140.651 Problem Set 2 Solutions

Problem 1

Using the rules of expectations, prove that $Var(X) = E[X^2] - E[X]^2$ where $Var(X) = E[(X-\mu)^2]$.

$$\begin{aligned}
E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\
&= E[X^2] - E[2\mu X] + E[\mu^2] && \text{by linearity of expectation} \\
&= E[X^2] - 2\mu E[X] + \mu^2 && \text{since 2 and } \mu \text{ are not random} \\
&= E[X^2] - 2E[X]^2 + E[X]^2 && \text{since } \mu = E[X] \\
&= E[X^2] - E[X]^2
\end{aligned}$$

Problem 2

Let $g(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)$ where f_1, f_2, and f_3 are densities with associated means and variances μ_1, σ_1^2; μ_2, σ_2^2; μ_3, σ_3^2, respectively. Here $\pi_1, \pi_2, \pi_3 \geq 0$ and $\sum_{i=1}^3 \pi_i = 1$. Show that g is a valid density. What is it's associated mean and variance?

To show that g is a valid density, we need to show that (1) $g \geq 0$ for all x and (2) $\int g(x)dx = 1$.

(1) Since $\pi_1, \pi_2, \pi_3 \geq 0$ and f_1, f_2, f_3 are valid densities, $g(x) \geq 0$ for all x.

(2) Verifying that $\int g(x)dx = 1$,

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} \pi_1 f_1 + \pi_2 f_2 + \pi_3 f_3 dx = \pi_1 \int_{-\infty}^{\infty} f_1 dx + \pi_2 \int_{-\infty}^{\infty} f_2 dx + \pi_3 \int_{-\infty}^{\infty} f_3 dx = \pi_1 + \pi_2 + \pi_3 = 1$$

Thus, $g(x)$ is a valid density. The associated mean is given by,

$$\begin{aligned}
E_g[X] &= \int_{-\infty}^{\infty} x g(x)dx = \pi_1 \int_{-\infty}^{\infty} x f_1(x)dx + \pi_2 \int_{-\infty}^{\infty} x f_2(x)dx + \pi_3 \int_{-\infty}^{\infty} x f_3(x)dx \\
&= \pi_1 E_{f_1}[X] + \pi_2 E_{f_2}[X] + \pi_3 E_{f_3}[X] \\
&= \pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3 \qquad (1)
\end{aligned}$$

To calculate the variance, recall from problem (1) that $Var(X) = E[X^2] - E[X]$. Consequently, we know that $E[X^2] = Var(X) + E[X]^2$ and thus,

$$E_{f_1}[X^2] = \sigma_1^2 + \mu_1^2 \qquad E_{f_2}[X^2] = \sigma_2^2 + \mu_2^2 \qquad E_{f_3}[X^2] = \sigma_2^2 + \mu_2^2$$

Furthermore,

$$\begin{aligned}
E_g[X^2] &= \int_{-\infty}^{\infty} x^2 g(x)dx \\
&= \pi_1 \int_{-\infty}^{\infty} x^2 f_1(x)dx + \pi_2 \int_{-\infty}^{\infty} x^2 f_2(x)dx + \pi_3 \int_{-\infty}^{\infty} x^2 f_3(x)dx \\
&= \pi_1 E_{f_1}[X^2] + \pi_2 E_{f_2}[X^2] + \pi_3 E_{f_3}[X^2] \\
&= \pi_1(\sigma_1^2 + \mu_1^2) + \pi_2(\sigma_2^2 + \mu_2^2) + \pi_3(\sigma_2^2 + \mu_2^2) \qquad (2)
\end{aligned}$$

Combining (1) and (2) we obtain,

$$\text{Var}_g[X] = E_g[X^2] - E_g[X]^2 = \pi_1(\sigma_1^2 + \mu_1^2) + \pi_2(\sigma_2^2 + \mu_2^2) + \pi_3(\sigma_2^2 + \mu_2^2) - (\pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3)^2$$

Problem 3

Suppose that a density is of the form $(k+1)x^k$ for some constant $k > 1$ and $0 < x < 1$.

a. What is the mean of this distribution?

Let X be a random variable with density $f(x) = (k+1)x^k$ for some constant $k > 1$ and $0 < x < 1$. Then, the mean is given by,

$$E[X] = \int_0^1 x(k+1)x^k dx = \int_0^1 (k+1)x^{k+1} dx = \frac{k+1}{k+2}x^{k+1}\Big|_0^1 = \frac{k+1}{k+2}$$

b. What is the variance?

