S520 Instructor's Solutions

Spring 2023 STAT-S 520

February 14th, 2023

Q1

1a.

Let X be a random variable that assigns the waiting time in minutes, hence $X \sim Uniform(0, 20)$ and $EX =$ 10 minutes (the x-coordinate value that corresponds to the middle point of the rectangle). Alternatively, observe that $f(x) = 0.05$ for $0 < x < 20$ and equal to zero otherwise, so

$$EX = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{20} xf(x)dx = \int_{0}^{20} x(0.05)dx = 0.05\left.\frac{x^2}{2}\right|_0^{20} = 0.05\left(\frac{400}{2} - \frac{0}{2}\right) = 10$$

1b.

Let Y be the random variable that assigns the waiting time in minutes. Since I can arrive at any point, it is equally likely that I arrive between the top of the hour and 20 minutes past (waiting at most 20 minutes), or between 20 and 40 minutes past (waiting again at most 20 minutes), or between 40 and 60 minutes past (waiting at most 40 minutes but not less than 20 minutes); i.e., it is twice as likely I only need to wait between 0 and 20 minutes than between 20 and 40 minutes. The pdf of Y looks like

$$f(x) = \begin{cases} 2c & 0 \le x < 20 \\ c & 20 \le x < 40 \\ 0 & \text{otherwise.} \end{cases}$$

for some constant c; and $20 * 2c + 20 * c = 1$ so $c = 1/60$. The pdf for Y is then

$$f(x) = \begin{cases} 1/30 & 0 \le x < 20 \\ 1/60 & 20 \le x < 40 \\ 0 & \text{otherwise.} \end{cases}$$

and the expected value is the weighted average of the x-coordinate values representing the middle points for both areas (rectangles), where the weights are the corresponding areas. We have:

$$10 * (20 * 1/30) + 30 * (20 * 1/60) = 10 * 2/3 + 30 * 1/3 = 50/3 \approx 16.67$$

minutes (solving with integrals should get you the same result)

1c.

This is an extension of 1b and the logic remains the same. We have now three intervals form 0 to 10, from 10 to 20, and form 20 to 30 (waiting times). the PDF is $3c$ for the first, $2c$ for the second, and c for the third interval respective, concluding that $c = 1/60$. The expected value is the weighted average of the balance point for each region (rectangles)

$$(10 - 0) \cdot 3c \cdot 5 + (20 - 10) \cdot 2c \cdot 15 + (30 - 20) \cdot c \cdot 5 \approx 11.67$$

Q2 ISI Section 5.6 exercise 3.

a The pdf is plotted in Figure 1. c must be nonnegative for f to be a pdf. The total area under f is the

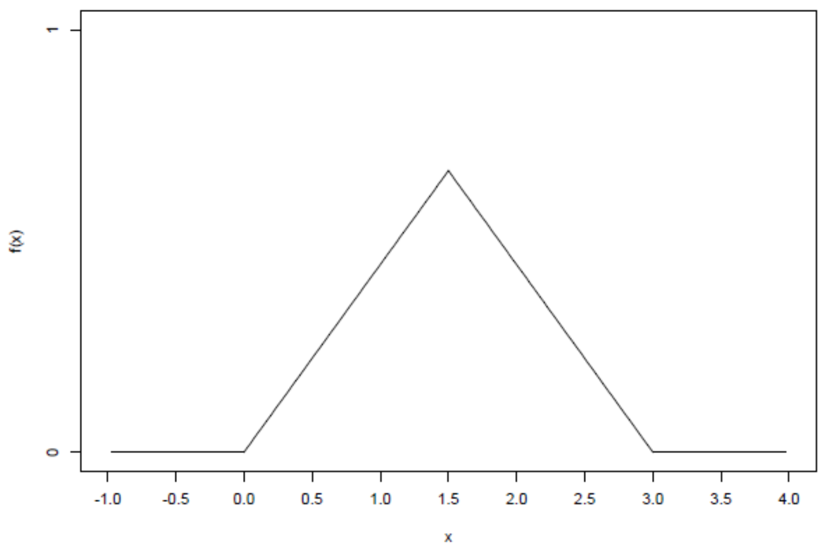


Figure 1: pdf for Exercise 5.6.3.

