PSTAT 120A, Summer 2022: Practice Problems 10: Final Review, Part II

Week 2

Conceptual Review

(a) Review the conceptual questions from the previous Discussion Worksheets!

Problem 1: Sums and Money

When run once, a particular random number generator generates a number from the set $\{0, 1, 2, 4\}$ with the following probabilities:

number	0	1	2	4
prob.	$1/2$	$1/4$	$1/12$	$1/6$

Suppose that if the number generated is even, I gain \$1; otherwise I lose \$1. Let W_i denote the amount of money I gain/loose after one run of this generator.

(a) Find $p_{W_i}(w)$, the p.m.f. of W_i.
(b) Compute $\mathbb{E}[W_i]$ and $\text{Var}(W_i)$.
(c) Now suppose I run this generator 100 times, each time winning/losing money according to the aforementioned scheme. What is the probability that I earn more than \$55 at the end of these 100 runs?

Problem 2: Faces of a Die

A fair die is rolled three times; let X denote the number of distinct faces that appear.

(a) Find $p_X(x)$, the p.m.f. of X
(b) Compute $\mathbb{E}[X]$.

Problem 3: Conditioning

Suppose (X, Y) is a continuous bivariate random vector with joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{y} & \text{if } 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that $f_{X,Y}(x, y)$ is a valid joint p.d.f.
(b) Identify the conditional distribution of $(X \mid Y)$.
(c) Identify the marginal distribution of Y.
(d) Compute $\mathbb{E}[X]$ and $\text{Var}(X)$.

Problem 4: The $\delta-$Method

Though the LOTUS is a very powerful tool, the integrals that it requires us to evaluate aren't always the nicest (nor are they always solvable in terms of elementary antiderivatives!) As such, the so-called **$\delta-$method** was introduced as a way of approximating expectations and variances of functions of random variables. The idea is as follows: we can utilize Taylor Series Expansions to derive approximations to quantities like $\text{Var}[f(X)]$. Let X be a random variable with mean μ and variance σ^2.

(a) Write down the first-order Taylor Series expansion of $f(x)$ about the point μ.

Recall: "first-order" means the second term in the sum

Solution:

$$f(x) \approx f(\mu) + (x - \mu) f'(x)$$

(b) Use part (a) to derive the approximation

$$\text{Var}[f(X)] \approx \sigma^2 \cdot [f'(\mu)]^2$$

Solution:

$$\begin{aligned} f(X) &\approx f(\mu) + (X - \mu) f'(x) \\ \text{Var}[f(X)] &\approx \text{Var}\left[f(\mu) + (X - \mu) f'(\mu)\right] \\ &= \text{Var}[(X - \mu) f'(\mu)] \\ &= \text{Var}(X - \mu) \cdot [f'(\mu)]^2 \\ &= \text{Var}(X) \cdot [f'(\mu)]^2 = \sigma^2 \cdot [f'(\mu)]^2 \end{aligned}$$

Problem 5: MGF

Let X be a random variable with Moment-Generating Function (MGF) given by

$$M_X(t) = \exp\left\{2t + \frac{3}{2}t^2\right\}$$

Compute $\mathbb{P}(X \geq 4)$.

Solution: We recognize $X \sim \mathcal{N}(2, 3)$ meaning

$$\mathbb{P}(X \geq 4) = 1 - \mathbb{P}\left(\frac{X - 4}{\sqrt{3}} < \frac{4 - 2}{\sqrt{3}}\right) = 1 - \Phi\left(\frac{2}{\sqrt{3}}\right)$$

Problem 6: Convergence

If $X_n \sim \text{Gamma}(n, n)$, show that $X_n \xrightarrow{\text{p}} 1$.

Solution: There are many ways to do this. The first is to note that

$$\text{Var}(X_n) = \frac{n}{n^2} = \frac{1}{n} \to 0 \text{ as } n \to \infty$$

meaning, by Problem 3 in Worksheet 9,

$$X_n \xrightarrow{\text{p}} \mathbb{E}[X_n] = 1$$

Alternatively, note that if $X_i \sim \text{Gamma}(1, 1)$ then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, n) \stackrel{\text{d}}{=} X_n$$

meaning, by the LLN,

$$X_n \xrightarrow{\text{p}} \mathbb{E}[X_1] = 1$$

Problem 7: The Flash vs. The Reverse-Flash

The time it takes the *Flash* to complete an obstacle course is an exponentially distributed random variable with mean 30 minutes. The time it takes the *Reverse-Flash* to complete the same obstacle course is an exponentially distributed random variable with mean 32 minutes. If the *Flash* and *Reverse-Flash* start the obstacle course at the same time, and finish independently of each other, what is the probability that the two finish the course within 2 minutes of each other?

