STAT 244 (Statistical Theory/Method I), Winter 2016

Homework 2 solutions

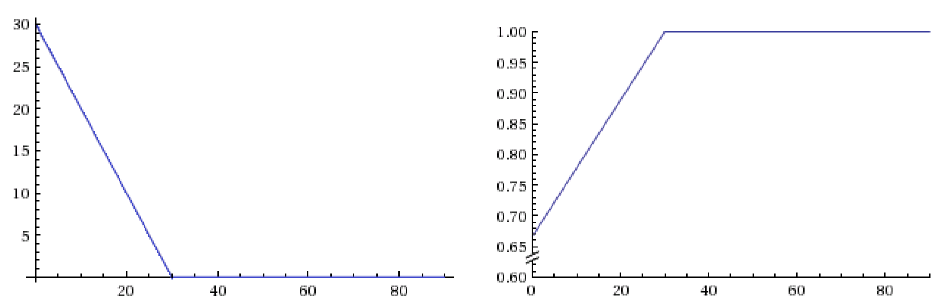
January 19, 2016

100 points total. Marks are assigned based on progress made.

Problem 1 (10 points)

Let X be your "random" arrival time, uniformly distributed on $[0, 90]$, and let $f(x) = \max(30 - x, 0)$, so that the random variable $f(X)$ is your wait time in seconds (shown in Figure 1). The CDF of $f(X)$ is

$$F_{f(X)}(x) = \mathbb{P}(f(X) \le x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{60+x}{90} & \text{if } 0 \le x \le 30 \\ 1 & \text{if } x > 30. \end{cases} \tag{1}$$



$f(X)$ CDF of $f(X)$

Figure 1

8 points for justification, 2 points for correct answer.

Problem 2 (15 points)

(a) Since $f_{X,Y}(x, y)$ is the joint density of X and Y, it must integrate to 1 on its domain, so c must satisfy the condition

$$\int_0^\infty \int_{-x}^{x} c(x^2 - y^2)e^{-x} \, \mathrm{d}y\mathrm{d}x = \int_0^\infty \frac{4}{3} c x^3 e^{-x} \, \mathrm{d}x = \frac{4}{3} c\Gamma(4) = 8c = 1, \tag{2}$$

which means $c = \frac{1}{8}$.

(b) The domain of the $f_{X,Y}(x, y)$ is shown in Figure 2. To find the marginal density of X, fix x and integrate over all possible values of y, and similarly for the marginal density of Y:

$$f_X(x) = \int_{-x}^{x} \frac{1}{8}(x^2 - y^2)e^{-x}\ \mathrm{d}y = \frac{1}{6}x^3 e^{-x} \tag{3}$$

$$f_Y(y) = \int_{|y|}^{\infty} \frac{1}{8}(x^2 - y^2)e^{-x}\ \mathrm{d}x = \frac{1}{4}(|y|+1)e^{-|y|} \tag{4}$$

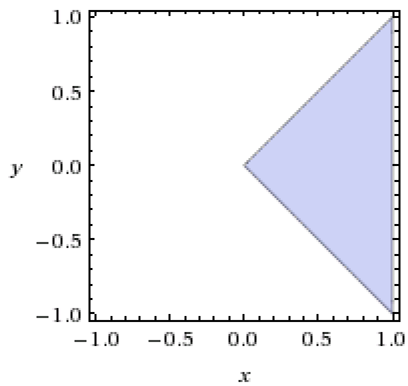


Figure 2: Domain of $f_{X,Y}(x,y)$

(c) Using (3) and (4), the conditional densities are:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\frac{1}{8}(x^2-y^2)e^{-x}}{\frac{1}{4}(|y|+1)e^{-|y|}} = \frac{(x^2-y^2)e^{-x+|y|}}{2(|y|+1)} \tag{5}$$

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{\frac{1}{8}(x^2-y^2)e^{-x}}{\frac{1}{6}x^3 e^{-x}} = \frac{3(x^2-y^2)}{4x^3} \tag{6}$$

5 points for each part. Within each part: 4 points for setup, 1 point for correct answer.

Problem 3 (10 points)

Note that $Y = \tan(X)$, and since the tangent function is differentiable and bijective, the density of Y is

$$f_Y(y) = f_X(\arctan(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y}(\arctan(y)) \right| = \frac{1}{\pi(1+y^2)} \tag{7}$$

by Proposition B on page 62 of the textbook. This is the standard Cauchy distribution.

8 points for setup, 2 points for correct answer.