sum of two triangles rectangles (just draw a vertical line on the pdf at x = 1.5 to see them):

$$\frac{1.5 \cdot 1.5c}{2} + \frac{1.5 \cdot 1.5c}{2} = \frac{9}{4}c.$$

This has to equal 1 for a pdf, so

$$\begin{aligned} \frac{9}{4}c &= 1 \\ c &= \frac{4}{9}. \end{aligned}$$

b Looking at Figure 1, it's evident the f is symmetric about $x = 1.5$. So the expected value of X must be 1.5.

c $P(X > 2)$ is the area under the pdf between 2 and 3, which is the area of a triangle. The base of the triangle is $3 - 2 = 1$ and the height of the triangle is $f(2) = c = 4/9$. The area is $1/2 \times 1 \times 4/9 = 2/9$.

d Figure 3 plots the two pdfs on top of each other. Both have the same expected value: $EX = EY = 1.5$. However, the values of X tend to cluster a bit nearer 1.5 than do the values of Y. So Y has the larger variance.

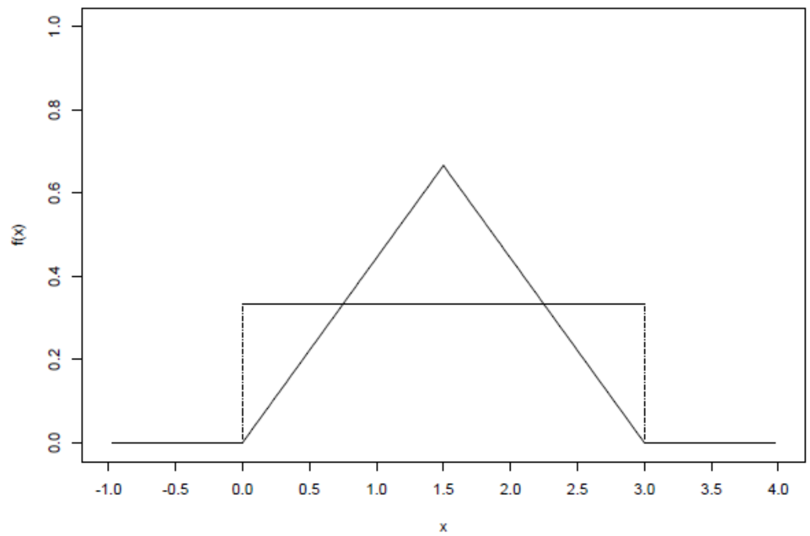


Figure 2: pdfs for Exercise 5.6.3(d).

e Firstly, if $y < 0$, then $F(y) = 0$, and if $y > 3$, then $F(y) = 1$.

If $0 \leq y \leq 1.5$, then $F(y)$ is the area of a triangle:

$$\begin{aligned} F(y) &= P(X \leq y) \\ &= \frac{1}{2} \cdot y \cdot cy \\ &= \frac{2y^2}{9}. \end{aligned}$$

If $1.5 \leq y \leq 3$, then $F(y)$ is one minus the area of a triangle. The base of the triangle is $3 - y$ and the height is $c(3 - y)$.

$$\begin{aligned} F(y) &= 1 - P(X > y) \\ &= 1 - \frac{1}{2} \cdot (3 - y) \cdot c(3 - y) \\ &= 1 - \frac{c}{2}(3 - y)^2 \\ &= 1 - \frac{2}{9}(3 - y)^2 \end{aligned}$$

One way of writing all of this down formally is:

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{2y^2}{9} & 0 \leq y < 1.5 \\ 1 - \frac{2}{9}(3 - y)^2 & 1.5 \leq y < 3 \\ 1 & y \geq 3 \end{cases}.$$

And here is the graph:

