Assignment 4

STAT321

Connor Braun

November 24, 2021

Problem 1

Let

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $E[X]$ and $VAR[X]$.

$$E[X] = \int_0^1 xf(x)dx = \int_0^1 x(2x)dx = \int_0^1 2x^2dx = \frac{2x^3}{3}\bigg|_0^1 = \frac{2}{3}$$

Let $E[X] = \mu$. Then we have that

$$\begin{aligned} VAR[X] &= E[x^2] - \mu^2 = \int_0^1 x^2f(x)dx - \mu^2 = \int_0^1 x^2(2x)dx - \mu^2 = \int_0^1 2x^3dx - \mu^2 = \frac{x^4}{2}\bigg|_0^1 - \mu^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 \\ &= \frac{9}{18} - \frac{8}{18} = \frac{1}{18}. \end{aligned}$$

Solution: $E[X] = \frac{2}{3}$ and $VAR[X] = \frac{1}{18}$.

Problem 2

Let X be a continuous random variable such that $X \sim uniform(10, 20)$. If $X \geq 13$, compute $P(X \leq 18)$.

$$P(X \leq 18 | X \geq 13) = \frac{P(X \leq 18 \cap X \geq 13)}{P(X \geq 13)}.$$

By the law of total probability, we have that

$$\begin{aligned} P(X \leq 18) &= P(X \leq 18 \cap X \geq 13) + P(X \leq 18 \cap X < 13) \\ \Rightarrow P(X \leq 18 \cap X \geq 13) &= P(X \leq 18) - P(X < 13) \\ &= \frac{18-10}{20-10} - \frac{13-10}{20-10} \\ &= \frac{18-13-10+10}{20-10} \\ &= \frac{5}{10} = \frac{1}{2}. \end{aligned}$$

Substituting into the conditional probability equation:

$$\begin{aligned}
P(X \leq 18|X \geq 13) &= \frac{P(X \leq 18 \cap X \geq 13)}{P(X \geq 13)} \\
&= \frac{\frac{1}{2}}{P(X \geq 13)} \\
&= \frac{\frac{1}{2}}{1 - P(X < 13)} \\
&= \frac{\frac{1}{2}}{\frac{10}{10} - \frac{13-10}{20-10}} \\
&= \frac{\frac{1}{2}}{\frac{10}{10} - \frac{3}{10}} \\
&= \frac{1}{2} \times \frac{10}{7} \\
&= \frac{10}{14} = \frac{5}{7}.
\end{aligned}$$

Solution: Given that $X \geq 13$, $P(X \leq 18) = \frac{5}{7}$.

Problem 3

Let X be a normally-distributed continuous random variable with $\mu = 131.9$ and $\sigma = 20.3$ such that $X \sim \mathit{normal}(\mu, \sigma)$. Contextually, suppose X models the finishing time for half-marathon participants, and that both μ, σ are in units minutes.

a) Compute $P(X < 120)$, the probability that a given runner's finishing time is less than 2 hours.

Since X is continuous, we have that $P(X < 120) = P(X \leq 120)$, so we compute the answer directly with the following line in R:

```
pnorm(120, 131.9, 20.3)
```

Solution: we have that $P(X < 120) = 0.2788682$.

b) What is the fastest a runner can finish and still be in the slowest 20% of finishers?

We seek to find the finishing time x_0 such that $P(X \leq x_0) = 0.2$. We compute the answer directly with the following line in R:

```
qnorm(0.2, 131.9, 20.3)
```

Solution: we have that x_0 such that $P(X < x_0) = 0.2$ is $x_0 = 114.8151$ minutes.

c) What is the probability that exactly 4 of 10 finishers (randomly selected) had a finishing time of less than 2 hours?

From 3.a, we know that

$$P(X < 120) = 0.2788682.$$

Let Y be the discrete random variable for the number of runners out of 10 randomly selected participants who finished the half-marathon in under two hours. Then we have that

$$Y \sim \mathit{binomial}(10, P(X < 120)) = \mathit{binomial}(10, 0.2788682).$$

So we compute $P(Y = 4)$ directly.

$$P(Y = 4) = \binom{10}{4}(0.2788682)^4(1 - 0.2788682)^6 = 0.1786089231.$$

Solution: the probability that exactly 4 of 10 randomly selected participants finish in under 2 hours is $P(Y = 4) = 0.1786089231$.

Problem 4

Let X be a continuous random variable for the number of minutes it takes to observe 10 pedestrians using a particular crosswalk. Suppose that $\alpha = 10$ and $\beta = \frac{1}{2}$ such that

$$X \sim gamma(\alpha, \beta) = gamma(10, \frac{1}{2}).$$

a) Find $P(X < 10)$.

This probability is most easily computed directly using the following line in R:

```
pgamma(10, 10, 2)
```

Solution: The probability that it takes fewer than 10 minutes to observe 10 pedestrians is $P(X < 10) = 0.9950046$.

b) Compute $E[X]$.

For any $\xi \sim gamma(v, \omega)$, we have that $E[\xi] = v\omega$. Therefore

$$E[X] = \alpha\beta = 10 \times \frac{1}{2} = 5.$$

Solution: We'd expect it to take about 5 minutes to observe 10 pedestrians on the crosswalk.

Problem 5

Let $c \in \mathbb{R}$ and

$$f(x) = \begin{cases} cx^4 e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Find the value of c that makes $f(x)$ a valid probability density function.