Note that,

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

Using the identity derived in problem 1, the variance is given by,

$$Var(X) = E[X^2] - E[X]^2 = \frac{k+1}{k+3} - \left(\frac{k+1}{k+2}\right)^2 = \frac{k+1}{(k+3)(k+2)^2}$$

Problem 4

Suppose that the time in days until hospital discharge for a certain patient population follows a density $f(x) = \frac{1}{3.3}\exp(-x/3.3)$ for $x > 0$.

a. Find the mean and variance of this distribution.

We will use the solutions for the mean and variance derived in the general case below (part b). Plugging in $\beta = 3.3$, we get

$$E[X] = \beta = 3.3$$
$$Var(X) = \beta^2 = 3.3^2 = 10.89$$

b. The general form of this density (the exponential density) is $f(x) = \frac{1}{\beta}\exp(-x/\beta)$ for $x > 0$ for a fixed value of β. Calculate the mean and variance of this density.

To calculate the mean and variance of the exponential density, we can use integration by parts. That is,

$$\int_a^b u\ dv = uv\Big|_a^b - \int_a^b v\ du$$

The mean of the exponential density is,

$$\begin{aligned}
E[X] &= \int_0^\infty x\frac{1}{\beta}e^{-x/\beta}dx \\
&= -xe^{-x/\beta}\Big|_0^\infty + \int_0^\infty e^{-x/\beta}dx && \text{integration by parts with } u = x \text{ and } dv = -\frac{1}{\beta}e^{-x/\beta} \\
&= -xe^{-x/\beta}\Big|_0^\infty + -\beta e^{-x/\beta}\Big|_0^\infty \\
&= (0-0) + (0+\beta) = \beta
\end{aligned}$$

To calculate the variance, recall that $Var(X) = E[X^2] - E[X]^2$. Again, using integration by parts,

$$\begin{aligned}
E[X^2] &= \int_0^\infty x^2 \frac{1}{\beta} e^{-x/\beta} dx \\
&= -x^2 e^{-x/\beta}\Big|_0^\infty + \int_0^\infty 2x e^{-x/\beta} dx && \text{integration by parts with } u = x^2 \text{ and } dv = -\frac{1}{\beta} e^{-x/\beta} \\
&= -x^2 e^{-x/\beta}\Big|_0^\infty + 2\beta \int_0^\infty x \frac{1}{\beta} e^{-x/\beta} dx \\
&= -x^2 e^{-x/\beta}\Big|_0^\infty + 2\beta E[X] && \text{recognizing that } \int_0^\infty x e^{-x\beta} dx = E[X] \\
&= (0 - 0) + 2\beta^2 = 2\beta^2
\end{aligned}$$

Thus, the variance of the exponential density is given by,

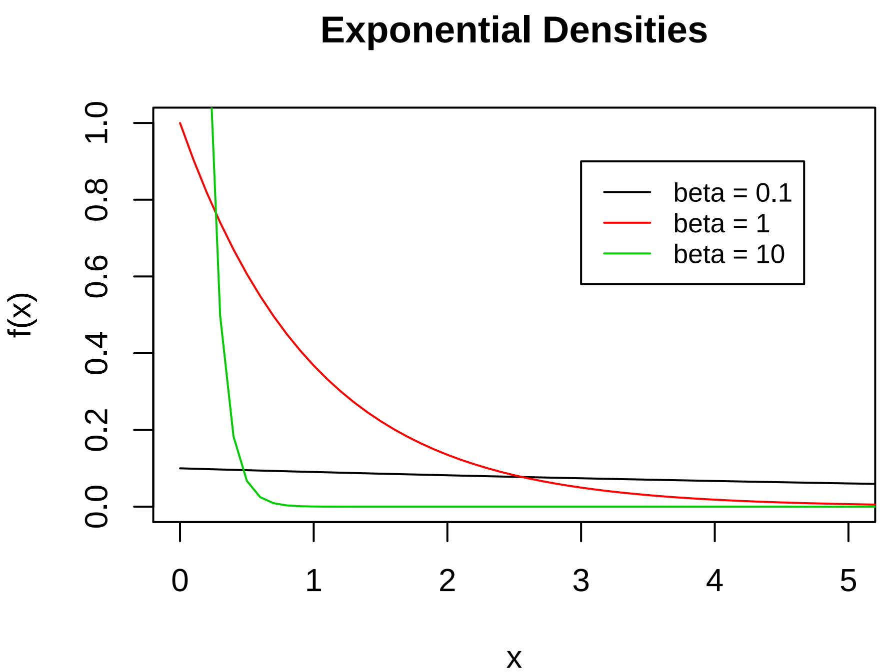
$$Var(X) = E[X^2] - E[X]^2 = 2\beta^2 - \beta^2 = \beta^2$$

c. Plot the exponential pdf for $\beta = 0.1, 1, 10$.