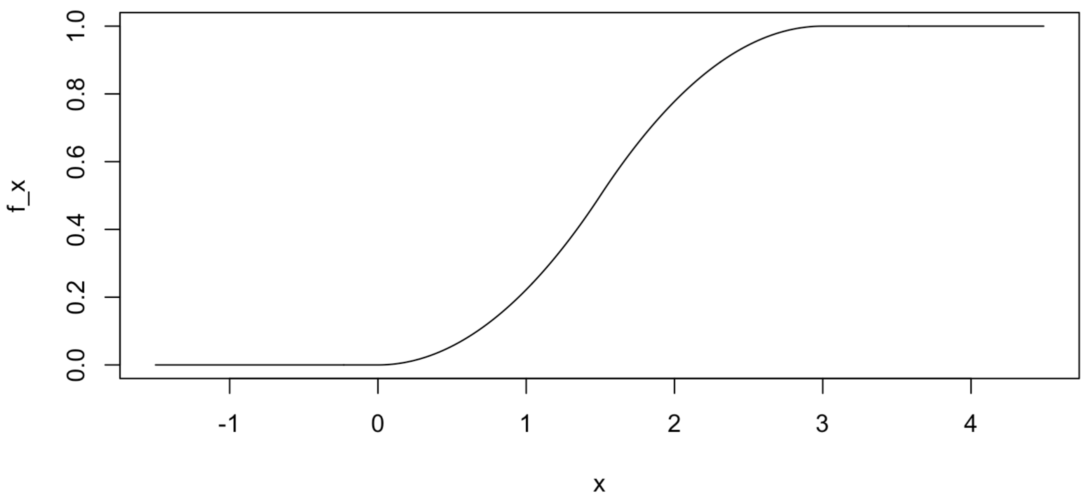


Figure 3: pdfs for Exercise 5.6.3(e).

Q3 ISI Section 5.6 exercise 7.

Let X be a normal random variable with mean $\mu = -5$ and standard deviation $\sigma = 10$.

a. P(X < 0)

```
pnorm(0,-5, 10)
```

```
## [1] 0.6914625
```

b. P(X > 5)

```
1 - pnorm(5,-5,10)
```

```
## [1] 0.1586553
```

c. P(-3< X < 7)

```
pnorm(7,-5,10) - pnorm(-3,-5,10)
```

```
## [1] 0.3056706
```

d. P(|X+ 5|<10) = P(-15 < x < 5)

```
pnorm(5,-5,10) - pnorm(-15,-5,10)
```

```
## [1] 0.6826895
```

e. P(|X-3|>2)

P(x > 5) + P(x < 1)

```
(1 - pnorm(5,-5,10)) + pnorm(1,-5,10)
```

```
## [1] 0.8844021
```

4

We have $X_1 \sim Normal(69.2, 2.5^2)$ and $X_2 \sim Normal(63.8, 2.7^2)$.

a. $P(X_1 > 72) = 1 - P(X_1 <= 72) = 1 - F(72)$. Using R

```
1 - pnorm(72, 69.2, 2.5)
```

```
## [1] 0.1313569
```

b. Y follows exactly a normal distribution, with mean $EY = EX_1 + EX_2 = 69.2 + 63.8 = 133$ and $VarY = VarX_1 + VarX_2 = 2.5^2 + 2.7^2 = 13.54$, or $Y \sim Normal(133, 13.54)$
c. $P(Y > 144) = 1 - P(Y <= 144) = 1 - F_Y(144)$

```
1 - pnorm(144, 133, sqrt(13.54))
```

```
## [1] 0.001397651
```

d. Yes, D is a random variable that follows the normal distribution, with mean $ED = EX_1 + E((-1)X_2) = 69.2 - 63.8 = 5.4$ and $VarD = VarX_1 + Var((-1)X_2) = 2.5^2 + (-1)^2 \cdot 2.7^2 = 13.54$, or $D \sim Normal(5.4, 13.54)$
e. $P(X_1 < X_2) = P(X_1 - X_2 < 0) = P(D < 0) = P(D \leq 0) = F(0)$

```
pnorm(0, 5.4, sqrt(13.54))
```

```
## [1] 0.07111714
```

5

We need some reference to determine whether this looks closer or not. Histograms can be used for visual inspection but since we've talked about the 68-95-99.t rule we can also use it as our reference (but other measures could be used as well)

So, let's focus on the 95 part of the rule: if $X \sim Normal(\mu, \sigma^2)$ then the probability that X assigns values within 2 standard deviation from the mean should be about 95% or

