

## Math Camp - Homework 10 Solutions

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

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

**Solution** A useful property to know here is that  $E[E[X]] = E[X]$ .

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])^2] \\&= E[X^2 - (2E[X] \times X) + E[X]^2] \\&= E[X^2] - 2E[E[X] \times X] + E[E[X]^2] \\&= E[X^2] - 2E[X]E[X] + E[X]^2 \\&= E[X^2] - 2E[X]^2 + E[X]^2 \\&= E[X^2] - E[X]^2\end{aligned}$$

**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*?

## Solutions

(a)  $P(X \leq 3) \approx .43$

$$\begin{aligned}
 P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
 P(X = 0) &= e^{-4} \frac{4^0}{0!} = 0.01831564 \\
 P(X = 1) &= e^{-4} \frac{4^1}{1!} = 0.07326256 \\
 P(X = 2) &= e^{-4} \frac{4^2}{2!} = 0.1465251 \\
 P(X = 3) &= e^{-4} \frac{4^3}{3!} = 0.1953668 \\
 P(X \leq 3) &= 0.01831564 + 0.07326256 + 0.1465251 + 0.1953668 = 0.4334701
 \end{aligned}$$

(b)  $P(10 < X < 13) \approx .003$

$$\begin{aligned}
 P(10 < X < 13) &= P(X = 11) + P(X = 12) \\
 P(X = 11) &= \frac{4^{11}}{11!} = 0.001924537 \\
 P(X = 12) &= \frac{4^{12}}{12!} = 0.0006415123 \\
 P(10 < X < 13) &= 0.001924537 + 0.0006415123 = 0.002566049
 \end{aligned}$$

(c)  $P(X > 5) = 1 - P(X \leq 5) \approx 0.22$

We already know  $P(X \leq 3) = .43$  so now we only need to find  $P(X=4)$  and  $P(X=5)$

$$\begin{aligned}
 P(X \leq 3) &= 0.4334701 \\
 P(X = 4) &= \frac{4^4}{4!} = 0.1953668 \\
 P(X = 5) &= \frac{4^5}{5!} = 0.1562935 \\
 P(X \leq 5) &= 0.4334701 + 0.1953668 + 0.1562935 = 0.7851304 \\
 P(X > 5) &= 1 - 0.7851304 = 0.2148696
 \end{aligned}$$

(d) Since *Responsive* is dichotomous,  $E[\text{Responsive}] = P(\text{Responsive} = 1) = P(X \leq 5)$ . From the previous question we know that  $P(5 < X) = 0.2148696$  and  $P(X \leq 5) = 0.7851304$ . So  $E[\text{Responsive}] = 0.7851304$ .

(e) *Responsive* is distributed Bernoulli.  $\text{Var}(\text{Bernoulli}) = p(1 - p)$ . In this case  $p = 0.7851304$ . So  $\text{Var}(\text{Responsive}) = (0.7851304 \times (1 - 0.7851304)) = 0.1687007$ .

**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  $\theta$  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 Republican's 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, \$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?

## Solutions

- (a) The election outcomes are distributed according to a binomial distribution so we find the expected value given our two parameters, n and  $\theta$ .

$$\begin{aligned} E[Binomial] &= n \times \theta \\ &= 4 \times .55 \\ &= 2.2 \end{aligned}$$

- (b)  $P(k=0) \approx 0.04100625$

$$\begin{aligned} P(k = 0) &= \binom{4}{0} .55^0 (1 - .55)^4 \\ &= 1 \times 1 \times 0.04100625 \\ &= 0.04100625 \end{aligned}$$

(c)  $P(2 < k) \approx .39$ .

$$\begin{aligned}
P(k = 3) &= \binom{4}{3} .55^3 (1 - .55)^1 = 4 \times 0.07486875 = 0.299475 \\
P(k = 4) &= \binom{4}{4} .55^4 (1 - .55)^0 = 1 \times 0.09150625 = 0.09150625 \\
P(2 < k) &= 0.299475 + 0.09150625 = 0.3909813
\end{aligned}$$

(d) We can should take this bet as we will net an expected return of \$6.32 from it.