For $f(x)$ to be a valid probability density function, we require that

$$\int_0^{+\infty} f(x)dx = 1.$$

We will use the fact that $\forall_\alpha \in \mathbb{Z}$ and $\beta \in \mathbb{R}$,

$$\int_0^{+\infty} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^\alpha (\alpha - 1)!.$$

Using our requirement and identity, we can compute c directly.

$$1 = \int_0^{+\infty} cx^4 e^{-\frac{x}{2}} dx = c \int_0^{+\infty} x^4 e^{-\frac{x}{2}} dx$$

$$
\begin{aligned}
&= c\left(2^5(4)!\right)\\
&= c \times 32 \times 24\\
&= 768c\\
\Rightarrow c &= \frac{1}{768}
\end{aligned}
$$

Solution: for $f(x)$ to be a valid probability density function, we require that $c = \frac{1}{768}$.

Problem 6

Let X be a continuous random variable for the length of a call to an emergency medical services dispatch centre in minutes. Suppose that $\beta = 2.25$ so that

$$X \sim exponential(\beta).$$

a) Compute $P(X > 4)$.

First, $P(X > 4) = 1 - P(X \leq 4)$, which is most easily computed using the following line in R:

```
1 - pexp(4, 1/2.25)
```

Solution: The probability that a call will last more than 4 minutes is $P(X > 4) = 0.1690133$.

b) Compute $P(X < 1|X < 3)$.

By the definition of conditional probability

$$
\begin{aligned}
P(X < 1|X < 3) &= \frac{P(X < 1 \cap X < 3)}{P(X < 3)}\\
&= \frac{P(X < 1)}{P(X < 3)}, \qquad \text{(since if } X < 1\text{, then } X < 3.)
\end{aligned}
$$

We can compute this ratio most easily using the following line in R:

```
pexp(1, 1/2,25)/pexp(3,1/2.25)
```

Solution: given that a particular call lasted less than 3 minutes, the probability that the same call lasted less than 1 minute is $P(X < 1|X < 3) = 0.5343126$.

c) Compute call duration x_0 such that $P(X > x_0) = 0.1$.

Firstly, we have that $P(X > x_0) = 1 - P(X \leq x_0)$, where $P(X \leq x_0) = 1 - 0.1 = 0.9$. Hence, we can then compute x_0 directly using the following line in R:

```
qexp(0.9, 1/2.25)
```

Solution: 10% of calls to the dispatch centre will last longer (or equivalently, 90% will be shorter) than 5.180816 minutes.

Problem 7

Let X be the discrete random variable for the number of times an office printer is used every hour. Suppose that $\lambda = 4.3$ such that

$$X \sim poisson(\lambda).$$

Next, let W be the continuous random variable for the duration of any given interval between printer uses in hours.

a) What type of distribution best models the behavior of W? What are the parameters for this example?

Solution: Let $\beta = \frac{1}{\lambda}$, then $W \sim exponential(\beta = \frac{1}{\lambda} = \frac{1}{4.3})$.

b) Compute $P(W > 1)$.

We have that $P(W > 1) = 1 - P(W \leq 1)$, which we can compute directly using the following line in R:

```
1 - pexp(1, 4.3)
```

Solution: $P(W > 1) = 0.01356856$.

c) Compute $P(W < \frac{1}{2})$.

$P(W < \frac{1}{2})$ can be computed directly using the following line of code in R:

```
pexp(1/2, 4.3)
```

Solution: $P(W < \frac{1}{2}) = 0.8835158$.

Problem 8

Let X be the continuous random variable modeling the proportion of a large sample population of people with an advanced stage of a certain type of cancer who are still alive after five years of monitoring. Suppose that $\alpha = 0.66$ and $\beta = 0.34$ such that

$$X \sim beta(\alpha, \beta).$$

What is the probability that more than 75% of the initial population are still alive after the five year period? In other words, compute $P(X > 0.75)$.

First, we have that $P(X > 0.75) = 1 - P(X \leq 0.75)$. Then we can compute $P(X > 0.75)$ most easily using the following line in R:

```
1 - pbeta(0.75, 0.66, 0.34)
```

Solution: The probability that more than 75% of the initial population survived the five year period is $P(X > 0.75) = 0.5243433$.

Problem 9

Let $c \in \mathbb{R}$ and $f(x, y)$ be the multivariate function given by

$$f(x,y) = \begin{cases} \frac{x}{5} + cy, & 0 < x < 1, 0 < y < 5 \\ 0, & elsewhere \end{cases}$$

Find the value of c which makes $f(x, y)$ a valid probability density function.

For $f(x, y)$ to be a valid probability density function, we require that

$$1 = \int_0^5 \int_0^1 f(x,y)dxdy$$

Which we can compute directly as follows

$$\begin{aligned}
1 &= \int_0^5 \int_0^1 f(x,y)dxdy \\
&= \int_0^5 \int_0^1 (\frac{x}{5} + cy)dxdy \\
&= \int_0^5 \int_0^1 \frac{x}{5}dxdy + c\int_0^5 \int_0^1 ydxdy \\
&= \int_0^5 \frac{x^2}{10}\bigg|_0^1 dy + c\int_0^5 yx\bigg|_0^1 dy \\
&= \int_0^5 \frac{1}{10}dy + c\int_0^5 ydy \\
&= \frac{y}{10}\bigg|_0^5 + c\left(\frac{y^2}{2}\bigg|_0^5\right) \\
&= \frac{1}{2} + c\frac{25}{2} \\
\Rightarrow 2 &= 25c + 1 \\
\Rightarrow c &= \frac{1}{25}
\end{aligned}$$

Solution: for $f(x,y)$ to be a valid probability density function, we require that $c = \frac{1}{25}$.