

APMA 1650 – Review Session 2

Monday, July 25, 2016

1. The following series of questions refers to the SAT, everyone's favorite standardized test.
 - (a) The mean score on the SAT math section is 511. What is an upper bound on the probability that a student scores over 700?

Let X be the score on the SAT math section. Since we only know the mean, the best estimate we can get comes from Markov's Inequality:

$$\begin{aligned}\mathbb{P}(X \geq 700) &\leq \frac{\mathbb{E}(X)}{700} = \frac{511}{700} \\ \mathbb{P}(X \geq 700) &\leq 0.73\end{aligned}$$

This is not a very tight bound, but it is an upper bound nonetheless.

- (b) The standard deviation of the SAT math section is 120. Can you get a better upper bound on the probability that a student scores over 700?

The variance is the square of the standard deviation, so $\text{Var}[X] = 120^2$. To get a score over 700, we need to exceed the mean by 189 points. Chebyshev's inequality gives us the probability of deviating from the mean by a certain amount, not the probability of exceeding the mean. We can, however, still use it. Chebyshev's inequality says that:

$$\begin{aligned}\mathbb{P}(|X - 511| \geq 189) &\leq \frac{\text{Var}(X)}{189^2} = \frac{120^2}{189^2} \\ \mathbb{P}(|X - 511| \geq 189) &\leq 0.40\end{aligned}$$

Since $\mathbb{P}(X \geq 700) < \mathbb{P}(|X - 511| \geq 189)$, we can also conclude that:

$$\mathbb{P}(X \geq 700) \leq 0.40$$

This is still not the greatest bound, but it is much better than the bound from Markov's inequality. If we happen to know (or suspect) that the distribution of SAT scores is symmetric about its mean, then we could divide this by 2 to get:

$$\mathbb{P}(X \geq 700) = \frac{1}{2}\mathbb{P}(|X - 511| \geq 189) = \frac{0.40}{2} = 0.20$$

- (c) The College Board makes great effort to ensure that SAT scores are roughly normally distributed. Assuming this is the case, what is the probability that a student scores over 700 on the SAT math section?

$$\begin{aligned}
\mathbb{P}(X \geq 700) &= \mathbb{P}\left(Z \geq \frac{700 - 511}{120}\right) \\
&= \mathbb{P}(Z \geq 1.575) \\
&= 0.0571
\end{aligned}$$

where we used the Z table and rounded 1.575 up to 1.58. The take-home message from this problem (and the similar one on the homework) is that the more information we know about a probability distribution, the better bound we can obtain on outlier probabilities.

2. You are a barista at a local coffee shop. The average number of customers per hour who enter your shop is 10. Assume customers arrive one-at-a-time and their arrivals are independent from each other.

- (a) What is the probability that fewer than 3 customers will enter your coffee shop in one hour?

Let X be the number of customers who enter your coffee shop in one hour. The best distribution to model X is a Poisson distribution with parameter $\lambda = 10$. Then

$$\begin{aligned}
\mathbb{P}(X < 3) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\
&= \frac{e^{-10}10^0}{0!} + \frac{e^{-10}10^1}{1!} + \frac{e^{-10}10^2}{2!}
\end{aligned}$$

- (b) What is the average time between the arrival of two customers?

Let Y be the time between the arrival of two customers. Then Y is an exponential random variable with the same parameter as the Poisson random variable X , i.e. an exponential random variable with parameter $\lambda = 10$. Using the table of common distributions, we have $E(Y) = 1/\lambda = 1/10$ hours (i.e. 6 minutes)

- (c) What is the probability that there will be an interval of 10 minutes or more between the arrival of one customer and the next?

Using the exponential density, and using the fact that 10 minutes is $1/6$ hours,

$$\begin{aligned}
\mathbb{P}(Y > 10) &= \int_{1/6}^{\infty} 10e^{-10y} dy \\
&= \lim_{t \rightarrow \infty} -e^{-10y} \Big|_{1/6}^t \\
&= e^{-10/6} = e^{-5/3} \approx 0.189
\end{aligned}$$

3. Let X and Y have a joint density function given by

$$f(x, y) = \begin{cases} cx & 0 \leq y \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of c such that this is valid joint density function.

Integrate the joint density and set the integral equal to 1. Don't forget to draw the region so that you have the correct limits of integration.

$$\begin{aligned} 1 &= \int_0^3 \int_0^x cxdydx \\ &= \int_0^3 cxy \Big|_0^x dx \\ &= \int_0^3 cx^2 dx \\ &= \frac{cx^3}{3} \Big|_0^3 \\ &= 9c \end{aligned}$$

From this we conclude that $c = 1/9$.