Problem 4 (14 points)

(a) The possible values that X can take are 0, 1, and 2, and their probabilities can be calculated using the fact that X_1 and X_2 are independent as follows:

$$\mathbb{P}(X=0) = \mathbb{P}(\{X_1=0\} \cap \{X_2=0\}) = (1-p_1)(1-p_2) \tag{8}$$

$$\mathbb{P}(X=1) = \mathbb{P}\left((\{X_1=0\} \cap \{X_2=1\}) \cup (\{X_1=1\} \cap \{X_2=0\})\right) = p_2(1-p_1) + p_1(1-p_2) \tag{9}$$

$$\mathbb{P}(X=2) = \mathbb{P}(\{X_1=1\} \cap \{X_2=1\}) = p_1 p_2 \tag{10}$$

(b) The difference between the left-hand and right-hand sides is

$$\mathbb{P}(X=1) - \mathbb{P}(Y=1) = (1-p_1)p_2 + p_1(1-p_2) - 2\left(\frac{p_1+p_2}{2}\right)\left(1 - \frac{p_1+p_2}{2}\right) = \frac{(p_1-p_2)^2}{2}, \tag{11}$$

which is strictly greater than 0 as long as $p_1 \neq p_2$, so $\mathbb{P}(X=1) > \mathbb{P}(Y=1)$ as long as the distribution parameters p_1 and p_2 are different.

(c) A basketball team can either

1. play a "home-and-home" series as described, where p_1 is the probability that the team wins at home and p_2 is the probability that the team wins away with $p_1 > p_2$, or
2. play two consecutive games at some other location that is not the home stadium of either team, where the probability that the team wins is instead the average $\frac{p_1+p_2}{2}$.

Then X and Y represent respectively the number of games a team wins in the first and second scenarios. Hence part (b) means it is always more likely for there to be a draw in a "home-and-home" series.

5 points for (a) and (b), 4 points for (c).

Problem 5 (16 points)

(a) To get the CDF of $\sin(U)$, we are interested in where $\sin(U)$ is less than a certain value $x \in [-1, 1]$. The values of U for which this occurs are indicated in yellow in Figure 3. Then the CDF of $\sin(U)$ is

$$F_{\sin(U)}(x) = \mathbb{P}(\sin(U) \leq x) = \frac{\pi + 2\arcsin(x)}{2\pi}, \tag{12}$$

so its density is the derivative

$$f_{\sin(U)}(x) = \frac{\mathrm{d}F_{\sin(U)}}{\mathrm{d}x} = \frac{1}{\pi\sqrt{1-x^2}}. \tag{13}$$

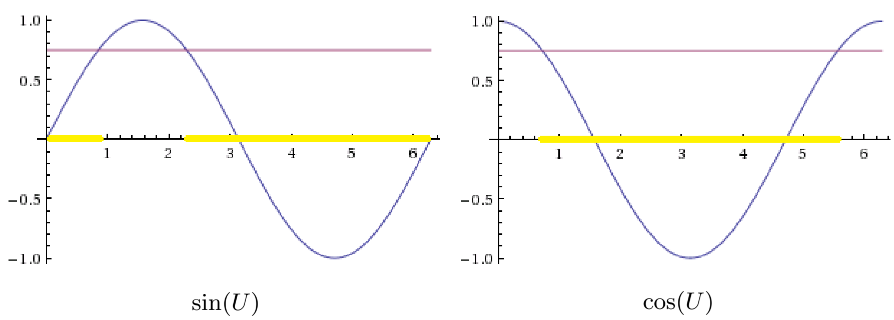


Figure 3

(b) As in part (a) (see again Figure 3), the CDF of $\cos(U)$ is

$$F_{\cos(U)}(x) = \mathbb{P}(\cos(U) \leq x) = \frac{2\pi - 2\arccos(x)}{2\pi}, \tag{14}$$

so its density is

$$f_{\cos(U)}(x) = \frac{\mathrm{d}F_{\cos(U)}}{\mathrm{d}x} = \frac{1}{\pi\sqrt{1-x^2}}. \tag{15}$$

(c) Both the CDFs and densities of $\sin(U)$ and $\cos(U)$ are the same, which we could have concluded without doing the calculations in parts (a) and (b) because the sine curve is just the cosine curve shifted over to the right by $\frac{\pi}{2}$.

(d) Note that the random variable $\sin^2(U) + \cos^2(U)$ takes the value 1 on the entire probability space, so its CDF is the step function

$$F_{\sin^2(U)+\cos^2(U)}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geq 1, \end{cases} \tag{16}$$

which is shown in Figure 4.

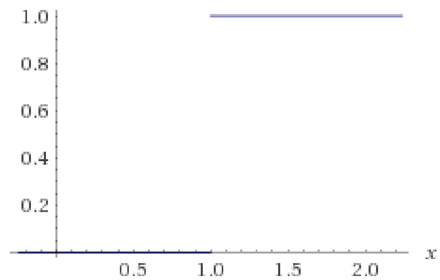


Figure 4: CDF of $\sin^2(U) + \cos^2(U)$

4 points for each part.

