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

Homework: The third homework will be due on Friday 8 December, at 12pm, the *start* of the lecture. It will consist of questions:

8 (from supp. problems) and 5.

1. From the textbook, chapter 2, problems 7, 8, 11, 13, 15, 16.

2. From the supplementary problems, chapter 2, problem 6, 8, 9, 14. Problem 8 has been typed out below since there was a typo in the supplementary problems as posted on the textbook website.

3. If X is an exponential random variable with parameter λ find both PDF and CDF of the random variable $Y = e^X$.

Solution: The range of X is the interval $[0, \infty)$ so the range for Y is the interval $[1, \infty)$. Therefore $F_Y(t) = 0$ and $f_Y(t) = 0$ for $t < 1$. For $t \geq 0$ we have $\log t \geq 0$ and since we know the PDF of X we get easily

$$\mathbb{P}(Y \leq t) = \mathbb{P}(X \leq \log t) = 1 - e^{-\lambda \log t} = 1 - t^{-\lambda},$$

and $f_Y(t) = \lambda t^{-\lambda - 1}$.

4. A random variable X is said to have arcsine law if it's CDF is

$$F_X(t) = \frac{2}{\pi} \arcsin(\sqrt{t}).$$

   (a) What are the possible values that X can have?

   (b) Find the median of X.

   (c) Find the PDF of X (you might want to recall some derivatives of typical functions).

   (d) Find the mean $\mathbb{E}(X)$ (to compute the integral you might want to use the trigonometric substitution and trigonometric double angle identities).

Solution:

1) The CDF $F_X(t)$ isn't completely specified here. However, since domain for arcsin is $[-1, 1]$, the domain for $F_X(t)$ is $[0, 1]$ and $F_X(0) = 0$ and $F_X(1) = 1$ and so $F_X(t)$ is constant equal to zero for $t < 0$ and constant equal to 1 for $t > 1$ and so the range of X is the interval $[0, 1]$.

2) The median m is the number such that $\mathbb{P}(X \leq m) = 1/2$ so we need to find m such that $F_X(m) = 1/2$ that is $\frac{2}{\pi} \arcsin(\sqrt{m}) = 1/2$ or $\arcsin(\sqrt{m}) = \pi/4$, that is $\sqrt{m} = 1/\sqrt{2}$ so $m = 1/2$.

3) PDF $f_X(t)$ is obtained by taking the derivative of $F_X(t)$ which is

$$f_X(t) = \frac{2}{\pi} \frac{1}{\sqrt{1-t}} \frac{1}{2\sqrt{t}} = \frac{1}{\pi\sqrt{t(1-t)}}.$$

4) Using substitution $t = \sin^2\theta$, for $0 \leq \theta \leq \pi/2$ we get

$$\mathbb{E}(X) = \int_0^1 \frac{t}{\pi\sqrt{t(1-t)}}\, dt = \int_0^{\pi/2} \frac{\sin\theta}{\pi\cos\theta} 2\sin\theta\cos\theta\, d\theta = \frac{1}{\pi}\int_0^{\pi/2} 2\sin^2\theta\, d\theta$$

Using the double angle formula $\cos(2\theta) = 1 - 2\sin^2\theta$ we get

$$\mathbb{E}(X) = \frac{1}{\pi}\int_0^{\pi/2} 1 - \cos(2\theta)\theta \; d\theta = \frac{1}{2}.$$

5. Let A be the set of all pairs (x, y) which satisfy each of the conditions

$$x \geq 0, \;\; y \geq 0, \;\; 1 \leq x + y \leq 2.$$

Let the random variables X and Y be jointly continuous with the joint PDF

$$f_{X,Y}(x,y) = \begin{cases} C(x+y), & \text{if } (x,y) \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Find the value of the constant C. Find the marginal PDFs and CDFs.