We can use the `dexp()` function to plot the exponential pdf.

```
x <- seq(0, 1000, by = 0.1)

plot(x, dexp(x, rate = 0.1), xlim = c(0, 5), ylim = c(0,1), type = "l",
     main = "Exponential Densities", xlab = "x", ylab = "f(x)")
lines(x,dexp(x, rate = 1), col=2)
lines(x,dexp(x, rate = 10), col=3)
legend(x = 3, y = 0.9, legend = c("beta = 0.1", "beta = 1",
                                   "beta = 10"),
       col = c(1, 2, 3), lty = c("solid"),
       cex = 0.8)
```



Problem 5

The Gamma density is given by

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta) \quad \text{for} \quad x > 0$$

for fixed values of α and β.

a. Derive the mean and variance of the gamma density. You can assume the fact (proved in HW 1) that the density integrates to 1 for any $\alpha > 0$ and $\beta > 0$.

Let f be the gamma density given above, and recall that

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

The mean of the gamma density is given by,

$$\begin{aligned}
E[X] &= \int_0^\infty x \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty \left(\frac{x}{\beta}\right)^\alpha e^{-x/\beta} dx \\
&= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty u^\alpha e^{-u} du && \text{change of variables for } u = x/\beta \text{ so } du = \frac{1}{\beta} dx \\
&= \frac{\beta \Gamma(\alpha+1)}{\Gamma(\alpha)} && \text{since } \Gamma(\alpha+1) = \int_0^\infty u^\alpha e^{-u} du \\
&= \frac{\alpha \Gamma(\alpha) \beta}{\Gamma(\alpha)} && \text{since } \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \\
&= \alpha\beta
\end{aligned}$$

To calculate the variance, recall that $Var(X) = E[X^2] - E[X]^2$. Thus,

$$\begin{aligned}
E[X^2] &= \int_0^\infty x^2 \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx \\
&= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha+1} e^{-x/\beta} dx \\
&= \frac{\beta^2}{\Gamma(\alpha)} \int_0^\infty u^{\alpha+1} e^{-u} dx && \text{change of variables for } u = x/\beta \text{ so } du = \frac{1}{\beta} dx \\
&= \frac{\beta^2 \Gamma(\alpha+2)}{\Gamma(\alpha)} && \text{since } \Gamma(\alpha+2) = \int_0^\infty u^{\alpha+1} e^{-u} du \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)\beta^2}{\Gamma(\alpha)} && \text{since } \Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1) \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)\beta^2}{\Gamma(\alpha)} && \text{since } \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \\
&= (\alpha+1)\alpha\beta^2
\end{aligned}$$

Hence, the variance of the gamma density is given by,

$$Var(X) = E[X^2] - E[X]^2 = (\alpha+1)\alpha\beta^2 - (\alpha\beta)^2 = \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

b. The Chi-squared density is the special case of the Gamma density where $\beta = 2$ and $\alpha = p/2$ for some fixed value of p (called the "degrees of freedom"). Calculate the mean and variance of the Chi-squared density.

Using the formulas derived in part (a) for the mean and variance when $\beta = 2$ and $\alpha = p/2$, we have

$$E[X] = \alpha\beta = \frac{p}{2} \cdot 2 = p \qquad \text{and} \qquad Var(X) = \alpha\beta^2 = \frac{p}{2} \cdot 2^2 = 2p$$

Problem 6

The Beta density is given by

$$\frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for} \quad 0 < x < 1$$

and $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

a. Derive the mean of the beta density. Note the following is useful for simplifying results: $\Gamma(c+1) = c\Gamma(c)$ for $c > 0$.

$$\begin{aligned}
E[X] &= \int_0^1 x \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} && \text{since } B(\alpha+1,\beta) = \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \Bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} && \text{since } B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta) \\
&= \frac{\alpha\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta+1)\Gamma(\alpha)\Gamma(\beta)} && \text{since } \Gamma(c+1) = c\Gamma(c) \text{ for } c > 0 \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

b. Derive the variance of the beta density.

To calculate the variance, recall that $Var(X) = E[X^2] - E[X]^2$. Thus,

$$\begin{aligned}
E[X^2] &= \int_0^1 x^2 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} && \text{since } B(\alpha+2,\beta) = \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \Bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} && \text{since } B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta) \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)\Gamma(\beta)}{(\alpha+\beta+1)\Gamma(\alpha+\beta+1)} \Bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} && \text{since } \Gamma(c+1) = c\Gamma(c) \text{ for } c > 0 \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \Bigg/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} && \text{since } \Gamma(c+1) = c\Gamma(c) \text{ for } c > 0 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

Hence, the variance of the beta density is given by,

$$Var(X) = E[X^2] - E[X]^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Problem 7

The Poisson mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for} \quad x = 0, 1, 2, 3, \ldots$$

a. Derive the mean of this mass function.