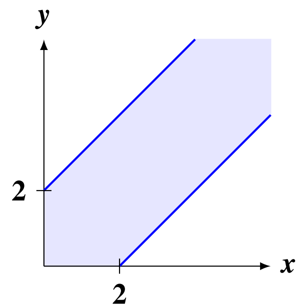
Solution: Let X denote the time it takes the *Flash* to complete the course, and let Y denote the time it takes the *Reverse-Flash* to complete the course. Using the rate parameterization of the exponential distribution, we then have

$$f_X(x) = \lambda e^{-\lambda \cdot x} \cdot \mathbb{1}_{\{x \geq 0\}}; \qquad f_Y(y) = \mu e^{-\mu \cdot y} \cdot \mathbb{1}_{\{y \geq 0\}}$$

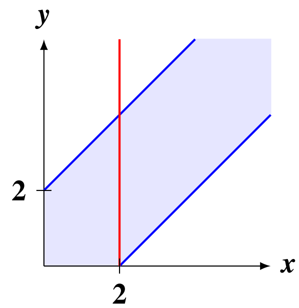
where $\lambda = 1/30$ and $\mu = 1/32$. Additionally, since $X \perp Y$ we know that the joint density function of (X, Y) is

$$f_{X,Y}(x, y) = \begin{cases} \lambda\mu e^{-\lambda \cdot x} e^{-\mu \cdot y} & \text{if } x \geq 0,\ y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We seek $\mathbb{P}(|X - Y| < 2)$; let's sketch the region of integration.



Either order of integration is fine, and will result in the same amount of work. I shall demonstrate the dy dx order below: note that we have to split the region into two subregions:



$$
\begin{aligned}
\mathbb{P}(|X-Y|<2) &= \int_0^2 \int_0^{x+2} \lambda\mu e^{-\lambda\cdot x} e^{-\mu\cdot y} \ \mathrm{d}y \ \mathrm{d}x + \int_2^\infty \int_{x-2}^{x+2} \lambda\mu e^{-\lambda\cdot x} e^{-\mu\cdot y} \ \mathrm{d}y \ \mathrm{d}x \\
&= \int_0^2 \lambda\mu e^{-\lambda x} \int_0^{x+2} \mu e^{-\mu y} \ \mathrm{d}y \ \mathrm{d}x + \int_2^\infty \lambda e^{-\lambda x} \int_{x-2}^{x+2} \mu e^{-\mu y} \ \mathrm{d}y \ \mathrm{d}x \\
&= \int_0^2 \lambda e^{-\lambda x}(1-e^{-\mu(x+2)}) \ \mathrm{d}x + \int_2^\infty \lambda e^{-\lambda x}(e^{-\mu(x-2)} - e^{-\mu(x+2)}) \ \mathrm{d}x \\
&= \int_0^2 \lambda e^{-\lambda x} \ \mathrm{d}x - e^{-2\mu} \int_0^2 \lambda e^{-(\lambda+\mu)x} \ \mathrm{d}x + e^{2\mu} \int_2^\infty \lambda e^{-(\lambda+\mu)x} \ \mathrm{d}x - e^{-2\mu} \int_2^\infty \lambda e^{-(\lambda+\mu)x} \ \mathrm{d}x \\
&= (1-e^{-2\lambda}) - \frac{\lambda e^{-2\mu}}{\lambda+\mu}(1-e^{-2(\lambda+\mu)}) + \frac{\lambda e^{2\mu}}{\lambda+\mu} e^{-2(\lambda+\mu)} - \frac{\lambda e^{-2\mu}}{\lambda+\mu} e^{-2(\lambda+\mu)} \\
&= (1-e^{-2\lambda}) - \frac{\lambda e^{-2\mu}}{\lambda+\mu} + \cancel{\frac{\lambda e^{-2\mu}}{\lambda+\mu} e^{-2(\lambda+\mu)}} + \frac{\lambda e^{2\mu}}{\lambda+\mu} e^{-2(\lambda+\mu)} - \cancel{\frac{\lambda e^{-2\mu}}{\lambda+\mu} e^{-2(\lambda+\mu)}} \\
&= \boxed{(1-e^{-2\lambda}) + \frac{\lambda}{\lambda+\mu}\left(e^{-2\lambda} - e^{-2\mu}\right)}
\end{aligned}
$$

