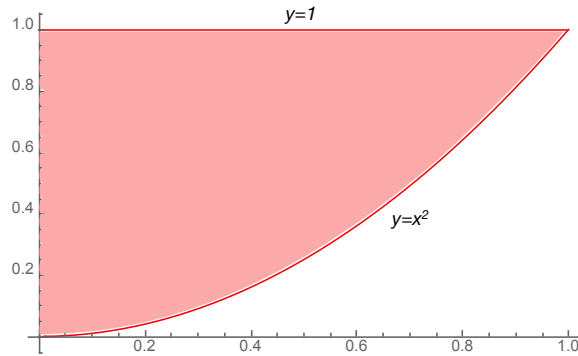1. (10 points) Let X and Y be continuous random variables with joint probability density given by:

$$f(x, y) = \begin{cases} cx & x \geq 0, 0 \leq x^2 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of c for which $f(x, y)$ is a valid joint probability density function?

First step is to draw a picture of the region, then integrate the joint density over the region and set the integral equal to 1.

$$1 = \int_0^1 \int_0^{\sqrt{y}} cx\, dx\, dy$$
$$= c \int_0^1 \frac{x^2}{2} \Big|_0^{\sqrt{y}} dy$$
$$= \frac{c}{2} \int_0^1 y\, dy$$
$$= \frac{c}{2} \frac{y^2}{2} \Big|_0^1$$
$$= \frac{c}{4}$$

From this we conclude $c = 4$.

(b) Find the expected value of X.

First we find the marginal density of X by integrating the joint density over y.

$$f_X(x) = \int_{x^2}^1 4x\, dy$$
$$= 4xy \Big|_{x^2}^1$$
$$= 4x(1 - x^2)$$
$$= 4(x - x^3)$$

1

The bounds on the marginal density are from 0 to 1. For the expected value,

$$\mathbb{E}(X) = \int_0^1 x4(x - x^3)dx$$

$$= 4\int_0^1 (x^2 - x^4)dx$$

$$= 4\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_0^1$$

$$= 4\left(\frac{1}{3} - \frac{1}{5}\right)$$

$$= \frac{8}{15}$$

(c) Find the conditional density of X given $Y = y$. Do not forget to include appropriate bounds for the density function.

First we need the marginal density of Y.

$$f_Y(y) = \int_0^{\sqrt{y}} 4xdx$$

$$= 2x^2\Big|_0^{\sqrt{y}}$$

$$= 2y$$

The bounds on the marginal density are from 0 to 1. For the conditional density, we divide the joint density by the marignal density.

$$f(x|Y = y) = \frac{4x}{2y}$$

$$= 2\frac{x}{y}$$

If $Y = y$, then X can only range from 0 to $\sqrt{y}$, thus we have:

$$f(x|Y = y) = \begin{cases} 2\dfrac{x}{y} & 0 \leq x \leq \sqrt{y} \\ 0 & \text{otherwise} \end{cases}$$

2

2. (10 points) You are interested in the proportion p of registered voters in Pennsylvania who prefer Clinton. You take a sample of n voters from the population. Let Y be the number of voters in your sample who prefer Clinton. We showed in class that $Y \sim \text{Binomial}(n, p)$ and that $\hat{p} = \frac{Y}{n}$ is an unbiased estimator for p. The standard estimator for the variance of Y is:

$$\hat{V} = n\hat{p}(1 - \hat{p}) = n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$$

(a) Show that $\hat{V}$ is a biased estimator for the variance of Y.

$$\mathbb{E}(\hat{V}) = \mathbb{E}\left[n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)\right]$$
$$= \mathbb{E}\left[Y\left(1 - \frac{Y}{n}\right)\right]$$
$$= \mathbb{E}(Y) - \frac{1}{n}\mathbb{E}(Y^2)$$

By the Magic Variance Formula,

$$\mathbb{E}(Y^2) = Var(Y) + [\mathbb{E}(Y)]^2$$

Substituting this in, and using the known expected value and variance of a binomial random variable,

$$\mathbb{E}(\hat{V}) = \mathbb{E}(Y) - \frac{1}{n}(Var(Y) + [\mathbb{E}(Y)]^2)$$
$$= np - \frac{1}{n}[np(1-p) + (np)^2]$$
$$= np - p(1-p) - np^2$$
$$= p(n - (1-p) - np)$$
$$= p[n(1-p) - (1-p)]$$
$$= p[(n-1)(1-p)]$$
$$= (n-1)p(1-p)$$

Since this is not equal to $Var(Y) = np(1-p)$, this is biased.

(b) Modify $\hat{V}$ so that you obtain an unbiased estimator for the variance of Y.

If we multiply $\hat{V}$ by $\frac{n}{n-1}$, we get an unbiased estimator.

3. (10 points) Suppose you choose a point uniformly at random from the interval $[0, 1]$. After you do so, your friend chooses a point uniformly at random between 0 and your point.