We can think of  $bet$  as a variable with the associated values:

$$bet = \begin{cases} -15 & \text{if Republican majority;} \\ 20 & \text{otherwise.} \end{cases}$$

$$\begin{aligned}
E[bet] &= E[bet|Republican Majority] + E[bet|Otherwise] \\
&= -15 \times P(Republican Majority) + 20 \times P(Otherwise) \\
&= -15 * (0.3909813) + 20 * (1 - 0.3909813) \\
&= -5.864719 + 12.18037 \\
&\approx 6.3
\end{aligned}$$

(e) We should take this bet as well since we will net an expected \$9.18.

We can think of  $bet_2$  as a variable that takes the following values:

$$bet_2 = \begin{cases} 100 & \text{if Republican majority;} \\ 50 & \text{if tie;} \\ -200 & \text{otherwise.} \end{cases}$$

$$\begin{aligned}
P(Rep Majority) &= 0.3909813 \\
P(tie) &= \binom{4}{2} .55^2 (1 - .55)^2 = 6 \times 0.06125625 = 0.3675375 \\
P(Dem Majority) &= 1 - P(Rep Majority) - P(tie) \\
&= 1 - 0.3909813 - 0.3675375 \\
&= 0.2414812 \\
E[bet_2] &= E[bet_2|Rep Majority] + E[bet_2|Tie] + E[bet_2|Otherwise] \\
&= 100 \times P(Republican Majority) + 50 \times P(Tie) \\
&\quad + -200 \times P(Republicans don't win majority) \\
&= 100 * 0.3909813 + 50 * 0.3675375 - 200 * 0.2414812 \\
&\approx 9.2
\end{aligned}$$

**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  $\theta_1 = .6, \theta_2 = .35$  and  $\theta_3 = .05$ .

- Based on this model, what is the probability one Independent, two Republicans and two Democrats win elections?
- What is the expected number of seats that Independent candidates will win in the upcoming election?
- What is the probability that one Independent candidate wins a seat in the election?
- Suppose general dissatisfaction with the two major parties boosts  $\theta_3$  from .05 to .06 with  $\theta_1$  and  $\theta_2$  each reduced by .005. How would this change you answer to the previous question?

## Solutions

- $P(2,2,1) \approx .07$

$$\begin{aligned} P(W_1 = 2, W_2 = 2, W_3 = 1) &= \frac{5!}{2!2!1!} .6^2 .35^2 .05^1 \\ &= 0.06615 \end{aligned}$$

- We're dealing with a multinomial distribution so, from the slides.

$$\begin{aligned} E[W_3] &= n \times \theta_3 \\ &= 5 \times .05 \\ &= .25 \end{aligned}$$

- $P(W_3 = 1) \approx .2$ .

The easiest way to go about this is to rewrite the multinomial as a binomial, since we now only have two relevant outcomes: one for where *either* a Rep or Dem wins ( $W_{1 \cup 2}$ ) and a second for where an Independent wins ( $W_3$ )

$$\begin{aligned} P(W_{1 \cup 2} = 4, W_3 = 1) &= \frac{5!}{4!1!} .95^4 .05^1 \\ &= 0.2036266 \end{aligned}$$

- (d) The .01 increase in the win probability for the independents in each election produces an increase of about .03 that one Independent candidate will win a seat.

$$\begin{aligned} P(W_1 \cup 2 = 4, W_3 = 1) &= \frac{5!}{4!1!} .94^4 .06^1 \\ &= 0.2342247 \end{aligned}$$

### Question 5:

The random variable  $X$  has a probability density 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  $\lambda$ ). 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.

