Problem Set 2 Solutions (Main classes 5-7)

Izzy Grosof

October 3, 2025 – Due October 10th (Friday)

1. Let X_1, X_2 be i.i.d. continuous random variables, both having CDF $F_X(x)$ and density $f_X(x)$. Let $M = \min(X_1, X_2)$ be the minimum of the two random variables.

   (a) Prove that the CDF of M is:

   $$F_M(x) = 1 - (1 - F_X(x))^2$$

   Solution. We want to find $F_M(x) = P(M \leq x)$. First, note that $P(M \leq x) = 1 - P(M > x)$. Next, note that the event $M > x$ is equivalent to the event $X_1 > x \& X_2 > x$. Next, by independence, note that $P(X_1 > x \& X_2 > x) = P(X_1 > x)P(X_2 > x)$. Next, note hat $P(X_1 > x) = P(X_2 > x) = 1 - F_X(x)$.

   Putting it all together, we find that

   $$F_M(x) = 1 - (1 - F_X(x))^2. \qquad \square$$

   (b) What is the density $f_M(x)$ of M in terms of the density $f_X(x)$ and CDF $F_X(x)$ of X_1 and X_2?

   Solution.

   $$f_M(x) = \frac{d}{dx}F_M(x) = \frac{d}{dx}1 - (1 - F_X(x))^2 = 2f_X(x)(1 - F_X(x))$$

   $\square$

   (c) Suppose that $X_1, X_2 \sim Exp(\lambda)$. What is the CDF $F_M(x)$ of M?
   Note: Info on the exponential random variable is available in IPC 7.2.

   Solution.

   $$F_M(x) = 1 - (1 - F_X(x))^2 = 1 - (e^{-\lambda x})^2 = 1 - e^{-2\lambda x}$$

   $\square$

   (d) Is M an exponential random variable? Why?

   Solution. Yes, M is an exponential random variable. Specifically, $M \sim Exp(2\lambda)$, $\square$

2. You've just shown upon to a train platform, where trains arrive every 15 minutes. When you arrive let T denote the time single the last train. In this case, $T \sim Unif(0, 15)$ minutes.

   Let S denote the number of people who you see at the train platform, when you arrive. People have been arriving to the train platform single the last train arrived at a rate of 2 people per minute, on average. As a result, $S \sim Poisson(2T)$.

   Note: Info on the Poisson distribution can be found in IPC 3.2.4. Info on the Uniform distribution can be found in IPC 7.2. Note also that $\mathbb{E}[Poisson(\lambda)] = \lambda$ and $\text{Var}[Poisson(\lambda)] = \lambda$.

(a) What is the mean number of people you see at the train platform, $\mathbb{E}[S]$?

Solution.

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid T]]$$
$$[S \mid T = t] \sim Poisson(2t)$$
$$\mathbb{E}[S \mid T] = 2T$$
$$\mathbb{E}[\mathbb{E}[S \mid T]] = 2\mathbb{E}[T] = 2\mathbb{E}[Uniform(0, 15)] = 15 \qquad \square$$

(b) What is the variance of the number of people you see at the train platform, $Var[S]$?

Solution. We'll calculate $\mathbb{E}[S^2]$ first.

$$[S^2 \mid T = t] \sim [Poisson(2t)^2]$$
$$\mathbb{E}[S^2 \mid T = t] = \mathbb{E}[Poisson(2t)^2] = Var[Poisson(2t)] + \mathbb{E}[Poisson(2t)]^2 = 2t + (2t)^2 = 2t + 4t^2$$
$$\mathbb{E}[S^2 \mid T] = 2t + 4T^2$$
$$\mathbb{E}[S^2] = \mathbb{E}[\mathbb{E}[S^2 \mid T]] = \mathbb{E}[2T + 4T^2] = 15 + 4\mathbb{E}[Uniform(0, 15)^2]$$

Now, we calculate the second moment of $Uniform(0, 15)$:

$$\mathbb{E}[Uniform(0, 15)^2] = \int_{t=0}^{1} 5t^2 \frac{1}{15} = \frac{1}{15}[t^3/3]_{t=0}^{1} 5 = \frac{1}{15} 15^3/3 = 15^2/3 = 225/3 = 75$$

Now, we can finish up the problem:

$$\mathbb{E}[S^2] = 15 + 4\mathbb{E}[Uniform(0, 15)^2] = 315$$
$$Var[S] = \mathbb{E}[S^2] - \mathbb{E}[S]^2 = 315 - 15^2 = 315 - 225 = 90 \qquad \square$$

3. You've been selected as a contestant on Jeopardy, a game show. You'll play in the first round, where you might win or lose. Each round, if you win, you go on to the next round, and play again. Otherwise, you're done.

You're pretty good at Jeopardy, so you have an 80% probability of winning each round, independent of all other rounds.

Let R denote the total number of rounds you play.

(a) What's the standard name for the distribution of the random variable R? **Note:** Standard names for discrete distributions can be found in the textbook, in chapter 3.2.

Solution. This is a geometric distribution: We have an i.i.d. chance of continuing or stopping. The parameter inside a geometric distribution is the stopping probability, which is 0.2 in this case. So $R \sim Geometric(0.2)$. $\qquad \square$

(b) You've won your first 5 rounds. Let Q denote the number of rounds you will play after these first five. In terms of R, we can write Q as follows:

$$Q = [R - 5 \mid R > 5]$$

What is the PMF of Q, $\mathbb{P}\{Q = i\}$ for each possible value i?

