Midterm Review Session problems

October 6, 2025

Exercise 1 (Filling a contingency table from conditionals)**:** You are told that (X, Y) is a pair of random variables where $X \in \{0, 1\}$ and $Y \in \{0, 1, 2\}$. The joint distribution is not given, but you are told the following information:

$$P(Y = 0) = 1/2, \quad P(Y = 1) = 1/4, \quad P(Y = 2) = 1/4$$

and the conditional probabilities of X given Y are

$$P(X = 1 \mid Y = 0) = 1/2, \quad P(X = 1 \mid Y = 1) = 1/4, \quad P(X = 1 \mid Y = 2) = 1/4$$

Fill in the contingency table for the joint distribution $P(X = x, Y = y)$:

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$			
$X = 1$			

Exercise 2 (Covariance of two Bernoulli random variables)**:** Let (X, Y) be a pair of Bernoulli random variables with joint distribution given by

	$Y = 0$	$Y = 1$
$X = 0$	0.3	0.2
$X = 1$	0.1	0.4

Compute $\text{cov}(X, Y)$.

Exercise 3 (Sample distribution)**:** Consider the problem of estimating the mean of a normal random variable $X \sim \text{Normal}(\mu, \sigma^2)$ with known variance σ^2. You have N iid samples $Y_1, \ldots, Y_N \sim \text{Normal}(\mu, 1)$ where μ is unknown. Write down an estimator whose sample distribution is

$$\hat{\mu}_{☺} \sim \text{Normal}\left(\mu a + c/N, \frac{a^2\sigma^2}{N}\right), \quad a, c \neq 0$$

Exercise 4: In class we learned that given N samples $X_1, \ldots, X_N$ from a Normal distribution with unknown mean

$$\hat{\sigma}_1^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

is a biased estimator of the variance. Explain (without writing out any code) how you could test that this is indeed biased in Python. Specifically, describe what data you would generate and a plot you could make.

Exercise 5 (Translating code to math)**:** What will the following code print?

```
> import numpy as np
> n = 1000
> x = np.random.binomial(3,1/2,size=n)
> y = np.random.normal(1+ x,1,n)
> print(np.mean(y))
```

Exercise 6 (Translating math to code)**:** Consider the model

$$
\begin{aligned}
X_1 &\sim \text{Bernoulli}(1/2) \\
X_2|X_1 &\sim \text{Bernoulli}(X_1/2 + 1/4) \\
X_3|X_2 &\sim \text{Normal}(X_1 - X_2, X_1 + 1)
\end{aligned}
$$

Write a python code to approximate $P(X_3 < 2|X_2 = 0)$.

Exercise 7: Suppose

$$X \sim \text{Bernoulli}(1/2)Y|X \sim \text{Bernoulli}$$

Exercise 8 (Recovering regression parameters)**:** Let

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

Given the marginal mean, expectation and correlation coefficient

$$\mu_X = 2, \qquad \sigma_X^2 = 1, \qquad \mu_Y = 5, \qquad \sigma_Y^2 = 4, \qquad \rho = 1/2$$

FInd β_1, β_0 and σ_ε^2.

Exercise 9: The following is a plot from the paper *Lithium deficiency and the onset of Alzheimer's disease* published in Nature. It shows Lithium levels in two groups (NCI = no cognitive impairment, AD = Alzheimer's disease). Why might a linear regression model with the group as the predictor X and Lithium level as the response Y be inappropriate for this data?