Answer:

With a Poisson variable, the pdf supplies probability of a given number (count) of events, i.e.  $Pr(X = x)$ . From the prompt, we know that  $Pr(X = 0) = .135$ . Using this, we can solve for  $\lambda$ , which is the average rate of “success” in a given interval, (i.e. the average number of times an event occurs in a given interval). Here, the event is an error occurring and the interval is the page.

$$\begin{aligned} f(x; \lambda) &= \frac{\lambda^x \exp(-\lambda)}{x!} \\ f(x = 0; \lambda) &= \frac{\lambda^0 \exp(-\lambda)}{0!} = .135 \\ \frac{\exp(-\lambda)}{1} &= .135 \\ \frac{1}{\exp(\lambda)} &= .135 \\ \exp(\lambda) &= \frac{1}{.135} \\ \exp(\lambda) &= 7.407 \\ \lambda &= 2.002 \end{aligned}$$

Now that we know the value of  $\lambda$ , we know that on average, a page contains 2.002 errors. (Note: If we want to change the interval to say, 10 pages, we just adjust  $\lambda$  accordingly. If there are 2.002 errors per page, there are  $2.002 \times 10 = 20.02$  errors per 10 pages.) Now we can solve for  $Pr(X = 1)$ .

$$f(x = 1; \lambda) = \frac{2.002^1 \exp(-2.002)}{1!} = .27$$

### 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.<sup>1</sup> 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  $\pi_i$  (note the subscript), with the following PMF:

$$Pr(y_i = 1 \mid \pi_i) = (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}$$

To incorporate the predictor, we’ll let  $\pi_i$  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  $\beta_0$  and  $\beta_1$  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  $\hat{\beta}_0 = .35$  and  $\hat{\beta}_1 = -0.9$ .

- (a) Using the estimates of  $\beta_0$  and  $\beta_1$ , 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  $\pi_i$  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  $\hat{\beta}_0 = .2$ ,  $\hat{\beta}_1 = -0.1$ , 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)?

---

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

## Solution

- (a) We can plug in the estimates to the formula for  $\pi$  to figure out the relevant PMF. For oil-dependent states,  $x_i = 1$ , so we have

$$\begin{aligned}\pi_{x_i=1} &= \frac{\exp(0.35 - 0.9(1))}{1 + \exp(0.35 - 0.9(1))} \\ &= \frac{\exp(-0.55)}{1 + \exp(-0.55)} \\ &\approx 0.37\end{aligned}$$

So the PMF for oil-dependent states is  $f(y_i | x_i = 1) = 0.37^{y_i}(0.63)^{1-y_i}$ , so the probability of democracy for oil-dependent states is  $f(1 | x_i = 1) = 0.37$ .

For non-oil-dependent states,  $x_i = 0$  so we have

$$\begin{aligned}\pi_{x_i=0} &= \frac{\exp(0.35 - 0.9(0))}{1 + \exp(0.35 - 0.9(0))} \\ &= \frac{\exp(0.35)}{1 + \exp(0.35)} \\ &\approx 0.59\end{aligned}$$

So the PMF for oil-dependent states if  $f(y_i | x_i = 0) = 0.59^{y_i}(0.41)^{1-y_i}$ , so the probability of democracy for non-oil-dependent states is  $f(1 | x_i = 0) = 0.59$ .

- (b) We can re-calculate  $\pi$  for each combination of oil-dependence and Middle East:

$x_i$	$z_i$	$\pi_i$	$f(y_i   \pi_i)$	$f(1   \pi_i)$
0	0	$\frac{\exp(0.2)}{1+\exp(0.2)}$	$0.55^{y_i} 0.45^{1-y_i}$	0.55
1	0	$\frac{\exp(0.2-0.1)}{1+\exp(0.2-0.1)}$	$0.52^{y_i} 0.48^{1-y_i}$	0.52
0	1	$\frac{\exp(0.2-1.2)}{1+\exp(0.2-1.2)}$	$0.27^{y_i} 0.73^{1-y_i}$	0.27
1	1	$\frac{\exp(0.2-0.1-1.2)}{1+\exp(0.2-0.1-1.2)}$	$0.25^{y_i} 0.75^{1-y_i}$	0.25

- (c) Based on this regression, it looks like oil dependence doesn't have much of an effect on the probability that a country is a democracy. A country that's not in the Middle East is only 3 percentage points less likely to be a democracy if it's oil dependent, and a country that is in the Middle East is only 2 percentage points less likely to be a democracy.