Problem 6 (15 points)

First suppose $X \sim \mathcal{N}(0,1)$. The third and fourth moments of X can be computed using integration by parts:

$$\begin{aligned}
\mathbb{E}(X^3) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 e^{-\frac{x^2}{2}} \, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \left(x e^{-\frac{x^2}{2}} \right) \, \mathrm{d}x \\
&= -\frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 2x e^{-\frac{x^2}{2}} \, \mathrm{d}x = \mathbb{E}(X) = 0 \qquad (17) \\
\mathbb{E}(X^4) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} \, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 \left(x e^{-\frac{x^2}{2}} \right) \, \mathrm{d}x \\
&= -\frac{1}{\sqrt{2\pi}} x^3 e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 3x^2 e^{-\frac{x^2}{2}} \, \mathrm{d}x = 3\mathbb{E}(X^2) = 3 \qquad (18)
\end{aligned}$$

Alternatively, we can see that $\mathbb{E}(X^3) = 0$ because x^3 is an odd function and the standard normal density function is symmetric about $x = 0$. Now if $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $X = \frac{Y-\mu}{\sigma}$ is a standard normal random variable, so using the binomial theorem and linearity of expectation, the n^{th} moment of Y is given in terms of the moments of X by

$$\mathbb{E}(Y^n) = \mathbb{E}((\sigma X + \mu)^n) = \mathbb{E}\left(\sum_{k=0}^{n} \binom{n}{k} (\sigma X)^k \mu^{n-k} \right) = \sum_{k=0}^{n} \binom{n}{k} \sigma^k \mu^{3-k} \mathbb{E}(X^k). \qquad (19)$$

Plugging in $n = 3$ and $n = 4$, we get the third and fourth moments of Y:

$$\mathbb{E}(Y^3) = \mu^3 + 3\sigma^2\mu \qquad (20)$$

$$\mathbb{E}(Y^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \qquad (21)$$

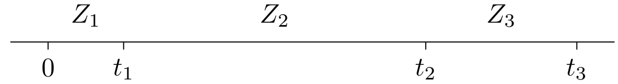
7 points for standard case, 8 points for general case.

Problem 7 (20 points)

(a) Recall that the mean and variance of a Poisson random variable with parameter λ are both λ (see Example C on page 117 of the textbook and Example A on page 156 of the textbook), so

$$\mathbb{E}(Y - X) = \mathbb{E}(Y) - \mathbb{E}(X) = \lambda(t_3 - t_1) - \lambda t_2 = \lambda(t_3 - t_1 - t_2). \qquad (22)$$

Unfortunately, the variance is not as straightforward because X and Y are not independent. It would be convenient if we could write X and Y in terms of independent random variables because then we would be able to add variances.



$X = Z_1 + Z_2 \qquad Y = Z_2 + Z_3$

Figure 5: Split up the interval into disjoint components.

In Figure 5, note that X is the sum of Z_1 (a Poisson random variable with parameter λt_1) and Z_2 (a Poisson random variable with parameter $\lambda(t_2 - t_1)$), and similarly for Y. Now calculating the variance becomes manageable:

$$\begin{aligned}\mathrm{Var}(Y - X) &= \mathrm{Var}((Z_1 + Z_2) - (Z_2 + Z_3)) = \mathrm{Var}(Z_1 - Z_3) \\ &= \mathrm{Var}(Z_1) + \mathrm{Var}(Z_3) - 2\mathrm{Cov}(Z_1, Z_3) = \lambda t_1 + \lambda(t_3 - t_2) = \lambda(t_1 + t_3 - t_2)\end{aligned} \tag{23}$$

(b) Assuming the same setup as in Figure 5, the conditional expectation is

$$\mathbb{E}(Y \mid X) = \mathbb{E}(Z_2 + Z_3 \mid X) = \mathbb{E}(Z_2 \mid X) + \mathbb{E}(Z_3 \mid X). \tag{24}$$

Since Z_3 and X are independent, $\mathbb{E}(Z_3 \mid X) = \mathbb{E}(Z_3)$ (being provided information about X will not change what we expect Z_3 to be). On the other hand, $\mathbb{E}(Z_2 \mid X = x)$ is a binomial random variable with parameter $n = x$ and $p = \frac{t_2 - t_1}{t_2}$ (see Example A on page 147 of the textbook) with mean np, so

$$\mathbb{E}(Y \mid X = x) = x\left(\frac{t_2 - t_1}{t_2}\right) + \lambda(t_3 - t_2), \tag{25}$$

and furthermore

$$\begin{aligned}\mathbb{E}\left(\mathbb{E}(Y \mid X)\right) &= \sum_{x=0}^{\infty} \mathbb{E}(Y \mid X = x)\mathbb{P}(X = x) = \sum_{x=0}^{\infty}\left(x\left(\frac{t_2 - t_1}{t_2}\right) + \lambda(t_3 - t_2)\right)\frac{(\lambda t_2)^x}{x!}e^{-\lambda t_2} \\ &= t_2\lambda\left(\frac{t_2 - t_1}{t_2}\right) + \lambda(t_3 - t_2) = \lambda(t_3 - t_1) = \mathbb{E}(Y).\end{aligned} \tag{26}$$

8 points for (a), 12 points for (b). An intuitive justification for why $\mathbb{E}(Z_2 \mid X)$ is binomial is sufficient.