
- Chapter 7.1.