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} && \text{we can drop the } x = 0 \text{ term since } x\frac{e^{-\lambda}\lambda^x}{x!} = 0 \text{ when } x = 0 \\
&= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} \\
&= e^{-\lambda}\lambda\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda}\lambda\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} && \text{change of variables for } i = x - 1 \\
&= \lambda e^{-\lambda}e^{\lambda} = \lambda && \text{since } e^{\lambda} = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \text{ (Taylor expansion of } e^{\lambda})
\end{aligned}$$

b. Derive the variance of this mass function. Hint, consider $E[X(X-1)]$.

To calculate the variance, recall that $Var(X) = E[X^2] - E[X]^2$. To find $E[X^2]$, we can use the hint given since $E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X]$ (by linearity of expectation), and we already know $E[X]$.

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} && \text{we can drop the } x = 0, 1 \text{ terms since } x\frac{e^{-\lambda}\lambda^x}{x!} = 0 \text{ when } x = 0, 1 \\
&= \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-2)!} \\
&= \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \\
&= \lambda^2 e^{-\lambda}\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} && \text{change of variables for } i = x - 2 \\
&= \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2 && \text{since } e^{\lambda} = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \text{ (Taylor expansion of } e^{\lambda})
\end{aligned}$$

Thus, $E[X^2] = E[X(X-1)] + E[X] = \lambda^2 + \lambda$, and

$$Var(X) = E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Problem 8

Suppose that, for a randomly drawn subject from a particular population, the proportion of a their skin that is covered in freckles follows a uniform density (constant between 0 and 1).

a. What is the expected percentage of a (randomly selected) person's body that is covered in freckles? (Show your work.)

The expected percentage is given by,

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

b. What is the variance? (Show your work.)

The variance is given by,

$$Var(X) = E[X^2] - E[X]^2 = \int_0^1 x^2 dx - \left(\frac{1}{2}\right)^2 = \frac{1}{3}x^3 \bigg|_0^1 - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Problem 9

You have an MP3 player with a total of 1000 songs stored on it. Suppose that songs are played randomly *with replacement*. Let X be the number of songs played until you hear a repeated song.

a. What values can X take, and with what probabilities?

Let x be the number of songs played until you hear a repeated song. That is, $x-1$ distinct songs have been played and the xth song is a repeat of one of the previous $x-1$ songs. Note that,

- We must play at least 1 song before we can hear a repeated song, so $x \geq 2$.
- Since there are only 1000 unique songs, by song 1001, at least one song must be repeated.

Thus, X can take on values from 2 to 1001.

To determine the probabilities, suppose the xth song played is the first repeated song. Then,

$$P(X = x) = \frac{\text{\# ways } x \text{ songs are played with the } x\text{th song being the only repeated song}}{\text{\# ways to play } x \text{ songs}}$$

Let us first consider the denominator.

- For the first song played, we have 1000 options.
- Similarly, for the second song, we have 1000 options since we have no restrictions on repetition.
- Continuing, for th xth song played, we still have 1000 options.

Thus, the total number of ways to play x songs is 1000^x.

Now consider the numerator.

- For the first song played, we have 1000 options.
- However, for the second song, we only have 999 options since the first song cannot be played again, otherwise we would have a repetition.

- Continuing for the $(x-1)$th song, we have $1000-(x-2)$ options since $x-2$ songs have already been played.
- For the xth song there are $x-1$ options since it must repeat one of the previous $x-1$ songs.

Thus, the total number of ways to play x songs with no repeats is

$$[1000 \cdot (1000-1) \cdots (1000-(x-2))](x-1) = \frac{1000!}{(1000-(x-1))!} \cdot (x-1)$$

Combining, we have

$$P(X=x) = \frac{1000!(x-1)}{(1000-(x-1))! \cdot 1000^x} \qquad \text{for } x = 2, 3, ..., 1001$$

Note that the closed form solution for $P(X=x)$ is not very tractable due to the large factorials. For the sake of calculating the expectation and variance, it is more practical to define $P(X=x)$ recursively.

- For $x=2$, we pick 1 song out of 1000 for the first song played, and the second song must repeat that first song, so
$$P(X=2) = \frac{1}{1000} * 1$$
- For $x=2$, we can use Bayes' rule conditioning on the event that $P(X>2)$, so
$$P(X=3) = P(X=3|X>2)P(X>2) = \frac{2}{1000} * [1-P(X=2)]$$
where $P(X=3|X>2) = \frac{2}{1000}$ because we know that the first two songs are not repeated as $X>2$, and for $X=3$, the third song must repeat one of the previous two songs so $P(X=3|X>2) = \frac{2}{1000}$.
- Continuing, we have for $x = 2, 3, \cdots, 1001$,
$$P(X=x) = P(X=x|X>x-1)P(X>x-1) = \frac{x-1}{1000} \times \left[1 - \sum_{k=2}^{x-1} P(X=k)\right]$$

b. What is the expected value for X?