Solution: To find C integrate

$$C\iint_A (x+y)\; dydx = C\int_0^1\int_{1-x}^{2-x}(x+y)\; dydx + C\int_1^2\int_0^{2-x}(x+y)\; dydx$$
$$= C\int_0^1 3/2\; dx + C\int_1^2 2 - x^2/2\; dx = 7C/3 \;\Rightarrow\; C = 3/7.$$

Marginal PDF for X is $f_X(x)$ and for $x < 0$ and $x > 2$ we have $f_X(x) = 0$. For $0 \leq x \leq 1$

$$f_X(x) = \frac{3}{7}\int_{1-x}^{2-x}(x+y)\; dy = \frac{9}{14}.$$

For $1 < x \leq 2$

$$f_X(x) = \frac{3}{7}\int_0^{2-x}(x+y)\; dy = \frac{3}{7}(2 - \frac{x^2}{2}).$$

The PDF is $F_X(x) = 0$ for $x \leq 0$, $F_X(x) = 1$, for $x \geq 2$ and for $0 < x \leq 1$ it is

$$F_X(x) = \int_0^x \frac{9}{14}\; dx = \frac{9x}{14},$$

and for $1 < x \leq 2$

$$F_X(x) = \frac{9}{14} + \int_1^x \frac{6}{7} - \frac{3t^2}{14}\; dt = \frac{9}{14} + \frac{6x}{7} - \frac{6}{7} - \frac{x^3}{14} + \frac{1}{14} = \frac{6x}{7} - \frac{x^3}{14} - \frac{1}{7}.$$

6. A food store just got a delivery of 10000 potatoes. It is known that each potato is rotten with probability 0.1. What is the expected number of healthy potatoes? Use the normal approximation to estimate the probability that there are at least 8970 healthy potatoes.

Solution: The probability is

$$\sum_{k=8970}^{10000}\binom{10000}{k}0.9^k 0.1^{10000-k}.$$

Since the number of healthy potatoes is Binomial with parameters 10000 and 0.9, it has mean 9000 and variance 900 we approximate this by the probability that the standard Normal Y

$$\mathbb{P}(Y \geq \frac{8970-9000}{30}) = \mathbb{P}(Y \geq -1) = \mathbb{P}(Y \leq 1) \approx 0.8413.$$

7. (Problem 8 from supplementary problems, with typo fixed.) A signal of amplitude $s = 2$ is transmitted from a satellite but is corrupted by noise, and the received signal is $Z = s + W$, where W is noise. When the weather is good, W is normal with zero mean and variance 1. When the weather is bad, W is normal with zero mean and variance 4. Good and bad weather are equally likely. In the absence of any weather information:
   (a) Calculate the PDF of Z.
   (b) Calculate the probability that Z is between 1 and 3.

Solution:

(a) First we compute the CDF:

$$\begin{aligned} F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(W \leq t-2) &= \mathbb{P}(W \leq t-2|\text{weather good})\mathbb{P}(\text{weather good}) \\ &+ \mathbb{P}(W \leq t-2|\text{weahter bad})\mathbb{P}(\text{weather bad}) \\ &= \frac{1}{2}\left(\int_{-\infty}^{t-2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\ dx + \int_{-\infty}^{t-2} \frac{1}{2\sqrt{2\pi}} e^{-x^2/8}\ dx\right) \end{aligned}$$

Differentiating with respect to t we get the PDF

$$f_Z(t) = \frac{1}{4\sqrt{2\pi}}\left(2e^{-(t-2)^2/2} + e^{-(t-2)^2/8}\right).$$

Alternatively, you can avoid computing CDF first by using a version of the total probability law for the PDF (see the last formula in the boxed are on page 167).

(b) If the weather is good then

$$\mathbb{P}(1 \leq Z \leq 3|\text{weather good})) = \mathbb{P}(-1 \leq W \leq 1|\text{weather good}).$$

This is the probability that a standard normal random variable is between -1 and 1 which is equal to $2\times$ (the probability that a standard normal is less than 1) minus 1. Looking at the table this is approximately $2 \times 0.8413 - 1 = 0.6826$.

Similarly $\mathbb{P}(1 \leq Z \leq 3|\text{weather bad})$ is the probability that a Normal with mean 0 and variance 4 is between -1 and 1 which is the probability that a standard Normal is between -0.5 and 0.5. This is $2\times$ (CDF of standard Normal at 0.5) minus 1 which is approximately $2 \times 0.6915 - 1 = 0.383$. Thus

$$\mathbb{P}(1 \leq Z \leq 3) \approx \frac{1}{2}(0.6828 + 0.383) = 0.5329.$$