This week on the problem set you will get practice with continuous random variables. Especially challenging questions, or questions that are not appropriate for an exam, are indicated with one or more asterisks.

1. From the textbook, chapter 2, problems 1, 2.
2. From the supplementary problems, chapter 2, problem 2.
3. Which of the following functions can be a probability density function for some continuous random variable. Explain your reasoning.

   (a)
   $$f(t) = \begin{cases} e^t + 4t - e, & \text{for } 0 \le t \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

   (b)
   $$f(t) = \begin{cases} e^{-t}, & \text{for } t \ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

   (c)
   $$f(t) = e^{\sin(t)}, \text{ for all } t \in \mathbb{R}.$$

   (d)
   $$f(t) = \begin{cases} 0, & \text{for } t < 0 \\ \frac{1}{2}, & \text{for } 0 \le t \le 1 \\ 0, & \text{for } 1 < t < 2 \\ \frac{1}{2}, & \text{for } 2 \le t \le 3 \\ 0, & \text{for } 3 < t. \end{cases}$$

   Solution:

   1) No, f is continuous and we have $f(0) = 1 - e < 0$, and so f is negative on an interval. Probability density function cannot be negative at any point.
   2) Yes. This is the density of the parameter 1 Exponential random variable.
   3) No. For $0 \le t \le \pi$ we have $\sin(t) \ge 1$ and so for such ts we have $f(t) \ge 1$. Then $\int_0^\pi f(t) \ge \pi \cdot 1 = \pi$. This would mean that the probability that the random variable is between 0 and π is larger than π which is impossible.
   4) Yes. This function f satisfies both $f(t) \ge 0$ and $\int_{-\infty}^{\infty} f(t)\ dt = 1$. This suffices for the arguments, but if you want to have an example of such random variable, take Y to be a uniform random variable on $[0, 2]$. If $0 \le Y \le 1$ define X to be equal to Y and if $1 < Y \le 2$ define X to be equal to $Y + 1$. Then X is a continuous random variable and it's probability density function is exactly f from this part of the problem.

4. Continuous random variable X has probability density function $f_X(t)$ such that

   $$f_X(t) = \begin{cases} C\sin(\pi t/n), & \text{for } 0 \le t \le n \\ 0, & \text{otherwise.} \end{cases}$$

   Find the value of the constant C. Then find $\mathbb{E}(X)$ and $\text{var}(X)$.

Solution: For this to be a probability density function we would need to have

$$\int_0^n C\sin(\pi t/n)\ dt = 1.$$

Using substitution $u = \pi t/n$ we have $dt = \frac{n}{\pi} du$ and the integral becomes

$$\frac{Cn}{\pi}\int_0^\pi \sin(u)\ du = 1.$$

Since $\int_0^\pi \sin(u)\ du = [-\cos(u)]_0^\pi = 2$ we have $C = \frac{\pi}{2n}$. Then

$$\mathbb{E}(X) = \frac{\pi}{2n}\int_0^n t\sin(\pi t/n)\ dt.$$

Again using $u = \pi t/n$ we get

$$\mathbb{E}(X) = \frac{n}{2\pi}\int_0^\pi u\sin(u)\ du = \frac{n}{2\pi}\left[-u\cos(u)\right]_0^\pi + \frac{n}{2\pi}\int_0^\pi \cos(u)\ du = \frac{n}{2} + \frac{n}{2\pi}[\sin(u)]_0^\pi = \frac{n}{2}.$$

Here we used integration by parts with u and $\sin(u)\ du$. To find variance proceed similarly, but use integration by parts twice.

$$\begin{aligned}\mathbb{E}(X^2) &= \frac{\pi}{2n}\int_0^n t^2\sin(\pi t/n)\ dt = \frac{n^2}{2\pi^2}\int_0^\pi u^2\sin(u)\ du \\ &= \frac{n^2}{2\pi^2}\left[-u^2\cos(u)\right]_0^\pi + \frac{n^2}{2\pi^2}\int_0^\pi 2u\cos(u)\ du = \frac{n^2}{2} + \frac{n^2}{\pi^2}\int_0^\pi u\cos(u)\ du \\ &= \frac{n^2}{2} + \frac{n^2}{\pi^2}\left[u\sin(u)\right]_0^\pi - \frac{n^2}{\pi^2}\int_0^\pi \sin(u)\ du = \frac{n^2}{2} + \frac{n^2}{\pi^2}\left[\cos(u)\right]_0^\pi = \frac{n^2}{2} - \frac{2n^2}{\pi^2}.\end{aligned}$$

Then

$$\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{n^2}{2} - \frac{2n^2}{\pi^2} - \frac{n^2}{4} = \frac{\pi^2 - 8}{4\pi^2}n^2.$$

5. Let $p > 0$ be a real number. Consider function f which has values $f(t) = Ct^{-p}$ for $t \geq 1$, and $f(t) = 0$ for $t < 1$. For which values of p we can select constant C so that this function f is a probability density function for some continuous random variable? For which values of p will this random variable have finite expectation? For which values of p will this random variable have finite variance?

Solution: The function f is non-negative everywhere. Therefore we only need to make sure that $\int_{-\infty}^\infty f(t)\ dt = 1$. This becomes

$$C\int_1^\infty t^{-p}\ dt.$$

The anti-derivative for $p \neq 1$ is $\frac{1}{1-p}t^{1-p}$, but for $0 < p < 1$ we have

$$C\int_1^\infty t^{-p}\ dt = \frac{C}{1-p}\left[t^{1-p}\right]_1^\infty = \infty,$$

for any value of C. Similar conclusion holds for $p = 1$ since the anti-derivative is $\log t$. However for $p > 1$ we have

$$C\int_1^\infty t^{-p}\ dt = \frac{C}{1-p}\left[t^{1-p}\right]_1^\infty = \frac{C}{p-1},$$

so this indeed is a probability density function for $C = p - 1$. The answer for the first question is thus $p > 1$.