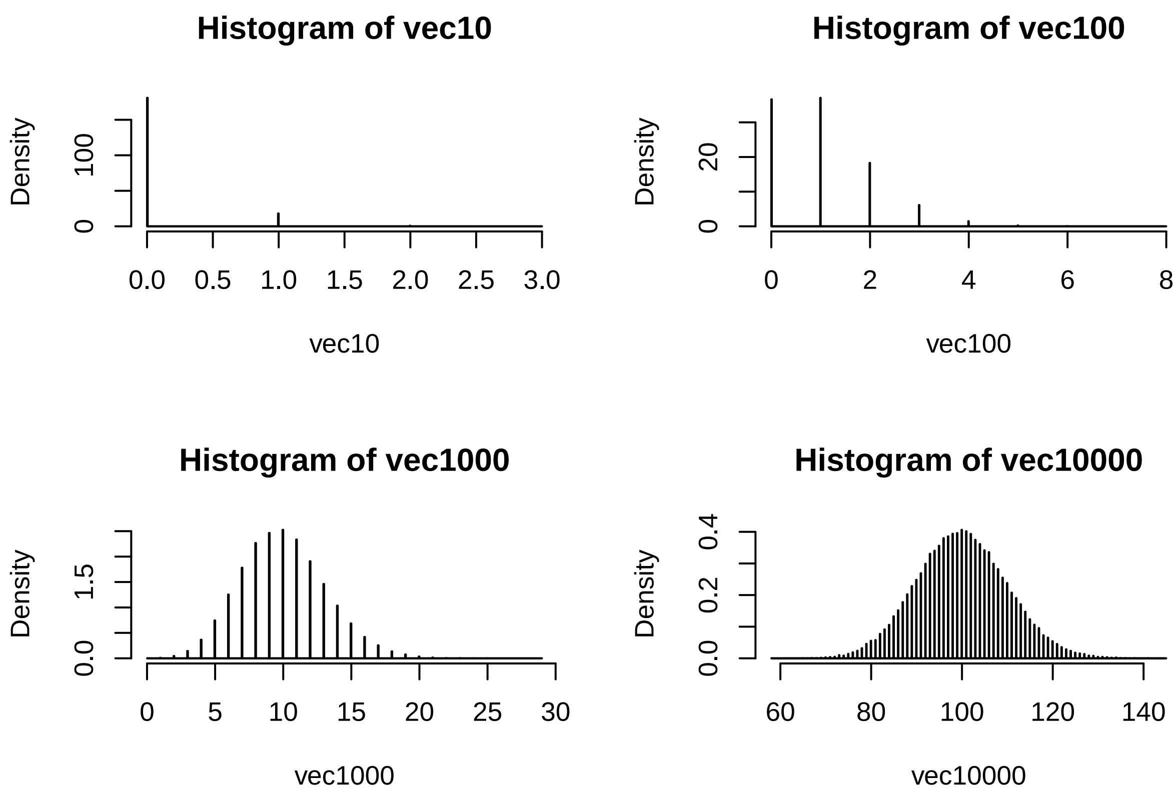
$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$$

```
vec10 = rbinom(10^5, 10, 0.01)
vec100 = rbinom(10^5, 100, 0.01)
vec1000 = rbinom(10^5, 1000, 0.01)
vec10000 = rbinom(10^5, 10000, 0.01)
vec100000 = rbinom(10^5, 100000, 0.01)
```

```
n = 10
p=0.01
mu = n*p
sigma = sqrt(n*p*(1-p))
p10 = mean(vec10 > mu - 2*sigma & vec10 < mu + 2*sigma )

n = 100
p=0.01
mu = n*p
sigma = sqrt(n*p*(1-p))
p100 = mean(vec100 > mu - 2*sigma & vec100 < mu + 2*sigma )


n = 1000
p=0.01
mu = n*p
sigma = sqrt(n*p*(1-p))
p1000 = mean(vec1000 > mu - 2*sigma & vec1000 < mu + 2*sigma )


n = 10000
p=0.01
mu = n*p
sigma = sqrt(n*p*(1-p))
p10000 = mean(vec10000 > mu - 2*sigma & vec10000 < mu + 2*sigma )

c(p10, p100, p1000, p10000)
```

```
## [1] 0.90466 0.92032 0.96312 0.94985
```

```
op =par(mfrow = c(2,2))
hist(vec10, freq = F, breaks = 1000)
hist(vec100, freq = F, breaks = 1000)
hist(vec1000, freq = F, breaks = 1000)
hist(vec10000, freq = F, breaks = 1000)
```




```
par(op)
```

At around $n = 10^3$ the curve seems to be close to a normal, although it is still slightly right skewed (longer right tail), while at $n = 10^5$ results really looks like a normal (and the probability within 2 standard deviations is also what you would expect).