ISyE 8813 - Quiz

January 7, 2017

1. Let A and B be two events on a probability space with positive probability. Categorize the following statements as either TRUE or FALSE. If TRUE, provide a brief justification; if FALSE, change one part of the statement to make it true.

   (a) If $\mathbb{P}(A) = 1/3$ and $\mathbb{P}(B^c) = 1/4$, then A and B cannot be disjoint.
   Answer: (Casella and Berger, 1.13) TRUE. If A and B were disjoint, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 1/3 + 3/4 = 13/12$, which is impossible. More generally, if A and B were disjoint, then $A \subseteq B^c$ and $\mathbb{P}(A) \leq \mathbb{P}(B^c)$. But here $\mathbb{P}(A) > \mathbb{P}(B^c)$, so A and B cannot be disjoint.

   (b) If A and B are independent events, then:

   $$\mathbb{P}(A|A \cup B) = \frac{\mathbb{P}(A)}{\mathbb{P}(A) + \mathbb{P}(B)}.$$

   Answer: FALSE, because $\mathbb{P}(A \cup B)$ does not necessarily equal $\mathbb{P}(A) + \mathbb{P}(B)$. The statement holds when "independent" is replaced with "mutually exclusive".

   (c) If $\mathbb{P}(B) = 1$, then $\mathbb{P}(A|B) = \mathbb{P}(A)$.
   Answer: (Casella and Berger, 1.38a) TRUE, because:

   $$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)}{\mathbb{P}(A \cup B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

2. Let X and Y be two random variables, and define:

   $$X \wedge Y = \min(X, Y) \quad \text{and} \quad X \vee Y = \max(X, Y).$$

   (a) Show that $\mathbb{E}(X \vee Y) = \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(X \wedge Y)$.
   Answer: (Casella and Berger, 2.15) The result follows by the linearity of expectation and the fact that $X \vee Y = X + Y - X \wedge Y$.

   (b) Suppose X and Y follow an exponential distribution with rate parameter 1. Use part (a) to compute $\mathbb{E}(X \vee Y)$.
   Answer: $X \wedge Y$ follows an exponential distribution with rate parameter 2, so $\mathbb{E}(X \vee Y) = 3/2$.

3. A manufacturer packages booklets in boxes of 100. It is known that, on average, each booklet weighs 1 ounce, with a standard deviation of .05 ounce. The manufacturer is interested in calculating:

$$\mathbb{P}(100 \text{ booklets weigh more than } 100.5 \text{ ounces}),$$

a probability which would help detect whether too many booklets are being put in a box. Approximate this probability, and mention any relevant theorems or assumptions needed. (Hint: $\mathbb{P}(|Z| \leq 1) = 0.68$ and $\mathbb{P}(|Z| \leq 2) = 0.95$, where $Z \sim N(0,1)$.)
Answer: (Casella and Berger, 5.29) Let X_i be the weight (in ounces) of the i-th booklet in the box. Assuming the X_i's are identically and independently distributed (i.i.d.) with $\mathbb{E}(X_i) = 1$ and $\text{Var}(X_i) = 0.05^2$. The desired probability can be rewritten as:

$$\mathbb{P}\left(\sum_{i=1}^{100} X_i > 100.5\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{100} X_i}{100} > 1.005\right) = \mathbb{P}(\bar{X} > 1.005).$$

Invoking the Central Limit Theorem, $\mathbb{P}(\bar{X} > 1.005) \approx \mathbb{P}(Z > \frac{1.005-1}{.05/\sqrt{100}}) = \mathbb{P}(Z > 1) = (1 - 0.68)/2 = 0.16$, *where* $Z \sim N(0,1)$.

4. Define the random variables $Y_1, \cdots, Y_n$ as:

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where $x_1, \cdots, x_n$ are fixed covariates, and $\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$.

(a) Define $\hat{\beta}_1 = (\sum_{i=1}^n Y_i)/(\sum_{i=1}^n x_i)$. Find an expression for $\mathbb{E}(\hat{\beta}_1)$ in terms of β.
Answer: (Casella and Berger, 7.20)

$$\mathbb{E}(\hat{\beta}_1) = \mathbb{E}\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}\right) = \frac{\sum_{i=1}^n \mathbb{E}(Y_i)}{\sum_{i=1}^n x_i} = \beta \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} = \beta.$$

(b) Define $\hat{\beta}_2 = [\sum_{i=1}^n (Y_i/x_i)]/n$. Find an expression for $\mathbb{E}(\hat{\beta}_2)$ in terms of β.
Answer: (Casella and Berger, 7.21)

$$\mathbb{E}(\hat{\beta}_2) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i}\right) = \frac{1}{n}\sum_{i=1}^n \frac{\mathbb{E}(Y_i)}{x_i} = \frac{1}{n}\sum_{i=1}^n \frac{\beta x_i}{x_i} = \beta.$$

(c) Would you prefer $\hat{\beta}_1$ or $\hat{\beta}_2$? Explain. (Hint: Compare the variance of the two estimators.)
Answer: (Casella and Berger, 7.20 and 7.21) Note that:

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}\right) = \frac{\sum_{i=1}^n \text{Var}(Y_i)}{(\sum_{i=1}^n x_i)^2} = \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2},$$

and:

$$\mathrm{Var}(\hat{\beta}_2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{x_i}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\mathrm{Var}(Y_i)}{x_i^2} = \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\frac{1}{x_i^2}.$$

But $\mathrm{Var}(\hat{\beta}_1) \leq \mathrm{Var}(\hat{\beta}_2)$*, because the arithmetic mean of* $\{x_i^2\}_{i=1}^n$ *is never less than the harmonic mean of* $\{x_i^2\}_{i=1}^n$*. Since both* $\hat{\beta}_1$ *and* $\hat{\beta}_2$ *are unbiased estimators of* β*,* $\hat{\beta}_1$ *should be preferred.*

5. Consider the typical regression set-up:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y} \in \mathbb{R}^n$ is the vector of observed response, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the model matrix of covariates, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the coefficient vector, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

(a) Let $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ be the least-squares estimator. Find an expression for $\mathbb{E}(\hat{\boldsymbol{\beta}})$ in terms of $\boldsymbol{\beta}$.
Answer:

$$\begin{aligned}\mathbb{E}(\hat{\boldsymbol{\beta}}) &= \mathbb{E}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})\\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\mathbf{y})\\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\beta}.\end{aligned}$$

(b) Find an expression for $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ in terms of $\boldsymbol{\beta}$, σ^2 and $\mathbf{X}$.
Answer:

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y})\\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.\end{aligned}$$