

Math Camp - Homework 10

Question 1: For a discrete random variable X , show that:

$$\text{Var}(X) = E[X^2] - E[X]^2$$

Question 2: X is a discrete random variable. It takes the value of the number of days required for a governmental agency to respond to a request for information. X is distributed according to the following PMF:

$$f(x) = e^{-4} \frac{4^x}{x!} \text{ for } x \in \{0, 1, 2, \dots\}$$

- (a) Given this information, what is the probability of a response from the agency in 3 days or less?
- (b) What is the probability the agency response takes more than 10 but less than 13 days?
- (c) What is the probability the agency response takes more than 5 days?
- (d) Suppose using X you generate a new variable, *Responsive*. *Responsive* equals 1 if an agency responds in 5 days or less and 0 otherwise. What is the expected value of *Responsive*?
- (e) What is the variance of *Responsive*?

Question 3: Suppose we've developed a model predicting the outcome of the upcoming midterm elections in a state with 4 Congressional districts. In each district there are two candidates, a Republican and a Democrat. We have reason to believe the following PMF describes the distribution of potential election results where $K \in \{0, 1, 2, 3, 4\}$ and is the number of seats won by Republican candidates in the upcoming election.

$$P(K = k|\theta) = \binom{4}{k} \theta^k (1 - \theta)^{4-k}$$

Based on polling information, we think the appropriate value for θ is 0.55.

- (a) What's the expected number of seats Republicans will win in the upcoming election?
- (b) Given this PMF, what's the probability that no Republican legislators win in the upcoming election?
- (c) What's the probability that Republican legislators win a majority of the seats in this state?
- (d) A prominent political pundit declares they are certain that Republicans will win a majority of seats in the next election and offers the following bet. If Republicans win a majority of the seats, we must pay the pundit \$15.00. If Republicans fail to win a majority of states, we will

win \$20.00. Based on our model, should we take this bet?

Hint: Think of the betting outcomes as a random variable. Find the expected value of this random variable.

- (e) Suppose we are offered a second bet with a more complicated structure. In this case we'll receive \$100 if the Republicans win a majority, we'll receive \$50 if neither party wins a majority, and we'll have to pay \$200 if the Democrats win a majority. Should we take this bet?

Question 4: Based on the success of our model in Question 3, we're now asked to model election outcomes in another state where there are 5 elections taking place. However, in this state there are three candidates (Democrat, Republican, and Independent) in each district, which requires a fundamental reworking of our model.

We believe the following PMF describes the distribution of potential election results where $W_i \in \{0, 1, 2, 3, 4, 5\}$ and is the number of seats won by candidates from a given party in the upcoming election. W_1 is the number of seats won by Democrats, W_2 is the number of seats won by Republicans, and W_3 is the number of seats won by Independent candidates.

$$P(W_i = w_i | \theta_i) = \frac{5!}{w_1! w_2! w_3!} \theta_1^{w_1} \theta_2^{w_2} \theta_3^{w_3}$$

Using polling information we believe $\theta_1 = 0.6$, $\theta_2 = 0.35$, and $\theta_3 = 0.05$, where θ_1 refers to the probability of a Democrat winning a given election, θ_2 refers to the probability of a Republican winning a given election, and θ_3 refers to the probability of an Independent winning a given election.

- (a) Based on this model, what is the probability one Independent, two Republicans, and two Democrats win elections?
- (b) What is the expected number of seats that Independent candidates will win in the upcoming election?
- (c) What is the probability that exactly one Independent candidate wins a seat in the election?
- (d) Suppose general dissatisfaction with the two major parties boosts θ_3 from 0.05 to 0.06 with θ_1 and θ_2 each reduced by 0.005. How would this change your answer to the previous question?

Question 5:

The random variable X has a probability mass function:

$$f(x; \lambda) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

for $x = 0, 1, 2, \dots$ (i.e. X has a Poisson distribution with parameter λ). In a lengthy manuscript, it is discovered that only 13.5 percent of the pages contain no typing errors.

If we assume that the number of errors per page is a random variable with a Poisson distribution, find the percentage of pages with exactly one error.

Question 6:

A prominent hypothesis in the comparative politics literature is the “resource curse”: in countries that depend economically on extraction of natural resources, there is a lower probability of democratic political institutions.¹ Suppose we want to investigate this hypothesis. For each country, we collect data on whether the country is democratic and whether it depends on oil exports. Denote the outcome variable (democracy) for country i as y_i , where $y_i = 1$ for democracies and $y_i = 0$ for non-democracies. Similarly, denote the predictor variable (dependence on oil) as x_i , with $x_i = 1$ for oil-dependent states and $x_i = 0$ for non-oil-dependent states.

For each country, we model y_i as a Bernoulli random variable with rate π_i (note the subscript), with the following PMF:

$$f(y_i | \pi_i) = (\pi)^{y_i} (1 - \pi)^{1-y_i}$$

To incorporate the predictor, we’ll let π_i vary as a function of x_i . We’ll model this using the inverse logit link function:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

Our goal in 450C will be to estimate β_0 and β_1 using maximum likelihood estimation. These parameters will tell us how the probability varies as a function of X . Let’s say for now we ran this regression and found that $\hat{\beta}_0 = 0.35$ and $\hat{\beta}_1 = -0.9$.

- (a) Using the estimates of β_0 and β_1 , what is the probability that an oil-dependent state is a democracy? What is the probability that a non-oil-dependent state is a democracy?
- (b) Mr. Critic comes along and tells you that your model is completely bogus because it doesn’t include region of the world as a predictor. He says you need to control for whether the country is in the Middle East or not. He suggests that you add another variable z_i where $z_i = 1$ if the country is in the Middle East and 0 otherwise. Now you model π_i as follows:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i + \beta_2 z_i)}{1 + \exp(\beta_0 + \beta_1 x_i + \beta_2 z_i)}$$

After running the regression again, we find $\hat{\beta}_0 = 0.2$, $\hat{\beta}_1 = -0.1$, and $\hat{\beta}_2 = -1.2$. How do your answers to (a) change with these new estimates?

- (c) Interpret the logit regression: based on these data, what should you conclude about the resource curse? What could account for the discrepancy between the results in (a) and (b)?

¹The empirical evidence for this hypothesis is shaky (Haber and Menaldo 2011).