Midterm Practice

September 30, 2025

Instructions

- You have the entire class period to complete the exam.
- You may have a one page (single-sided) "cheat sheet" which must be turned in with the exam, but no electronics (including calculators)
- Each problem is worth 5 points.

Exercise 1 (A conditional Bernoulli model)**:** Suppose

$$X \sim \text{Bernoulli}\big(\tfrac{1}{2}\big)\,, \qquad Y \mid X \sim \text{Bernoulli}\big(\tfrac{1}{4}\,(1-X) + \tfrac{3}{4}\,X\big)\,.$$

Compute $E[X \mid Y = 1]$.

Exercise 2 (Confidence intervals)**:** An experiment is conducted to see if a certain cancer drug works. The experiment involves treating N cells with a drug and counting the number Y that survive. The goal is to estimate the probability a cell survives. If the real probability is $q = 1/4$, how many cells are needed for there to be a 95% chance the estimate is within $1/10$ of this q?

Exercise 3: Let $X \sim \text{Bernoulli}(q)$ and $X_1, \ldots, X_N$ be iid samples of X. Let $Y = \sum_{i=1}^{N} X_i$. Consider the estimator

$$\hat{q}_L = \frac{Y+1}{N+2}.$$

Derive the mean squared error and identify the bias and variance.

Exercise 4 (Translating code to mathematics)**:** Up to a few significant digits, what will the following code print?

```
> import numpy as np
> def SomeModel(num_samples):
>   z = 0.0
>   for j in range(10):
>     z = z + np.random.binomial(10,0.5,num_samples)
>   return z
> z = SomeModel(10000000)
> print(np.var(z))
```

Exercise 5: Consider a linear regression model with two predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \qquad \varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2),$$

where

$$X_1 \sim \text{Normal}(\mu_1, \sigma_1^2), \qquad X_2 \sim \text{Normal}(\mu_2, \sigma_2^2).$$

Assume X_1 and X_2 are independent and both are independent of ε. What are the marginal mean and variance of Y? Is Y normal?

Solutions

1. **A conditional Bernoulli model.** We have

$$\mathbb{P}(Y=1) = \tfrac{1}{2}\cdot\tfrac{1}{4} + \tfrac{1}{2}\cdot\tfrac{3}{4} = \tfrac{1}{2}. \tag{1}$$

Also,

$$\mathbb{P}(X=1, Y=1) = \tfrac{1}{2}\cdot\tfrac{3}{4} = \tfrac{3}{8}. \tag{2}$$

Hence

$$\mathbb{E}[X \mid Y=1] = \mathbb{P}(X=1 \mid Y=1) = \frac{\frac{3}{8}}{\frac{1}{2}} = \tfrac{3}{4}. \tag{3}$$

2. **Confidence intervals.** Let $\hat{q} = Y/N$ for $Y \sim \text{Binomial}(N, q)$ with $q = \frac{1}{4}$. By the normal approximation,

$$\mathbb{P}\left(|\hat{q} - q| \le z\sqrt{\frac{q(1-q)}{N}}\right) \approx 0.95, \tag{4}$$

with $z \approx 2$ for 95% confidence.

We require

$$2\sqrt{\frac{q(1-q)}{N}} = 0.1. \tag{5}$$

Solving,

$$N = \frac{4\,q(1-q)}{0.1^2} = \frac{4\times\frac{3}{16}}{0.01} = \frac{0.75}{0.01} = 75. \tag{6}$$

3. **MSE of $\hat{q}_L = \dfrac{Y+1}{N+2}$ for** $Y \sim \text{Binomial}(N, q)$. First,

$$\mathbb{E}[\hat{q}_L] = \frac{\mathbb{E}[Y]+1}{N+2} = \frac{Nq+1}{N+2}, \tag{7}$$

so the bias is

$$\text{Bias}(\hat{q}_L) = \mathbb{E}[\hat{q}_L] - q = \frac{1-2q}{N+2}. \tag{8}$$

Next,

$$\text{Var}(\hat{q}_L) = \frac{\text{Var}(Y)}{(N+2)^2} = \frac{Nq(1-q)}{(N+2)^2}. \tag{9}$$

Therefore, the mean squared error is

$$\text{MSE}(\hat{q}_L) = \text{Bias}^2 + \text{Var} = \frac{(1-2q)^2 + Nq(1-q)}{(N+2)^2}. \tag{10}$$

4. **Translating code to mathematics.** Inside the loop we sum 10 independent draws of $\text{Binomial}(10, 0.5)$, hence

$$z \sim \text{Binomial}(100, 0.5) \text{ (elementwise over samples)}, \tag{11}$$

so

$$\text{Var}(z) = 100 \times 0.5 \times 0.5 = 25. \tag{12}$$

With 10^7 samples, `np.var(z)` will print approximately

$$25.0. \tag{13}$$

5. **Linear regression with independent normal predictors.** Using linearity of expectation and variance (for independent variables) we have

$$\mathbb{E}[Y] = \beta_0 + \beta_1\mu_1 + \beta_2\mu_2, \tag{14}$$
$$\text{Var}(Y) = \beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + \sigma_\varepsilon^2, \tag{15}$$

and by closure of the normal family under linear transformation

$$Y \sim \text{Normal}\big(\beta_0 + \beta_1\mu_1 + \beta_2\mu_2,\ \ \beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + \sigma_\varepsilon^2\big)\,. \tag{16}$$