or, plugging in $\lambda = 1/30$ and $\mu = 1/32$,

$$
\boxed{\left(1-e^{-1/15}\right) + \frac{\frac{1}{30}}{\frac{1}{30}+\frac{1}{32}}\left(e^{-1/15} - e^{-1/16}\right) = \left(1-e^{-1/15}\right) + \frac{30\cdot 32}{30+32}\left(e^{-1/15} - e^{-1/16}\right) \approx 0.06248}
$$

By the Way: The trick to evaluating these integrals is to continually relate them to the CDF of various Exponential distributions.

Problem 8: Roulette

A standard American roulette wheel consists of 38 equally spaced slots; 18 are red, 18 are black, and 2 are green. During a game of Roulette, the wheel is spun and a ball is released so that by the time the wheel stops spinning the ball will have landed into one of the slots. Assume the wheel is "fair" in the sense that the ball is equally likely to have landed in any of the slots.

Suppose n games of roulette are played, independently of each other. Let X_R denote the number of times the ball lands in a red slot, X_B denote the number of times the ball lands in a black slot, and X_G denote the number of times the ball lands in a green slot.

(a) If $n = 6$, what is the probability that the ball landed on red thrice, black twice, and green once?

Solution: Note that

$$(X_R, X_B, X_G) \sim \text{Multi}\left(n,\ 3,\ \frac{18}{38},\ \frac{18}{38},\ \frac{2}{38}\right)$$

We can therefore use the results of Homework 8 to conclude that

$$p_{X_R, X_B, X_G}(3, 2, 1) = \binom{6}{3,2,1} \cdot \left(\frac{18}{38}\right)^3 \left(\frac{18}{38}\right)^2 \left(\frac{2}{38}\right)^1$$

(b) Returning to the case of a general n, what is the correlation between the number of times the ball lands on Red and the number of times it lands on black?

Solution: From Homework 8, we know that

$$\text{Cov}(X_R, X_B) = -n\left(\frac{18}{38}\right)\left(\frac{18}{38}\right)$$

Additionally, we know that

$$X_R \sim \text{Bin}\left(n,\ \frac{18}{38}\right)$$

$$X_B \sim \text{Bin}\left(n,\ \frac{18}{38}\right)$$

meaning

$$\text{SD}(X_R) \cdot SD(X_B) = n\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)$$

and so

$$\text{Corr}(X_R, X_b) = \frac{\text{Cov}(X_R, X_B)}{\text{SD}(X_R)\cdot \text{SD}(X_B)} = \frac{-n\left(\frac{18}{38}\right)\left(\frac{18}{38}\right)}{n\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)} = -\frac{18}{20}$$

(c) What is the correlation between the number of times the ball lands on red and the number of times it lands on green?

Solution: We use a very similar procedure as we did in part (b), just with different numbers:

$$\text{Cov}(X_R, X_B) = -n\left(\frac{18}{38}\right)\left(\frac{2}{38}\right)$$

Additionally, we know that

$$X_R \sim \text{Bin}\left(n, \frac{18}{38}\right)$$
$$X_G \sim \text{Bin}\left(n, \frac{2}{38}\right)$$

meaning

$$\text{SD}(X_R) \cdot SD(X_G) = n\sqrt{\left(\frac{18}{38}\right)\left(\frac{20}{38}\right)\left(\frac{2}{38}\right)\left(\frac{36}{38}\right)} = n \cdot \frac{18\sqrt{5}}{361}$$

and so

$$\text{Corr}(X_R, X_b) = \frac{\text{Cov}(X_R, X_B)}{\text{SD}(X_R) \cdot \text{SD}(X_B)} = \frac{-n\left(\frac{18}{38}\right)\left(\frac{2}{38}\right)}{n\frac{18\sqrt{5}}{361}} = -\frac{\sqrt{5}}{10}$$