(a) Find the joint probability density function of the two points. Do not forget to include appropriate bounds for the density function.

Let X be your point, and Y be your friend's point. Then $X \sim$ Uniform$(0, 1)$. Y, conditioned on $X = x$, is a uniform random variable from 0 to x. Thus we have a marginal density for X:

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

And we have a conditional density for Y given $X = x$.

$$f(y|X = x) = \begin{cases} \dfrac{1}{x} & 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

Since the joint density is the product of the conditional and the marginal densities, we have:

$$f(x, y) = \begin{cases} \dfrac{1}{x} & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

You could also reason your way to this result by drawing a picture and coming up with a density with the desired properties.

(b) What is the expected value of the product of the two points?

$$\begin{aligned} \mathbb{E}(XY) &= \int_0^1 \int_0^x xy \frac{1}{x} dy dx \\ &= \int_0^1 \int_0^x y dy dx \\ &= \int_0^1 \frac{y^2}{2} \Big|_0^x dx \\ &= \frac{1}{2} \int_0^1 \frac{x^2}{d} x \\ &= \frac{1}{2} \frac{x^3}{3} \Big|_0^1 \\ &= \frac{1}{6} \end{aligned}$$

4

4. (10 points) During study breaks from APMA 1650, you play Pokémon GO. The average number of Pokémon a player catches in a 30-minute period is 6. Assume that Pokémon are caught independently from each other.

(a) Suppose you just caught a Pokémon. Modeling this with an appropriate probability distribution, what is the probability that it will take less than 10 minutes for you to catch your next Pokémon?

Let X be the time it takes to catch a new Pokémon. Since the average number of Pokémon caught in one minute is $6/30 = 1/5$, let $X \sim$ Exponential$(1/5)$. Then

$$
\begin{aligned}
\mathbb{P}(0 \leq X \leq 10) &= \int_0^{10} \frac{1}{5} e^{-(1/5)x} dx \\
&= -e^{-(1/5)x} \Big|_0^1 \\
&= -e^{-10/5} - (-1) \\
&= 1 - e^{-2}
\end{aligned}
$$

(b) Being a good statistician, you decide to play Pokémon Go long enough to catch 100 more Pokémon. You measure the amount of time it takes to catch each of the 100 Pokémon and compute the sample mean. What is the probability that this sample mean is between 4 and 6 minutes?

By the Central Limit Theorem, the sample $\bar{X}$ is a normal distirbution. The population is exponential with mean 5 and variance 5^2, so the population standard deviation is 5. (This can all be looked up in the reference table). Then the mean of $\bar{X}$ is 5 and the standard deviation of $\bar{X}$ is $5/\sqrt{100} = 1/2$. Thus $\bar{X} \sim$ Normal(5, 1/2).

$$
\begin{aligned}
\mathbb{P}(4 < \bar{X} < 6) &= \mathbb{P}\left(\frac{4-5}{1/2} < Z < \frac{5-6}{1/2} \right) \\
&= \mathbb{P}(-2 < Z < 2) \\
&= 0.9722 - 0.0228
\end{aligned}
$$

You could also have used the 68-95-99.7 rule since we want the probability of being within 2 standard deviations of the mean.

5. (10 points) Despite being sleep-deprived from all the Pokémon you caught in the previous problem, you decide to continue playing Pokémon GO. Once again, the average number of Pokémon a player catches in a 30-minute period is 6, and Pokémon are caught independently from each other.

(a) Modeling this with an appropriate probability distribution, what is the probability that you catch 1 or fewer Pokémon in a 10-minute period?

Let Y be the number of Pokémon caught in a 10-minute period. The average number of Pokémon caught in a 10-minute period is $6/3 = 2$, so let $Y \sim \text{Poisson}(2)$. Then

$$\mathbb{P}(Y \leq 1) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1)$$
$$= \frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!}$$
$$= 3e^{-2}$$

(b) Your friend, who also plays Pokémon GO, claims to have caught more than 4 (i.e. 5 or more) Pokémon in a 10-minute period. Can you be at least 75% certain that your friend is lying? Justify your answer mathematically. This does not require a calculator.

By Chebyshev's Inequality, since the mean and variance of Y is 2,

$$\mathbb{P}(Y \geq 5) \leq \mathbb{P}(|Y - 2| \geq 3)$$
$$\leq \frac{Var(Y)}{3^2}$$
$$= \frac{2}{9} < \frac{1}{4} = 0.25$$

Since the probability that $Y \leq 5$ is less than or equal to 0.25, you can be at least 75% certain that your friend is lying.

If you use Markov's Inequality, you get:

$$\mathbb{P}(Y \geq 5) \leq \frac{\mathbb{E}(Y)}{5} = \frac{2}{5} = 0.40$$

which is unfortunately not good enough. The moral of the story is that Markov's inequality is not very good, so in cases where you know the variance (like you do here), you want to use Chebyshev's Inequality.