Midterm Practice

October 3, 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 (Reading a contingency table: conditionals)**:** You are given the joint distribution of (X, Y) where $X \in \{0, 1\}$ and $Y \in \{a, b, c\}$. The table entries are $\mathbb{P}(X = x, Y = y)$ and the probabilities sum to 1.

	$Y = a$	$Y = b$	$Y = c$
$X = 0$	0.10	0.15	0.25
$X = 1$	0.20	0.10	0.20

Compute:

(a) $\mathbb{P}(X = 1 \mid Y = c)$.

(b) $\mathbb{P}(Y = b \mid X = 0)$.

Exercise 2 (Sampling distribution of an estimator for a binomial proportion)**:** Let $Y_1, \ldots, Y_N$ be iid samples from Binomial(M, q) with unknown $q \in (0, 1)$ (you can assume $M > 1$ is known). Write down an unbiased estimator of q and derive its standard error.

Exercise 3 (Translating code to math)**:** What will the following code print?

```
> import numpy as np
> n = 1000
> x = np.random.choice([0,1,10],p=[1/3,1/3,1/3],size=n)
> y = np.random.normal(x,1,n)
> print(np.var(y))
```

Exercise 4 (Translating math to code)**:** Consider the following probability model

$$\begin{aligned} X &\sim \text{Bernoulli}(q) \\ P(Y = 1|X = 0) &= 1/2 \\ P(Y = 2|X = 0) &= 1/4 \\ P(Y = 3|X = 0) &= 1/4 \\ P(Y = 1|X = 1) &= 1 \end{aligned}$$

Write a Python function which simulatemodel(q,n) generates n samples of X and Y.

Exercise 5 (Compute the correlation coefficient from a linear model)**:** Suppose (X, Y) follow the simple linear regression with intercept

$$Y|X \sim \text{Normal}(\beta_0 + \beta_1 X, \sigma_\epsilon^2) \tag{1}$$

Assume $\mathbb{E}[X] = \mu_X$ and $\text{Var}(X) = \sigma_X^2$. For the parameter values

$$\beta_0 = 1, \quad \beta_1 = 2, \quad \mu_X = 2, \quad \sigma_X^2 = 9, \quad \sigma_\varepsilon^2 = 16,$$

compute the following

(a) $\text{Cov}(X, Y)$

(b) $\text{Var}(Y)$.

(c) The correlation coefficient ρ.

Solutions

1. **Reading a contingency table: conditionals.**

$$\mathbb{P}(X = 1 \mid Y = c) = \frac{\mathbb{P}(X = 1, Y = c)}{\mathbb{P}(Y = c)} = \frac{0.20}{0.25 + 0.20} = \frac{0.20}{0.45} = \tfrac{4}{9}. \tag{2}$$

$$\mathbb{P}(Y = b \mid X = 0) = \frac{\mathbb{P}(X = 0, Y = b)}{\mathbb{P}(X = 0)} = \frac{0.15}{0.10 + 0.15 + 0.25} = \frac{0.15}{0.50} = 0.30. \tag{3}$$

2. **Sampling distribution of an estimator for a binomial proportion.** Consider the estimator

$$\hat{q} = \frac{1}{NM} \sum_{i=1}^{N} Y_i.$$

Since $Y_i \sim \text{Binomial}(M, q)$,

$$\mathbb{E}[\hat{q}] = \frac{1}{NM} \sum_{i=1}^{N} \mathbb{E}[Y_i] = \frac{1}{NM} \cdot N(Mq) = q, \tag{4}$$

so $\hat{q}$ is unbiased. Its variance is

$$\text{Var}(\hat{q}) = \frac{1}{N^2M^2} \sum_{i=1}^{N} \text{Var}(Y_i) = \frac{1}{N^2M^2} \cdot N\big(Mq(1-q)\big) = \frac{q(1-q)}{NM}. \tag{5}$$

Thus the standard error is

$$\text{se}(\hat{q}) = \sqrt{\tfrac{q(1-q)}{NM}}.$$

3. **Translating code to math.** The code samples

$$X \in \{0, 1, 10\}, \qquad \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 10) = \tfrac{1}{3},$$

and then $Y \mid X \sim \text{Normal}(X, 1)$ is a linear regression model. The code prints the marginal variance. We therefore use $\sigma_Y^2 = \beta_1^2\sigma_X^2 + \sigma_\epsilon^2$.

We compute

$$\mathbb{E}[X] = \tfrac{1}{3}(0 + 1 + 10) = \tfrac{11}{3}, \qquad \mathbb{E}[X^2] = \tfrac{1}{3}(0^2 + 1^2 + 10^2) = \tfrac{101}{3}.$$

Therefore

$$\text{Var}(X) = \frac{101}{3} - \left(\tfrac{11}{3}\right)^2 = \frac{101}{3} - \frac{121}{9} = \frac{182}{9}.$$

Thus the code prints approximately

$$\text{Var}(Y) = \frac{182}{9} + 1$$

4. **Translating math to code.** One correct Python implementation:

```
> import numpy as np
>
> def sample_XY(q, n):
>     x =np.random.choice([0,1],size=n)
>     y = []
>     for i in range(n)
>         if x[i] == 0:
>             y.append(np.random.choice([1,2,3], p=[0.5,0.25,0.25]))
>         else:
>             y.append(1)
>     return x,y
```

5. **Correlation from the linear model.** From the model $Y = \beta_0 + \beta_1 X + \varepsilon$

$$\mathrm{Cov}(X, Y) = \beta_1 \,\mathrm{Var}(X) = 2 \cdot 9 = 18, \tag{6}$$

$$\mathrm{Var}(Y) = \beta_1^2 \sigma_X^2 + \sigma_\varepsilon^2 = 4 \cdot 9 + 16 = 52. \tag{7}$$

Therefore

$$\rho = \frac{18}{\sqrt{9 \cdot 52}} = \frac{18}{3\sqrt{52}} = \frac{6}{\sqrt{52}} = \frac{3}{\sqrt{13}}$$