Problem Set 7 - Solutions

Statistics - NYU, Summer 2016
Ercan Karadas

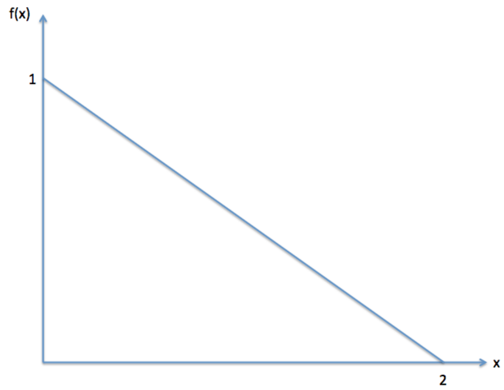
Section 1

[1] a) To show that $f(x)$ is a probability density function, we need to show that:

i. $f(x) \geq 0$ for all x

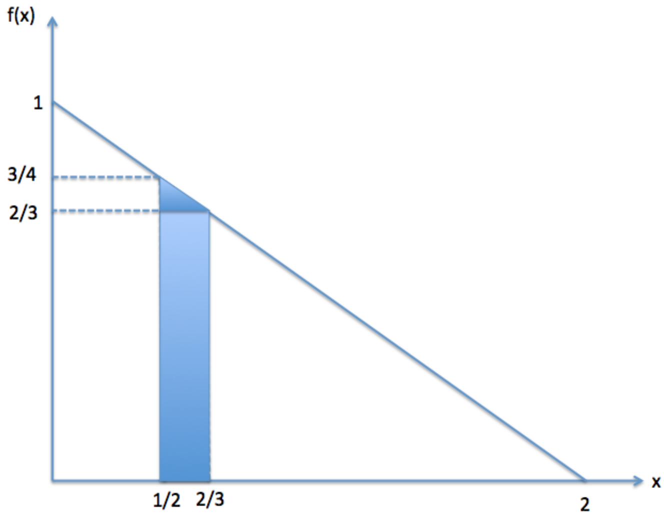
ii. $\int\limits_0^2 f(x)dx = 1$

From the plot below (or from definition of $f(x)$), we see that condition (i) is clearly satisfied. To check (ii), note that $\int\limits_0^2 f(x)dx =$Area of the triangle $= \frac{1}{2}(2)(1) = 1$. Therefore, $f(x)$ is a p.d.f.



b)

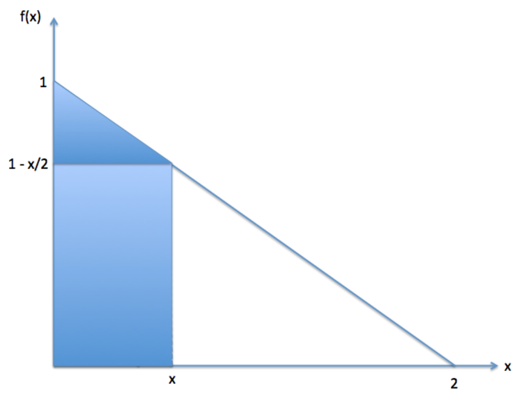
$$P\Big(\frac{1}{2} < X < \frac{2}{3}\Big) = \int_{1/2}^{2/3} f(x)dx = \int_{1/2}^{2/3} (1 - \frac{x}{2})dx$$
$$= \text{Area of trapezium}$$
$$= \frac{1}{2}\Big(\frac{2}{3} - \frac{1}{2}\Big)\Big(\frac{3}{4} + \frac{2}{3}\Big)$$
$$= \frac{17}{144}$$



Note that instead of evaluating the integral on the first line we just exploited the nice geometry of the corresponding area. Whenever possible we will do so in the rest of the problem set.

c)

$$
\begin{aligned}
F(x) = P(X < x) = \int_0^x f(x)dx = \int_0^x (1 - \frac{x}{2})dx \\
&= \text{Area of trapezium} \\
&= \frac{1}{2}x\Big(1 + \Big(1 - \frac{x}{2}\Big)\Big) \\
&= x - \frac{x^2}{4} \text{ for } x \in [0, 2]
\end{aligned}
$$



Thus

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x - \frac{x^2}{4} & \text{for } x \in [0,2] \\ 1 & \text{for } x > 2 \end{cases}$$

d)

$$\begin{aligned} F\Big(\frac{2}{3}\Big) - F\Big(\frac{1}{2}\Big) &= \Big[\frac{2}{3} - \frac{1}{4}\Big(\frac{2}{3}\Big)^2\Big] - \Big[\frac{1}{2} - \frac{1}{4}\Big(\frac{1}{2}\Big)^2\Big] \\ &= \frac{17}{144} \end{aligned}$$