(b) Find the marginal densities of X and Y .

$$\begin{aligned} f_X(x) &= \int_0^x \frac{1}{9}xdy \\ &= \frac{1}{9}xy \Big|_0^x \\ &= \frac{1}{9}x^2 \end{aligned}$$

With the appropriate bounds, this is

$$f_X(x) = \begin{cases} \frac{1}{9}x^2 & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} f_Y(y) &= \int_y^3 \frac{1}{9}xdx \\ &= \frac{1}{9} \frac{x^2}{2} \Big|_y^3 \\ &= \frac{1}{18}(9 - y^2) \end{aligned}$$

With the appropriate bounds, this is

$$f_Y(y) = \begin{cases} \frac{1}{18}(9 - y^2) & 0 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(c) Find the expected values of X and Y .

$$\begin{aligned} \mathbb{E}X &= \int_0^3 x f_X(x) dx \\ &= \int_0^3 x \frac{1}{9} x^2 dx \\ &= \frac{1}{9} \int_0^3 x^3 dx \\ &= \frac{1}{9} \frac{x^4}{4} \Big|_0^3 \\ &= \frac{1}{9} \frac{3^4}{4} \\ &= \frac{9}{4} \end{aligned}$$

$$\begin{aligned} \mathbb{E}Y &= \int_0^3 y f_Y(y) dy \\ &= \int_0^3 y \frac{1}{18} (9 - y^2) dy \\ &= \frac{1}{18} \int_0^3 (9y - y^3) dy \\ &= \frac{1}{18} \left(\frac{9y^2}{2} - \frac{y^4}{4} \right) \Big|_0^3 \\ &= \frac{2}{3^2} \left(\frac{3^4}{2} - \frac{3^4}{4} \right) \\ &= \frac{3^2}{2} \left(\frac{1}{2} - \frac{1}{4} \right) \\ &= \frac{9}{8} \end{aligned}$$

(d) Find the conditional density of Y given $X = x$.

$$F(y|x) = \frac{\frac{1}{9}x}{\frac{1}{9}x^2} = \frac{1}{x}$$

With the appropriate bounds, this becomes:

$$F(y|x) = \begin{cases} \frac{1}{x} & 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

In addition, we require that $x \neq 0$. We note that this is a Uniform $[0, x]$ random variable.

- (e) Find $\mathbb{P}(Y \leq 1/2|X = 1)$

Plugging 1 in for x in the conditional density, we have:

$$F(y|X = 1) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This is the density of a Uniform $[0,1]$ random variable! Thus $\mathbb{P}(Y \leq 1/2|X = 1) = 1/2$.

- (f) Find the conditional expected value $\mathbb{E}(Y|X = x)$.

$$\begin{aligned} \mathbb{E}(y|X = x) &= \int_0^x y \frac{1}{x} dy \\ &= \frac{1}{x} \frac{y^2}{2} \Big|_0^x \\ &= \frac{1}{x} \frac{x^2}{2} \\ &= \frac{x}{2} \end{aligned}$$

This makes sense given that the conditional distribution is a Uniform $[0, x]$ random variable.

4. Let X and Y be random variables with joint density given by

$$f(x, y) = \begin{cases} 6(1 - y) & 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the covariance of X and Y . Are X and Y independent?

To do this, we use the Magic Covariance Formula:

$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

This requires finding both marginal densities. Do not forget to draw the region first!

$$\begin{aligned} f_X(x) &= \int_x^1 6(1 - y) dy \\ &= (6y - 3y^2) \Big|_x^1 \\ &= (6 - 3) - (6x - 3x^2) \\ &= 3 - 6x + 3x^2 \end{aligned}$$

$$\begin{aligned}
f_Y(y) &= \int_0^y 6(1-y)dx \\
&= 6(1-y)x \Big|_0^y \\
&= 6y(1-y)
\end{aligned}$$

The bounds on both marginal densities are from 0 to 1. Now compute the expected values.

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^1 x(3-6x+3x^2)dx \\
&= \int_0^1 (3x-6x^2+3x^3)dx \\
&= \left(\frac{3x^2}{2} - 2x^3 + \frac{3x^4}{4} \right) \Big|_0^1 \\
&= \frac{1}{4}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(Y) &= \int_0^1 y6y(1-y)dy \\
&= \int_0^1 (6y^2-6y^3)dy \\
&= \left(2y^3 - \frac{6}{4}y^4 \right) \Big|_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

Finally we find $\mathbb{E}(XY)$. Recalling the formula for the expected value of a function of two random variables,

$$\begin{aligned}
\mathbb{E}(XY) &= \int_0^1 \int_0^y xy6(1-y)dx dy \\
&= \int_0^1 6y(1-y) \int_0^y x dx dy \\
&= \int_0^1 6y(1-y) \frac{x^2}{2} \Big|_0^y dy \\
&= \int_0^1 3y^3(1-y)dy \\
&= 3 \left(\frac{y^4}{4} - \frac{y^5}{5} \right) \Big|_0^1 \\
&= \frac{3}{20}
\end{aligned}$$

Thus the covariance is

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{3}{20} - \frac{1}{4} \frac{1}{2} = \frac{1}{40}$$

The covariance is not 0, so X and Y are not independent. Furthermore, the joint density is not the product of the marginal densities.

