Data 100, Spring 2023

Exam Prep Section #4 Solutions

Odd Odds

1. Suppose we have the following events, and their probabilities of occurring:

   - The probability of a person being born ambidextrous (can write with both hands) are around 1 in 100
   - The probability of shuffling a suit of cards into perfect numeric order is 1 in 6 billion

   The following parts ask you to compute the probabilities of various events of interest. Don't worry about actually computing your answer; it's OK to leave it as products of fractions, etc.

   (a) You sample a single individual at random from the world population. What is the probability that that individual will **not** be ambidextrous?

   Solution:

$$\begin{aligned} P(\text{not ambidextrous}) &= 1 - P(\text{ambidextrous}) \\ &= 1 - \frac{1}{100} \\ &= \frac{99}{100} \end{aligned}$$

   (b) Suppose you have a group of 50 Data 100 students, of which you know half are ambidextrous and half are not. If you sample 3 individuals **without replacement** (i.e., via a simple random sample), what is the probability that the first two sampled are ambidextrous, and the last is not ambidextrous?

   Solution:
   In my "population", 25 students are ambidextrous and 25 are not. If I were to sample 5 without replacement, the probability that I would observe an ambidextrous person on the first draw would be (25/50). Given an ambidextrous person was drawn first, the probability that I would draw an ambidextrous person on the second draw is (24/49). So, the probability that I would draw ambidextrous people on the first two draws is $(25/50) * (24/49)$. Given the results of my first two draws, the probability I would draw a non-ambidextrous person on

the next draw would be $(25/48)$, since there are only 48 individuals left in my sample, but all 25 non-ambidextrous people are still there. So, the probability of ambidextrous for the first two draws and non ambidextrous for the third is $(25/50) * (24/49) * (25/48)$.

(c) Suppose we ask everyone on planet Earth to shuffle a deck of cards. What is the probability that at least one person shuffles the cards into perfect numeric order? Assume that each individuals' shuffling of the cards are independent, and that there are 8 billion people in the world.

Solution:
Start with the following use of the complement rule:

$$P(\text{at least one perfect shuffling in world}) = 1 - P(\text{no perfect shuffling in world})$$

Now, let's rewrite the probability that no one on Earth will shuffle perfectly:

$$P(\text{no perfect shuffling in world}) = P(\text{single person does not shuffle perfectly})^{8\times 10^9}$$

The above uses the assumption that each individual's shuffling is independent, so the probability every individual on earth will not shuffle properly is just the probability each individual will not shuffle properly, but multiplied by itself 8 billion times.
Using the complement rule again and info in the problem to compute the probability an individual person will not properly shuffle:

$$P(\text{single person does not shuffle perfectly}) = 1 - P(\text{single person perfectly shuffles perfectly})$$
$$\ldots = 1 - \frac{1}{6 \times 10^9}$$

Plugging in all of these pieces back into the probability statement we started with, we obtain:

$$P(\text{at least one perfect shuffling in world}) = 1 - \left(1 - \frac{1}{6 \times 10^9}\right)^{8\times 10^9}$$

Standardized SLR

2. Consider the simple linear regression model without an intercept, $\hat{y} = \theta x$. Assume throughout the entirety of this problem that we are performing least squares linear regression (i.e. that we are choosing parameters that minimize average squared loss), and that we are not using regularization.

   (a) True or False: The point $(\bar{x}, \bar{y})$ always lies on the regression line $\hat{y} = \hat{\theta} x$.

   ◯ A. True

   ◯ B. False

   > **Solution:** It does not. With an intercept term this statement is true, but without one this property doesn't hold. A counterexample can be had by picking any set values of x whose mean is 0, and any set of y values whose mean is not 0. Then, 0 . $\hat{\theta}$ is 0, but this cannot be equal to $\bar{y}$, which we defined to be non-zero.

   (b) Suppose we create a new feature x' using $x' = 10x$. Let $r(x, y)$ be the correlation coefficient that we discussed in class. True or False: $r(x, y) = r(x', y)$.

   ◯ A. True

   ◯ B. False

   > **Solution:** r is a measure of linear association. Multiplying one set of values by 10 doesn't change the strength of their linear association with the other set of values.

   (c) Suppose we decide to now include an intercept term in our simple linear model. We fit two models. One uses x, and the other uses x' as defined above.

   - Model A: $\hat{y} = \theta_0 + \theta_1 x$
   - Model B: $\hat{y} = \beta_0 + \beta_1 x'$

   Is $\hat{\theta}_1 = \hat{\beta}_1$? If yes, explain in one sentence why. If no, give an expression for $\hat{\beta}_1$ in terms of $\hat{\theta}_1$.

   > **Solution:** No, the optimal slopes are different, and $\hat{\beta}_1 = \frac{1}{10}\hat{\theta}_1$. We know from lecture that the optimal is $r \cdot \frac{\sigma_y}{\sigma_x}$. The correlation coefficient doesn't change when we multiply a set of values by a constant (as established in part b), and $\sigma_x' = 10\sigma_x$.

   (d) Is $\hat{\theta}_0 = \hat{\beta}_0$? If yes, explain in one sentence why. If no, give an expression for $\hat{\beta}_0$ in terms of $\hat{\theta}_0$.

Solution: Yes. Recall from lecture, $\hat{\theta}_0 = \bar{y} - \hat{\theta}_1\bar{x}$. We established in the above part that $\hat{\beta}_1 = \frac{1}{10}\hat{\theta}_1$, and $\bar{x'} = 10\bar{x}$. $\bar{y}$ is the same for both models. Putting these facts together:

$$\begin{aligned}
\hat{\theta}_0 &= \bar{y} - \hat{\theta}_1\bar{x} \\
\hat{\theta}_0 &= \bar{y} - \frac{1}{10}.\hat{\theta}_1.(10\bar{x}) \\
&= \bar{y} - \hat{\beta}_1\bar{x'} \\
&= \hat{\beta}_0
\end{aligned}$$