We can use R to help us compute the expected value.

```
x   <- 1:1001          # vector of possible x values
p.x <- rep(0, 1001)    # vector of P(X = x)

# Fill in probabilities using our recursive definition
for(k in 2:1001){
  p.x[k] <- (k-1)/(1000) * (1 - sum(p.x[1:k-1]))
}

# Calculate E[X]
EX <- sum(x*p.x)
EX
```

```
[1] 40.30321
```

c. What is the variance for X?

Recall that $Var(X) = E[X^2] - E[X]^2$. Using R, we can calculate the variance.

```
EX2 <- sum(x^2*p.x)    # Calculate E[X^2]
VarX <- EX2 - EX^2     # Calculate Var(X)
VarX
```

```
[1] 415.9542
```

Problem 10

Moved to Problem Set 3

Problem 11

Assume that an act of intercourse between an HIV infected person and a non-infected person results in a 1/500 probability of spreading the infection. How many acts of intercourse would an uninfected person have to have with an infected persons to have a 10% probability of obtaining an infection? State the assumptions of your calculations.

Assume that infection is independent across acts of intercourse. Then we can model the number of acts of intercourse until infection as a geometric random variable X with probability 1/500. To determine the number of acts an uninfected person would have to have to have at least a 10% probability of obtaining an infection, we can find the value $x_{0.1}$ such that $P(X < x_{0.1}) \geq 0.1$. From homework 1, we know that

$$P(X < x) = 1 - \left(1 - \frac{1}{500}\right)^x$$

Setting $P(X < x) = 0.1$ and solving for x, we have,

$$\begin{aligned} 0.1 = 1 - \left(1 - \frac{1}{500}\right)^x \implies 0.9 &= \left(1 - \frac{1}{500}\right)^x \\ \implies \log(0.9) &= x \log\left(1 - \frac{1}{500}\right) \\ \implies x &= \frac{\log(0.9)}{\log\left(1 - \frac{1}{500}\right)} = 52.63 \end{aligned}$$

Since x takes on integer values, the uninfected individual must participate in at least 53 acts of intercourse with an infected individual to have a 10% probability of obtaining an infection.

Problem 12

Moved to Problem Set 3

Problem 13

Moved To Problem Set 3

Problem 14

Moved To Problem Set 3

Problem 15

Moved To Problem Set 3

Problem 16

Given below are the sexes of the children of 7,745 families of 4 children recorded in the archives of the Genealogical Society of the Church of Jesus Christ of Latter Day Saints in Salt Lake City, Utah. M indicates a male child and F indicates a female child.

Sequence	Freq	Sequence	Freq
MMMM	537	MFFM	526
MMMF	549	FMFM	498
MMFM	514	FFMM	490
MFMM	523	MFFF	429
FMMM	467	FMFF	451
MMFF	497	FFMF	456
MFMF	486	FFFM	441
FMMF	473	FFFF	408

a. Estimate the probability distribution of the number of male children, say X, in these families using the data below by calculating proportions.

We can estimate the distribution of the number of male children by using the observed proportions of families given in the table above.

x	$P(X = x)$	
0	$\frac{408}{7745}$	≈ 0.053
1	$\frac{429+451+456+441}{7745} = \frac{1777}{7745}$	≈ 0.229
2	$\frac{497+486+473+526+498+490}{7745} = \frac{2970}{7745}$	≈ 0.383
3	$\frac{549+514+523+467}{7745} = \frac{2053}{7745}$	≈ 0.265
4	$\frac{537}{7745}$	≈ 0.069

b. Find the expected value of X.

By the definition of expected value,

$$E[X] = \sum_{x=0}^{4} xP(X = x) = 0(0.053) + 1(0.229) + 2(0.383) + 3(0.265) + 4(0.069) = 2.066$$

c. Find the variance of X.

Recall that $Var(X) = E[X^2] - E[X]^2$.

$$E[X^2] = \sum_{x=0}^{4} x^2 P(X = x) = 0^2(0.053) + 1^2(0.229) + 2^2(0.383) + 3^2(0.265) + 4^2(0.069) = 5.25$$

Hence, $Var(X) = E[X^2] - E[X]^2 = 5.25 - 2.066^2 = 0.982$.

d. Find the probability distribution of $\hat{p}$, where $\hat{p}$ is the proportion of children in each family who are male. Find the expected value of $\hat{p}$ and the variance of $\hat{p}$.