[2]

a)

$$\begin{aligned} E(X) &= \int_0^2 xf(x)dx = \int_0^2 x(1 - \frac{x}{2})dx = \frac{1}{2}\int_0^2 x(2-x)dx \\ &= \frac{1}{2}\big[x^2 - \frac{x^3}{3}\big]_0^2 = \frac{2}{3} \end{aligned}$$

b)

$$\begin{aligned} E[g(X)] &= \int_0^2 g(x)f(x)dx \\ &= \int_0^2 (3x^2 - 4x + 7)(1 - \frac{x}{2})dx = \frac{1}{2}\int_0^2 (3x^2 - 4x + 7)(2-x)dx = \dots \end{aligned}$$

After expanding the expression inside, then evaluate the integral to find $E[g(X)]$. It involves a little algebra, but you should conclude it.

There is a second way to approach this problem: instead of first obtaining $g(x)f(x)$ explicitly and then evaluating the resultant expression, we could first split the integral into three and then evaluate each part separately as follows:

$$\begin{aligned}
E[g(X)] &= \int_0^2 g(x)f(x)dx \\
&= \int_0^2 (3x^2 - 4x + 7)f(x)dx \\
&= 3\int_0^2 x^2 f(x)dx - 4\int_0^2 xf(x)dx + 7\int_0^2 1.f(x)dx \\
&= 3\int_0^2 x^2 f(x)dx - 4E(X) + 7
\end{aligned}$$

On the third line, we used the fact that the integral in the middle is equal to $E(X)$ and the last is equal to 1, so we only need to evaluate the first integral. That is the main advantage of this second approach.

$$\begin{aligned}
\int_0^2 x^2 f(x)dx = \int_0^2 x^2(1 - \frac{x}{2})dx = \frac{1}{2}\int_0^2 x^2(2-x)dx = \frac{1}{2}\int_0^2 (2x^2 - x^3)dx \\
= \frac{1}{2}[2\frac{x^3}{3} - \frac{x^4}{4}]_0^2 = \frac{1}{2}[\frac{2^4}{3} - \frac{2^4}{4}] = \frac{2}{3}
\end{aligned}$$

Now we can compute $E[g(X)]$ as

$$\begin{aligned}
E[g(X)] &= 3\int_0^2 x^2 f(x)dx - 4E(X) + 7 \\
&= 3\frac{2}{3} - 4(\frac{2}{3}) + 7 = \frac{19}{3}
\end{aligned}$$

c)

$$Var(X) = \int_0^2 (x-\mu)^2 f(x)dx = \int_0^2 (x - \frac{2}{3})^2(1 - \frac{x}{2})dx = \dots$$

Again, instead of evaluating the integral above, we could proceed as in the previous part to obtain

$$Var(X) = \int_0^2 x^2 f(x)dx - \mu^2$$

And since we have already computed these on the right hand side, we conclude

$$Var(X) = \frac{2}{3} - (\frac{2}{3})^2 = \frac{2}{9}$$

[3] Let X represent a number picked at random from the interval (0,1). X is a uniformly distributed random variables with the below density function:

$$f(x) = \begin{cases} \frac{1}{1-0}, & 0 \leq x \leq 1 \\ 0, & \text{for all other x} \end{cases}$$

Then

$$P(X > 0.7) = \int_{0.7}^{1} f(x)dx = \int_{0.7}^{1} 1dx = \left[x\right]_{.7}^{1} = 0.3$$

Then, all five are be greater than 0.7 with $(0.3)^5$ probability. In this problem we can think as if X is drawn 5 times and the probability that $X > 0.7$ is independent over the trials.

[4] Let X denote the random variable for the amount of money generated in a week. We are given that $\mu_X = 700$ and $\sigma_X = 130$. Let P denote the random variable for the employees total pay in a week, then $P = 60 + 0.2X$ and

$E[P] = E[60 + 0.2X] = 60 + 0.2\mu_X = 200.$
$Var(P) = Var(60 + 0.2X) = 0.2^2\sigma_X^2$. Thus $\sigma_P = \sqrt{Var(P)} = 0.2\sigma_X = 26.$

[5] Let X denote the time for service calls, then $X \sim N(60, 10^2)$.