To find out when this random variable has a finite expectation compute (of course only for $p > 1$)

$$\mathbb{E}(X) = (p-1)\int_1^\infty t \cdot t^{-p}\, dt = (p-1)\int_1^\infty t^{1-p}\, dt.$$

Now argue similarly as before: the anti-derivative for $p \neq 2$ is $\frac{1}{2-p}t^{2-p}$, and for $1 < p < 2$ this function converges to ∞ as $t \to \infty$. So for $1 < p < 2$ the expectation does not exist. The same is true for $p = 2$. However, for $p > 2$ the aforementioned integral is convergent so the expectation is finite. Therefore, the answer for the 2nd question is $p > 2$.

Using the same reasoning one can conclude that the variance is finite if and only if $p > 3$.

6. Let X be a positive continuous random variable with the probability density function $f(t)$. If $g(t) = tf(t)$ is also a probability density function of some random variable Y what is $\mathbb{E}(X)$? Express $\mathrm{var}(X)$ in terms of $\mathbb{E}(Y)$.

Solution: Since X is positive we see that $f(t) = 0$ for $t < 0$. Then

$$\mathbb{E}(X) = \int_0^\infty tf(t)\, dt = \int_0^\infty g(t)\, dt = 1,$$

since $g(t)$ is density of the random variable Y. Similarly

$$\mathbb{E}(X^2) = \int_0^\infty t^2 f(t)\, dt = \int_0^\infty tg(t)\, dt = \mathbb{E}(Y).$$

Then $\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(Y) - 1$.

7. You are operating a train. Ticket for this train costs \$10. The train is late to the destination T minutes, where T is an exponential random variable with parameter 1. If the train is more than 2 minutes and less than 4 minutes late then each customer gets half of the ticket refunded. If the train is more than 4 and less than 10 minutes late each customer gets the full price of the ticket refunded. If the train is more than 10 minutes late each customer gets the full price of the ticket refunded and in addition gets \$$n$. For what values of n would your expected profit be positive? For what values of n would you expect to earn at least \$5 per ticket.

Solution: For each customer, you can have either \$10 income, \$5 income, no income or loss or \$$n$ loss. These events happen when the train is less than 2 min late, between 2 and 4 min late, between 4 and 10 min late and more than 10min late respectively. The probabilities of these events are respectively

$$\int_0^2 e^{-t}\, dt = 1 - e^{-2},\quad \int_2^4 e^{-t}\, dt = e^{-2} - e^{-4},\quad \int_4^{10} e^{-t}\, dt = e^{-4} - e^{-10},\quad \int_{10}^\infty e^{-t}\, dt = e^{-10}.$$

You expected income is then

$$10(1 - e^{-2}) + 5(e^{-2} - e^{-4}) - ne^{-10}.$$

For this to be positive we need

$$n < 10e^{10} - 5e^8 - 5e^6 \approx 203342.72.$$

For the above expression to be larger than 5

$$n < 5e^{10} - 5e^8 - 5e^6 \approx 93210.395.$$

8. If X is a continuous random variable which attains only positive values (that is $X \geq 0$). Show that

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)\ dt.$$

Note that this also holds for discrete random variables, you can try to prove this formula for discrete random variables as well.

Solution: We have $\mathbb{P}(X \geq t) = \int_t^\infty f_X(s)\ ds$ and so

$$\begin{aligned}\int_0^\infty \mathbb{P}(X \geq t)\ dt = \int_0^\infty \int_t^\infty f_X(s)\ dsdt = \int_0^\infty \int_0^s f_X(s)\ dtds \\ = \int_0^\infty f_X(s) \int_0^s\ dtds = \int_0^\infty s f_X(s)\ ds = \mathbb{E}(X).\end{aligned}$$

If X is discrete then say it has values $0 \leq a_1 < a_2 < a_3 < \dots$ with probabilities $p_i = \mathbb{P}(X = a_i)$. Then for $0 \leq t \leq a_1$ we have

$$\mathbb{P}(X \geq t) = \sum_{i \geq 1} p_i \ \Rightarrow\ \int_0^{a_1} \mathbb{P}(X \geq t)\ dt = a_1 \sum_{i \geq 1} p_i.$$

For $a_1 < t \leq a_2$ we have

$$\mathbb{P}(X \geq t) = \sum_{i \geq 2} p_i \ \Rightarrow\ \int_{a_1}^{a_2} \mathbb{P}(X \geq t)\ dt = (a_2 - a_1) \sum_{i \geq 2} p_i.$$

In general we get for $j \geq 2$

$$\int_{a_{j-1}}^{a_j} \mathbb{P}(X \geq t)\ dt = (a_j - a_{j-1}) \sum_{i \geq j} p_i.$$

Adding all these expression we get

$$\begin{aligned}\int_0^\infty \mathbb{P}(X \geq t)\ dt &= \int_0^{a_1} \mathbb{P}(X \geq t)\ dt + \int_{a_1}^{a_2} \mathbb{P}(X \geq t)\ dt + \int_{a_2}^{a_3} \mathbb{P}(X \geq t)\ dt + \dots \\ &= a_1 \sum_{i \geq 1} p_i + (a_2 - a_1) \sum_{i \geq 2} p_i + (a_3 - a_2) \sum_{i \geq 3} p_i + \dots \\ &= p_1 a_1 + p_2(a_1 + (a_2 - a_1)) + p_3(a_1 + (a_2 - a_1) + (a_3 - a_2)) + \dots \\ &= p_1 a_1 + p_2 a_2 + p_3 a_3 + \dots = \mathbb{E}(X).\end{aligned}$$