Observe that $\hat{p} = \frac{X}{4}$. Thus, the probability distribution is given by,

n	$P(\hat{p} = n)$	
0	$\frac{408}{7745}$	≈ 0.053
0.25	$\frac{429+451+456+441}{7745} = \frac{1777}{7745}$	≈ 0.229
0.5	$\frac{497+486+473+526+498+490}{7745} = \frac{2970}{7745}$	≈ 0.383
0.75	$\frac{549+514+523+467}{7745} = \frac{2053}{7745}$	≈ 0.265
1	$\frac{537}{7745}$	≈ 0.069

The expected value is,

$$E[\hat{p}] = E\left[\frac{X}{4}\right] = \frac{1}{4}E[X] = \frac{1}{4} \times 2.066 = 0.5165$$

And the variance is,

$$Var(\hat{p}) = Var\left(\frac{X}{4}\right) = \frac{1}{16}Var(X) = \frac{1}{16} \times 0.982 = 0.0614$$

Problem 17

Quality control experts estimate that the time (in years) until a specific electronic part from an assembly line fails follows (a specific instance of) the **Pareto** density

$$\frac{3}{x^4} \qquad \text{for } 1 < x < \infty.$$

a. What is the average failure time for components from this density? (Show your work.)

From the expectation for the general form of the Pareto density derived in part (c) below, the expectation for this instance of the Pareto density with $\alpha = 1$ and $\beta = 3$ is,

$$E[X] = \frac{\beta\alpha}{\beta - 1} = \frac{3 * 1}{3 - 1} = \frac{3}{2} = 1.5$$

b. What is the variance? (Show your work.)

From the expectation for the general form of the Pareto density derived in part (c) below, the variance for this instance of the Pareto density with $\alpha = 1$ and $\beta = 3$ is,

$$Var(X) = \frac{\beta\alpha}{(\beta-1)^2(\beta-2)} = \frac{3 * 1}{(3-1)^2(3-2)} = \frac{3}{2^2 * 1} = \frac{3}{4} = 0.75$$

c. The general form of the Pareto density is given by $\frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $0 < \alpha < x$ and $\beta > 0$ (for fixed α and β). Calculate the mean and variance of the general Pareto density.

Let X be a random variable with a Pareto density $\frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $0 < \alpha < x$ and $\beta > 0$ (for fixed α and β). Then, the mean is given by,

$$E[X] = \int_\alpha^\infty x\frac{\beta\alpha^\beta}{x^{\beta+1}}dx = \beta\alpha^\beta \int_\alpha^\infty x^{-\beta}dx = \beta\alpha^\beta \left(\frac{1}{1-\beta}x^{-\beta+1}\right)\Bigg|_\alpha^\infty$$

Note that if $\beta \leq 1$, then $x^{\beta+1}$ is positive, so the integral does not converge. Thus, we have

$$E[X] = \begin{cases} \infty & \text{if } \beta \leq 1 \\ \frac{\beta\alpha^\beta}{1-\beta}\left(0 - \alpha^{-\beta+1}\right) = \frac{\beta\alpha}{\beta-1} & \text{if } \beta > 1 \end{cases}$$

To calculate the variance, we have $Var(X) = E[X^2] - E[X]^2$ and,

$$E[X^2] = \int_\alpha^\infty x^2 \frac{\beta\alpha^\beta}{x^{\beta+1}} dx = \beta\alpha^\beta \int_\alpha^\infty x^{-\beta+1} dx = \beta\alpha^\beta \left(\frac{1}{2-\beta} x^{-\beta+2}\right)\Bigg|_\alpha^\infty$$

Again, note that if $\beta \leq 2$, then $x^{\beta+2}$ is positive, so the integral does not converge. Thus, we have

$$E[X^2] = \begin{cases} \infty & \text{if } \beta \leq 2 \\ \frac{\beta\alpha^\beta}{2-\beta}\left(0 - \alpha^{-\beta+2}\right) = \frac{\beta\alpha^2}{\beta-2} & \text{if } \beta > 2 \end{cases}$$

Combining, we get that

$$Var(X) = E[X^2] - E[X]^2 = \begin{cases} \infty & \text{if } \beta \leq 2 \\ \frac{\beta\alpha^2}{\beta-2} - \left(\frac{\alpha\beta}{\beta-1}\right)^2 = \frac{\beta\alpha}{(\beta-1)^2(\beta-2)} & \text{if } \beta > 2 \end{cases}$$

Problem 18

You are playing a game with a friend where you flip a coin and if it comes up heads you give her a dollar and if it comes up tails she gives you a dollar. You play the game ten times.

a. What is the expected total earnings for you? (Show your work; state your assumptions.)