a) $P(X > 65) = P(Z > \frac{65-60}{10}) = P(Z > .5) = 1 - F(.5) = .3085$

b) $P(50 < X < 70) = P(\frac{50-60}{10} < Z < \frac{70-60}{10}) = P(-1 < Z < 1) = 2F(1) - 1 = .6826$

c) We need to find an x value such that $P(X > x) = 0.25$. Let us first find the z-value that leaves 0.25% of the total area in the upper tail. From standard normal distribution table $P(Z > 0.675) = .25$, which means the area to the left and right of $z = 0.675$ is equal to 0.75, and 0.25, respectively. In order to find the x value that we were looking for we just need to solve $\frac{x-60}{10} = 0.675$ for x. Doing so yields $x = 66.75$. Here note that z=0.675 (which is not directly in the table) is the arithmetic mean of the two values from the table, z=0.67 and z=0.68, since none of these latter values gives the exact probability that we were looking for.

d) First note that the shortest range will be the interval centered on the mean. If this interval has 0.5 probability then in the lower and upper tails we should have 25% of the total area for each. For example let us focus on the upper tail, and find the z-value that leaves 0.25 probability in the upper tail, i.e. find the z-value such that $P(Z > z) = 0.25$. From the table, we find that $z = .675$. Then as in the previous part solving the equation $\frac{x-60}{10} = .675$ for x yields $x = 66.75$. This is the upper bound of the interval. But since the interval should be symmetric around 60, the lower bound of the interval is 53.25. Therefore the shortest range is $66.75 - 53.25 = 13.5$. In order to understand this solution you should definitely draw a picture, and then work on that to visualize the solution.

e) Let us define X as the number of service calls that takes more than 65 minutes. This is a binomial r.v. with the success probability $p = .3085$ (found in part (a)), and number of trials is 4, $n = 4$, and we want to find $P(X = 2)$. From the binomial formula
$P(X = 2) = \binom{4}{2}(.3085)^2(.6915)^2 = 0.273$

[6] Let $X = \#$ of people who do show up and $p = .9$ is the probability that a person who made a reservation will shop up. We would like to find $P(X \leq 80)$. Since $np(1-p) = 90(.9)(.1) = 8.1 > 5$ is satisfied we can use the normal approximation to Binomial. $\mu = np = 90(.9) = 81$, and $\sigma = \sqrt{np(1-p)} = \sqrt{90(.9)(.1)} = 2.85$, assuming normality we can say $X \backsim N(8.1, (2.85)^2)$ and then

$$P(X \leq 80) = P(Z < \frac{80 - 81}{2.85}) = P(Z < -0.35) = \dots$$

[7] (Skip!)

[8] Portfolio consists of 10 shares of stock A and 8 shares of stock B, so we need to define a new random variable $W = 10A + 8B$ for the portfolio value.

a) $\mu_w = 10\mu_A + 8\mu_B = 10(10) + 8(12) = 196$

$$\begin{aligned} Var(W) &= 10^2 Var(A) + 8^2 Var(B) + 2(10)(8) Corr(A, B)\sigma_A\sigma_B \\ &= 10^2(16) + 8^2(9) + 2(10)(8)(.3)(4)(3) \\ &= 2,752 \end{aligned}$$

b) Here Stock B is the same as in part (a), but we have two options for Stock A, and we are trying to calculate which option yields the lowest variance for the portfolio value.

Option 1: Let us denote Stock 1 with A_1. Then $\mu_{A_1} = 10$, $\sigma^2_{A_1} = 25$ and $Corr(A_1, B) = -(0.2)$.

$$\begin{aligned} Var(W_1) &= 10^2 Var(A_1) + 8^2 Var(B) + 2(10)(8) Corr(A_1, B)\sigma_{A_1}\sigma_B \\ &= 10^2(25) + 8^2(9) + 2(10)(8)(-0.2)(5)(3) \\ &= 2,596 \end{aligned}$$

Option 2: Let us denote Stock 2 with A_2. Then $\mu_{A_2} = 10$, $\sigma^2_{A_2} = 9$ and $Corr(A_2, B) = 0.5$.

$$\begin{aligned} Var(W_2) &= 10^2 Var(A_2) + 8^2 Var(B) + 2(10)(8) Corr(A_2, B)\sigma_{A_2}\sigma_B \\ &= 10^2(9) + 8^2(9) + 2(10)(8)(0.5)(3)(3) \\ &= 2,196 \end{aligned}$$

Therefore, choosing stock 2 reduces the variance (risk) of the portfolio.

[9] (Skip!)