Solution. Two useful facts: For R to be exactly x, we need x-1 wins and 1 loss, in that order, so $\mathbb{P}\{R = x\} = 0.8^{x-1}0.2$. For R to be greater than x, we need the first x rounds to be wins, so $\mathbb{P}\{R > x\} = 0.8^x$.

$$\mathbb{P}\{Q = i\} = \mathbb{P}\{R - 5 = i \mid R > 5\} = \frac{\mathbb{P}\{R = i + 5\}}{\mathbb{P}\{R > 5\}} = \frac{0.8^{i+5-1} \cdot 0.2}{0.8^5} = 0.8^{i-1} \cdot 0.2$$

This matches $\mathbb{P}\{R = i\}$, so $Q \sim R$. This is a proof that a geometric distribution is memoryless. $\square$

(c) What's the standard name for the distribution of the random variable Q?

Solution. Geometric(0.2), same as R. $\square$

4. Isabella is playing chess online, and keeping track of her rating. Her initial rating is 1000. Each day, her rating changes by an amount D_i, where $D_i \sim Uniform(-30, 30)$ is uniformly distributed between -30 and +30 rating. The rating changes D_i are i.i.d.

After 100 days, approximate Isabella's new rating with a Normal distribution. What's the approximate probability that her new rating is at least 1200?

Solution. Let $S = \sum_{i=1}^1 00D_i$ be the total change in Isabella's rating. We can approximate S by a Normal distribution using the Central Limit Theorem:

$$S \simeq Normal(100\mathbb{E}[D], \sqrt{100\mathrm{Var}[D]}$$

We have $\mathbb{E}[D] = 0$, and we can calculate $\mathrm{Var}[D]$ as follows:

$$\mathrm{Var}[D] = \mathbb{E}[D^2] - \mathbb{E}[D]^2 = \mathbb{E}[D^2]$$
$$= \int_{d=-30}^{30} \frac{1}{60}d^2 dd = \frac{1}{60}[d^3/3]_{d=-30}^{30} = \frac{1}{60}\frac{30^3 + 30^3}{3} = \frac{1}{2}\frac{2 \cdot 30^2}{3} = \frac{900}{3} = 300$$

So the distribution of S is approximately:

$$S \simeq Normal(0, \sqrt{300 \cdot 100} = Normal(0, 100\sqrt{3})$$

Now, we calculate the probability that Isabella's final rating is at least 1200:

$$\mathbb{P}\{1000 + S \geq 1200\} = \mathbb{P}\{S \geq 200\} \approx \mathbb{P}\{Normal(0, 100\sqrt{3}) \geq 200\}$$
$$= \mathbb{P}\{Normal(0, 1) \geq \frac{2}{\sqrt{3}}\} = 1 - F_{N(0,1)}(2/\sqrt{3})$$

This is fine as a solution – no further work is needed. However, we can also plug the final number into a Normal CDF calculator to get an approximate probability.

$$1 - F_{N(0,1)}(2/\sqrt{3}) \approx 12.41\% \qquad \square$$

5. This is a programming problem. Write a program in Python, as either a standalone file or as a Jupyter notebook. Include your answers in your solution directly, and also submit the code you write, either in the same document or as a separate upload.

Simulate the following situation:

Moss is flipping a fair coin (equally likely heads or tails, i.i.d.), until the result is heads.

For each coin flip, Moss waits $Exp(1)$ seconds before flipping the coin. The first coin flip happens at time $Exp(1)$, the second coin flip, if there is one, happens $Exp(1)$ time later, and so on.

Let T denote the time when the final coin, the one whose result is heads, is flipped.

Generate 10^6 samples of the random variable T.

(a) What are $\mathbb{P}\{T \geq 1\}, \mathbb{P}\{T \geq 2\}, \mathbb{P}\{T \geq 3\}, \mathbb{P}\{T \geq 4\}$?

Solution. See the solution program at the end of the document. I used `random`, it's fine if you used `numpy`.

My results were:

$$\mathbb{P}\{T \geq 1\} \approx 0.6067, \mathbb{P}\{T \geq 2\} \approx 0.3673, \mathbb{P}\{T \geq 3\} \approx 0.2231, \mathbb{P}\{T \geq 4\} \approx 0.1355$$

□

(b) Moss thinks that T is exponentially distributed, and wants to use your data to see if this might be true. If a random variable $X \sim Exp(\lambda)$ is exponentially distributed with rate λ, then for any threshold x, the following fraction will be constant:

$$\frac{\mathbb{P}\{X \geq x + 1\}}{\mathbb{P}\{X \geq x\}} = e^{-\lambda}$$

If X is not exponentially distributed, then the above fraction won't have a constant value across different thresholds x.

Use your data from (a) to check whether the above fraction is near-constant for the random variable T across thresholds $x = \{1, 2, 3\}$.

Does T seem exponentially distributed? If so, what is its rate parameter λ?

Solution. I found the following ratios:

$$\frac{\mathbb{P}\{T \geq 2\}}{\mathbb{P}\{T \geq 1\}} \approx 0.6054, \frac{\mathbb{P}\{T \geq 3\}}{\mathbb{P}\{T \geq 2\}} \approx 0.6074, \frac{\mathbb{P}\{T \geq 4\}}{\mathbb{P}\{T \geq 3\}} \approx 0.6074.$$

These are all similar, so T is probably exponentially distributed. With an average ratio of about 0.6068, I estimate that $\lambda \approx -\ln(0.6068) = 0.4996$. □