Assume that,

- The games are independent and identically distributed,
- The tossed coin is fair, i.e. $P(\text{heads}) = P(\text{tails}) = \frac{1}{2}$

Let X_i denote the result of the ith game for $i = 1, \cdots, 10$. That is,

$$X_i = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

where $+1$ indicates that you gained a dollar by winning the coin toss, and -1 indicates that you lost a dollar because you lost the coin toss. Note that for $i = 1, 2, \cdots, 10$,

$$E[X_i] = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0$$
$$Var(X_i) = E[X^2] - E[X]^2 = \left[1^2 \times \frac{1}{2} + (-1)^2 \times \frac{1}{2}\right] - 0^2 = 1$$

Furthermore, let $G = \sum_{i=1}^{10} X_i$ denote the winnings at the end of the game. Then,

$$\begin{aligned} E[G] &= E\left[\sum_{i=1}^{10} X_i\right] \\ &= \sum_{i=1}^{10} E[X_i] && \text{by linearity of expectation} \\ &= \sum_{i=1}^{10} 0 = 0 && \text{as } X_i \text{ are iid} \end{aligned}$$

b. What is the variance of your total earnings? (Show your work; state your assumptions.)

Under the same assumptions as part (a),

$$\begin{aligned}
Var(G) &= Var\left(\sum_{i=1}^{10} X_i\right) \\
&= \sum_{i=1}^{10} \text{Var}(X_i) && \text{since } X_i \text{ are independent} \\
&= \sum_{i=1}^{10} 1 = 10 && \text{since } X_i \text{ are identically distributed}
\end{aligned}$$

c. Suppose that the coin is biased and you have a .4 chance of winning for each flip. Repeat the calculations in parts *a* and *b*.

Since the coin is biased, we now have,

$$X_i = \begin{cases} +1 & \text{with probability } 0.4 \\ -1 & \text{with probability } 0.6 \end{cases}$$

and the mean and variance of X_i for $i = 1, 2, \cdots, 10$ are,

$$\begin{aligned}
E[X_i] &= 1 \times 0.4 + (-1) \times 0.6 = -0.2 \\
Var(X_i) &= E[X^2] - E[X]^2 = \left[1^2 \times 0.4 + (-1)^2 \times 0.6\right] - (-0.2)^2 = 0.96
\end{aligned}$$

Again, letting $G = \sum_{i=1}^{10} X_i$,

$$\begin{aligned}
E[G] &= E\left[\sum_{i=1}^{10} X_i\right] \\
&= \sum_{i=1}^{10} E[X_i] && \text{by linearity of expectation} \\
&= \sum_{i=1}^{10} -0.2 = -2 && \text{as } X_i \text{ are iid} \\
Var(G) &= Var\left(\sum_{i=1}^{10} X_i\right) \\
&= \sum_{i=1}^{10} \text{Var}(X_i) && \text{since } X_i \text{ are independent} \\
&= \sum_{i=1}^{10} 0.96 = 9.6 && \text{since } X_i \text{ are identically distributed}
\end{aligned}$$

Thus, by using a biased coin your are expected to lose \$2 at the end of the ten games, and the variance of your total earnings is \$9.60.

10 die rolls

1000

Problem 19

Note that the code

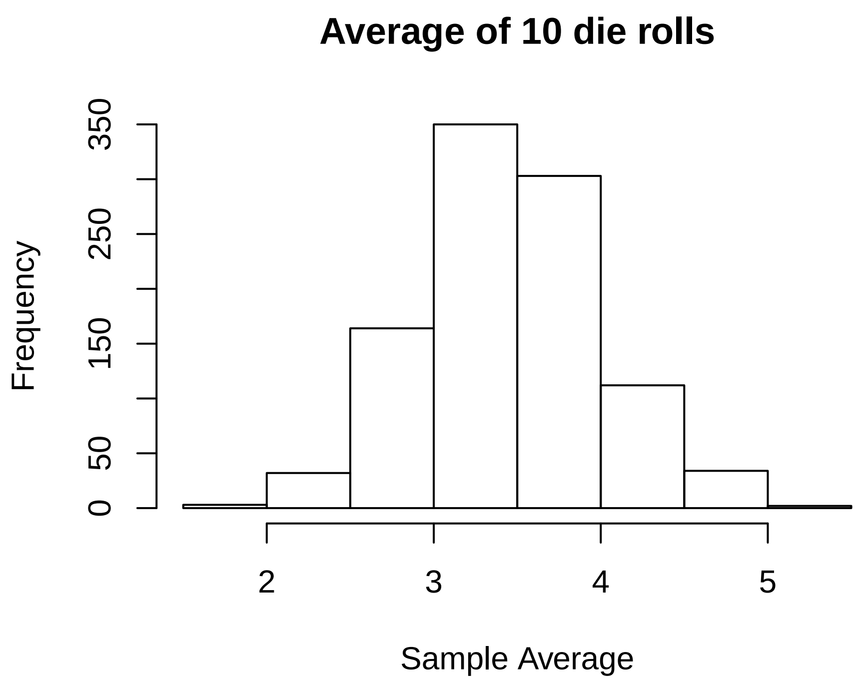
```
temp <- matrix(sample(1 : 6, 1000 * 10, replace = TRUE), 1000)
xBar <- apply(temp, 1, mean)
```