[10] a) Poisson distribution with $\lambda = 6$ per-hour; X: number of complaints
$P(X = 6) = \frac{e^{-6}6^{-6}}{6!} = 0.1606$ (You can use Individual Poisson Probabilities table in the book to find this value)

b) $P(X \leq 6) = P(X = 0) + \ldots + P(X = 6) = \frac{e^{-6}6^{0}}{0!} + \frac{e^{-6}6^{-1}}{1!} + \ldots + \frac{e^{-6}6^{-6}}{6!} =$ 0.6063 (You can use Cumulative Poisson Probabilities table in the book)

c) Skip!

d) Skip!

Section 2

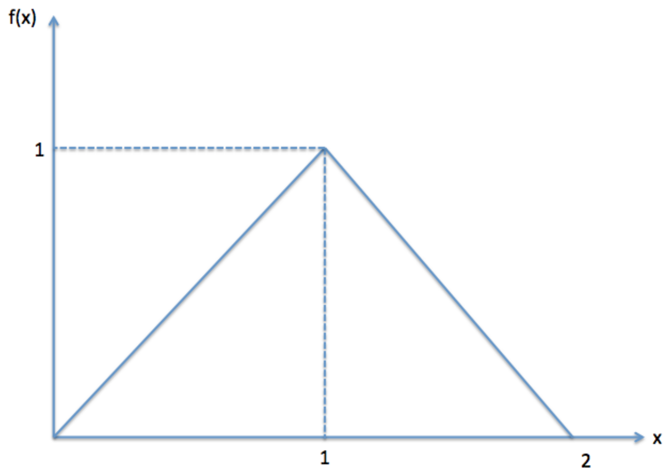
[11] a) To show that $f(x)$ is a probability density function, we need to show that:

i. $f(x) \geq 0$ for all x

ii. $\int_0^2 f(x)dx = 1$

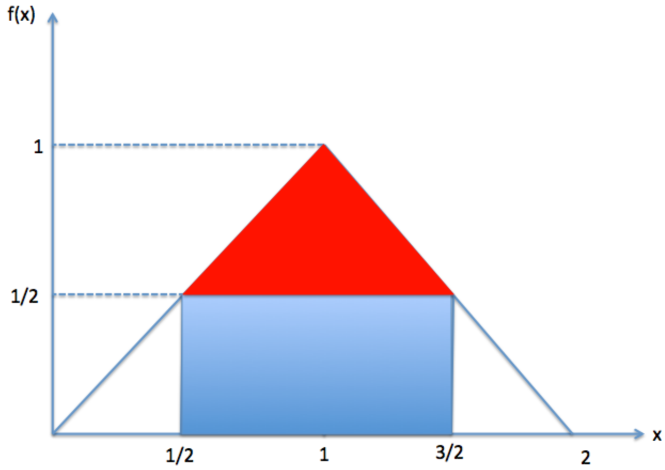
From the plot below, we see that condition (i) is satisfied.

Also, $\int_0^2 f(x)dx =$Area of the triangle $= \frac{1}{2}(2)(1) = 1$.



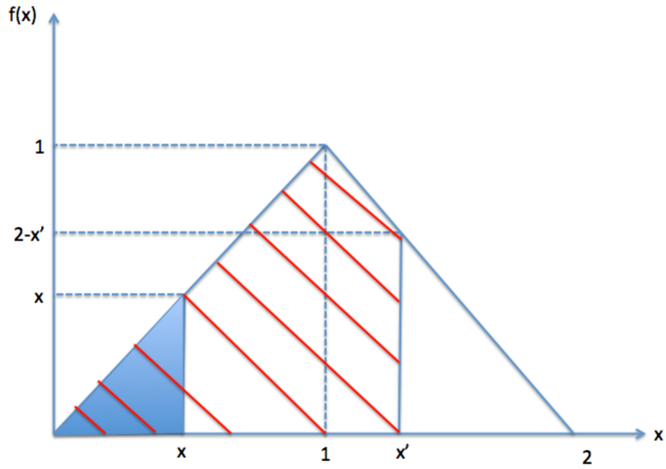
b)

$$\begin{aligned} P\left(\frac{1}{2} < X < \frac{3}{2}\right) &= \text{Area of rectangle} + \text{Area of triangle} \\ &= 1\left(\frac{1}{2}\right) + \frac{1}{2}\left(1\right)\left(\frac{1}{2}\right) \\ &= \frac{3}{4} \end{aligned}$$



c)

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}x^2 & \text{for } 0 \leq x \leq 1 \\ 1 - \frac{1}{2}(2-x)^2 & \text{for } 1 \leq x \leq 2 \\ 1 & \text{for } x > 2 \end{cases}$$



