Problem Section 9

Learning Outcomes

The problems are designed to build conceptual understanding and problem-solving skills. The emphasis is on learning to find, evaluate and build confidence. The specific tasks include:

- Find the mean and variance of a continuous distribution

- Use the law of the unconscious probabilist for finding expected values of functions of continuous random variables

- Back up and support work with relevant explanations

Exercises

1. Suppose $X \sim Unif(0, 1)$, that is it has PDF $f(x) = 1 \quad 0 \leq x < 1$

a. Find $F_X(x)$, the CDF of X.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

b. Use the CDF from part a to find $P(0.25 \leq X < 0.5)$.

Note that $P(0.25 \leq X < 0.5) = P(X < 0.5) - P(X < 0.25)$. However, since $P(X = a) = 0$ for any continuous random variable, we have

$$\begin{aligned} P(0.25 \leq X < 0.5) &= P(X \leq 0.5) - P(X \leq 0.25) = F(0.5) - F(0.25) \\ &= 0.5 - 0.25 = 0.25 \end{aligned}$$

c. Let $Y = X - a$ where a is some positive number. Write the CDF of Y.

The CDF of Y is given by $F_Y(y) = P(Y \leq y)$ for all real numbers y. Since $Y = X - a$, we can write

We can now find F_Y as follows.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X - a \leq y) \\ &= P\left(X \leq y + a\right) \\ &= F_X(y + a) \\ &= \begin{cases} 0 & (y + a) < 0 \\ y + a & 0 \leq (y + a) < 1 \\ 1 & 1 \leq (y + a) \end{cases} \end{aligned}$$

d. Find a PDF for Y.

Therefore a PDF for Y is

$$f_Y(y) = 1 \qquad 0 \leq (y + a) < 1$$

2. The time X (in seconds) between two randomly selected consecutive cars in a traffic flow model is modeled with PDF

$$f(x) = \frac{k}{x^4}, a \leq x < \infty$$

where $a > 0$.

a. Determine the value of k in terms of a.

We choose k so that $\int\limits_{-\infty}^{\infty} f(x)dx = 1$

$$\begin{aligned}\int\limits_{-\infty}^{\infty} f(x)dx &= \int\limits_{-\infty}^{\infty} f(x)dx, \\ &= k \left. \frac{x^{-3}}{-3} \right|_a^{\infty} \\ &= k\frac{a^{-3}}{3}.\end{aligned}$$

Therefore $k = 3a^3$.

b. Find $E[X]$ in terms of a

$E[X] = \frac{3a}{2}$

c. Find $SD[X]$, the standard deviation of X in terms of a.

$SD[X] = a \cdot \sqrt{\frac{3}{2}}$

d. Find the median of the distribution in terms of a. That is, find the number q so that $P(X \leq q) = 0.5$.

$q = a \, \left(\frac{3}{2}\right)^{1/3}$

3. Suppose the amount of propellant, X put into a can of spray paint is a random variable with PDF

$$f(x) = 3x^2 \quad 0 \leq x < 1$$

Experience has shown that the largest surface area that can be painted by a can having X amount of propellant is twenty times the area of a circle generated by a radius of X ft. Can you expect to spray paint a 5 feet by 8 feet wall with one can?

The area of the wall is 40 square feet. The most we can spray paint when the can holds X amount of propellant is a surface area of $t(X) = 20\pi X^2$. Let us find $E[t(X)]$ using the law of the unconscious probabilist.

$$
\begin{aligned}
E\left[t(X)\right] &= \int_0^1 20\pi x^2 \, f(x)\, dx \\
&= 60\pi \int_0^1 x^4 \, dx \\
&= 60\pi \left(\frac{x^5}{5} \bigg|_0^1 \right) \\
&= 12\pi.
\end{aligned}
$$

```r
12*pi
```

```
## [1] 37.69911
```

4. Recovering small quantities of calcium in the presence of magnesium can be a difficult problem for an analytical chemist. Suppose the amount of calcium X (in mg) to be recovered is uniformly distributed between 4 and 7 mg.

   The amount of calcium recovered by one method is the random variable $W_1 = 0.2281 + (0.9948)X$. A second procedure has random variable $W_2 = -0.0748 + (1.0024)X$.

   The better technique should have mean as close to the mean of X and a variance which is as small as possible. Compare the two methods on the basis of the mean and variance.

   Since $X \sim Unif(4, 7)$, it has PDF

$$
f(x) = \frac{1}{3} \quad 4 \leq x < 7
$$

   We know from Lemma 11.1 that $E\left[X\right] = \frac{4+7}{2} = 5.5$ and $Var\left[X\right] = \frac{(7-4)^2}{12} = \frac{3}{4}$.

   Using the results for linear transformations in Lemma 7.3 we can say

$$
\begin{aligned}
E\left[W_1\right] &= 0.2281 + 0.9948 \times 5.5 \\
Var\left[W_1\right] &= 0.9948^2 \times \frac{3}{4}.
\end{aligned}
$$

   Similarly

$$
\begin{aligned}
E\left[W_2\right] &= -0.0748 + 1.0024 \times 5.5 \\
Var\left[W_1\right] &= 1.0024^2 \times \frac{3}{4}.
\end{aligned}
$$

   These are calculated below.

```r
exp_w1 <- 0.2281 + 0.9948*5.5; exp_w1
```

```
## [1] 5.6995
```

```r
var_w1 <- 0.9948^2*0.75; var_w1
```

```
## [1] 0.7422203
```

```r
exp_w2 <- -0.0748 + 1.0024*5.5; exp_w2
```

```
## [1] 5.4384
```

```r
var_w2 <- 1.0024^2*0.75; var_w2
```

```
## [1] 0.7536043
```

$E[W_1]$ is further away from the target of 5.5 mg. Additionally, it has a slightly smaller variance than W_2. So all in all, we should prefer W_2.

A single number summary that may be useful here is the mean squared error or MSE which can be calculated as $(E[W] - 5.5)^2 + Var[W]$ for each method. The first term is the square of the bias in the method and the second is the variability. The method with the smaller MSE is preferred.

```r
mse_w1 <- (exp_w1 - 5.5)^2 + var_w1; mse_w1
```

```
## [1] 0.7820205
```

```r
mse_w2 <- (exp_w2 - 5.5)^2 + var_w2; mse_w2
```

```
## [1] 0.7573989
```

As we can see, W_2 does better overall.