In R produces $1,000$ averages of 10 die rolls. That is, it's like taking ten dice, rolling them, averaging the results and repeating this $1,000$ times.

a. Do this in R. Plot histograms of the averages.

```
set.seed(1000)
temp <- matrix(sample(1 : 6, 1000 * 10, replace = TRUE), 1000)
xBar <- apply(temp, 1, mean)

hist(xBar, xlab="Sample Average", main="Average of 10 die rolls")
```



b. Take the mean of `xBar`. What should this value be close to? (Explain your reasoning.)

```
mean(xBar)
```

```
[1] 3.4965
```

In class, we showed that the expected value of sample averages of iid random variables is approximately equal to the expectation for a single die roll. Thus, we expect the mean to be close to 3.5 as this is the expectation for a single die roll (shown in class).

c. Take the standard deviation of `xBar`. What should this value be close to? (Explain your reasoning.)

```
sd(xBar)
```

```
[1] 0.543685
```

Since the variance of a single die roll is 2.92, as shown in class, we expect the variance of the sample average of 10 dice rolls to be $\frac{2.92}{100}$, thus the standard deviation of these sample means should be approximately $\sqrt{\frac{2.92}{10}} = 0.54$.

Problem 20

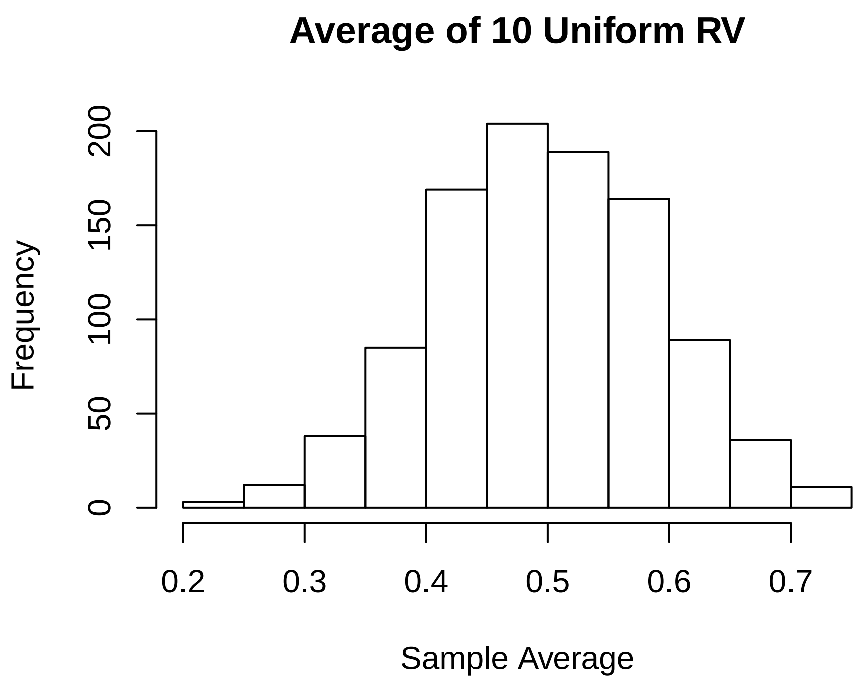
Note that the code

```
xBar <- apply(matrix(runif(1000 * 10), 1000), 1, mean)
```

produces $1,000$ averages of 10 uniforms.

a. Do this in R. Plot histograms of the averages.

```
xBar <- apply(matrix(runif(1000 * 10), 1000), 1, mean)
hist(xBar, xlab="Sample Average", main="Average of 10 Uniform RV")
```



b. Take the mean of `xBar`. What should this value be close to? (Explain your reasoning.)

```
mean(xBar)
```

```
[1] 0.4990916
```

In class, we showed that the expected value of sample averages of iid random variables is approximately equal to the expectation for a uniformly distributed random variable. Thus, we expect the mean to be close to 0.5 as this is the expectation for a uniformly distributed r.v. with parameters $a = 0, b = 1$.

c. Take the standard deviation of `xBar`. What should this value be close to? (Explain your reasoning.)

```
sd(xBar)
```

```
[1] 0.09111284
```

The standard deviation of the sample mean is $\sigma/\sqrt{n}$. Using the fact that the variance of a standard uniform random variable is $\frac{1}{12}$, we expect the standard deviation of the sample mean to be close to $\sqrt{\frac{1/12}{10}} = 0.091$.