Notice that for $x \in [0,1]$, $F(x)$ is the area of the triangle to the left of x under the curve. For $x' \in [1,2]$, $F(x')$ is the area under the curve to the left of x', which is the total area (1) minus the area of the triangle to the right of x'. Also note that saying $F(x) = \frac{x^2}{2}$, for all $x \in [0,2]$ would be wrong because for $x > 1$ the corresponding area includes a triangle that is actually outside the area we are interested in.

d)

$$F\Big(\frac{3}{2}\Big) - F\Big(\frac{1}{2}\Big) = \Big[1 - \frac{1}{2}\Big(2 - \frac{3}{2}\Big)^2\Big] - \Big[\frac{1}{2}\Big(\frac{1}{2}\Big)^2\Big]$$
$$= \frac{3}{4}$$

$F\Big(\frac{3}{2}\Big)$ represents the area under the curve to the left of $x = \frac{3}{2}$. Similarly, $F\Big(\frac{1}{2}\Big)$ represents the area under the curve to the left of $x = \frac{1}{2}$. The difference is thus the probability that x lies in $\Big[\frac{1}{2}, \frac{3}{2}\Big]$.

[12] a)

$$\int_0^{10} f(x)dx = \int_0^{10} \frac{x^3}{5000}(10 - x)dx$$
$$= \Big[\frac{x^4}{2000} - \frac{x^5}{25000}\Big]_0^{10}$$
$$= 1$$

b)

$$P(1 < X < 4) = \int_1^4 \frac{x^3}{5000}(10 - x)dx$$
$$= \Big[\frac{x^4}{2000} - \frac{x^5}{25000}\Big]_1^4$$
$$= 0.087$$

c)

$$P(X > 6) = \int_6^{10} \frac{x^3}{5000}(10 - x)dx$$
$$= \Big[\frac{x^4}{2000} - \frac{x^5}{25000}\Big]_6^{10}$$
$$= 0.66$$

[13] Let X denote the random variable for the number of units produced. We are given that $\mu_X = 500$ and $\sigma_X^2 = 900$. Let π denote the random variable for the profit, then $\pi = 2000 - 2X$ and therefore

$E[\pi] = E[2000 - 2X] = 2000 - 2\mu_X = 1000.$
$Var(\pi) = Var(2000 - 2X) = 2^2\sigma_X^2 = 3600.$

[14] X = Concert time $\Rightarrow X \sim N(200, (20)^2)$

a) $P(180 < X < 200) = ?$
$P(180 < X < 200) = P(\frac{180-200}{20} < Z < \frac{200-200}{20}) = P(-1 < Z < 0) = F(0) - F(-1) = F(0) - [1 - F(1)] = 0.5 + 0.8413 - 1 = 0.3413$

b) $P(X < 245) = P(\frac{X-200}{20} < \frac{245-200}{20}) = P(Z < 2.25) = F(2.25) = 0.9878$

c) Since now standard deviation is smaller the r.v. is more tightly distributed around the mean, and therefore the probability that X is higher than 245 is lower.

d) Find an x value such that $P(X < x) = 0.1$ is satisfied.
$P(X < x) = P(\frac{X-200}{20} < \frac{x-200}{20}) = P(Z < \frac{X-200}{20}) = 0.1$ which implies $F(\frac{x-200}{20}) = 0.1 \Rightarrow \frac{x-200}{20} = -1.28$ which yields $x = 174.4$

[15] Skip!

[16] Let the random variable X follow a normal distribution with mean $\mu = 48$ and variance $\sigma^2 = 60.84$ ($X \sim N(48, 60.84)$).

a)

$$P(X > 58) = P(Z > \frac{58-48}{\sqrt{60.84}}) = P(Z > 1.28) = 0.50 - 0.3997 = 0.1003$$

b)

$$P(36 < X < 60) = P(-1.54 < Z < 1.54) = 0.4382 + 0.4382 = 0.8764$$

c)

$$P(X > a) = 0.20 \Rightarrow P[Z > (a-48)/7.8] = 0.20 \Rightarrow (a-48)/7.8 = 0.84 \Rightarrow a = 54.552$$

d)

$$P(a < X < b) = 0.05 \Rightarrow P[(a-48)/7.8 < Z < (b-48)/7.8] = 0.05$$

$$\Rightarrow (a-48)/7.8 = -0.06 \text{ and } (b-48)/7.8 = 0.06 \Rightarrow a = 47.532 \text{ and } b = 48.468$$

e) $Y \sim N(110, 243.36)$ and therefore $P(Y < 110) = 0.5$