5. A forester studying the effects of fertilization on pine forests is interested in estimating the average basal area of pine trees (basal area is the area of a given section of land that is occupied by the cross-section of tree trunks at their base). She has discovered that these measurements (in square inches) are normally distributed with standard deviation of 4 square inches.

- (a) If she samples $n = 9$ trees, what is the probability that the sample mean will be within 2 square inches of the population mean.

The sample mean \bar{Y} is normally distributed with standard deviation of $\sigma/\sqrt{n} = 4/\sqrt{9} = 4/3$. The probability we want is:

$$\begin{aligned} \mathbb{P}(|\bar{Y} - \mu| \leq 2) &= \mathbb{P}\left(\left|\frac{\bar{Y} - \mu}{4/3}\right| \leq \frac{2}{4/3}\right) \\ &= \mathbb{P}(|Z| \leq 1.5) \\ &= \mathbb{P}(-1.5 \leq Z \leq 1.5) \\ &= 0.9332 - 0.0668 \\ &= 0.8664 \end{aligned}$$

- (b) If she would like the sample mean to be within 1 square inch of the population mean with probability 0.90, how many trees must she measure in order to ensure this degree of accuracy?

Here we want $\mathbb{P}(|\bar{Y} - \mu| \leq 1) = 0.90$. In this case the standard deviation is $4/\sqrt{n}$, where n is unknown. Dividing by this, we get:

$$\begin{aligned} 0.90 = \mathbb{P}(|\bar{Y} - \mu| \leq 1) &= \mathbb{P}\left(\left|\frac{\bar{Y} - \mu}{4/\sqrt{n}}\right| \leq \frac{1}{4/\sqrt{n}}\right) \\ &= \mathbb{P}(|Z| \leq \sqrt{n}/4) \end{aligned}$$

By symmetry, we have $\mathbb{P}(Z \leq -\sqrt{n}/4) = 0.05$. Looking up the Z value on the z table, we get a z value of -1.64 (taking the one for 0.0495 instead of 0.0505; either one will do.) Thus we have:

$$\begin{aligned} \sqrt{n}/4 &= 1.64 \\ \sqrt{n} &= 6.56 \\ n &= 43.03 \end{aligned}$$

So she needs to measure 44 trees (rounding this up to the nearest integer).

6. There are 3 boxes on a table, containing 0, θ , and $\theta + 1$ jellybeans. Each of n people opens a box uniformly at random and takes the amount of jellybeans in the box. (The boxes are reset after each person takes their turn.) Let X_1, \dots, X_n be the number of jellybeans taken by each of the n people.

(a) Show $\hat{\theta} = \bar{X} = (1/n) \sum_{i=1}^n X_i$ is a biased estimator for θ .

$$\begin{aligned} \mathbb{E}(\hat{\theta}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n \left(0 \cdot \frac{1}{3} + \theta \cdot \frac{1}{3} + (\theta + 1) \cdot \frac{1}{3}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{3}(2\theta + 1) \\ &= \frac{2}{3}\theta + \frac{1}{3} \end{aligned}$$

This is not equal to θ (unless $\theta = 1$), thus this is a biased estimator.

(b) Based on the above, how can we modify $\hat{\theta}$ to convert it into an unbiased estimator.

If we look at the expression for θ above, we can turn it into θ by first subtracting $1/3$ and then multiplying by $3/2$. This gives us our unbiased estimator:

$$\hat{\theta}_{\text{unbiased}} = \frac{3}{2} \left(\bar{X} - \frac{1}{3} \right)$$

7. Suppose that the number of minutes late a RIPTA bus arrives is uniformly distributed on an interval $[0, 12]$. Suppose you measured the delay of a RIPTA bus 64 times and computed the sample mean \bar{Y} . What is the probability that the sample mean is between 5 and 7 minutes?

The mean and variance of a Uniform $[0, 12]$ random variable are 6 and $(12-0)^2/12 = 12$. By the central limit theorem, since the number of samples is large (≥ 30), we can assume that \bar{Y} is (approximately) normally distributed with mean 6 and standard deviation $\sigma/\sqrt{n} = \sqrt{12}/\sqrt{64} = \sqrt{12}/8$. The probability that the sample mean is

between 5 and 7 minutes is:

$$\begin{aligned}\mathbb{P}(|\bar{Y} - 6| \leq 1) &= \mathbb{P}\left(\left|\frac{\bar{Y} - 6}{\sqrt{12}/8}\right| \leq \frac{1}{\sqrt{12}/8}\right) \\ &= \mathbb{P}(|Z| \leq 8/\sqrt{12}) \\ &= \mathbb{P}(|Z| \leq 2.31) \\ &= \mathbb{P}(-2.31 \leq Z \leq 2.31) \\ &= 0.9896 - 0.0104 \\ &= 0.9792\end{aligned}$$