Stat 3202: Homework 1

Nathan Johnson.9254

Due Saturday, January 21 by 11:59 pm

Setup:

```
library(tidyverse)
library(readr)
```

Instructions

- Replace "FirstName LastName (name.n)" above with your information.
- Provide your solutions below in the spaces marked "Solution:".
- Include any R code that you use to answer the questions; if a numeric answer is required, show how you calculated it in R.
- Knit this document to **pdf** and upload both the **pfd** file and your completed **Rmd** file to Carmen
- Make sure your solutions are clean and easy-to-read by
  - formatting all plots to be appropriately sized, with appropriate axis labels.
  - only including R code that is necessary to answer the questions below.
  - only including R output that is necessary to answer the questions below (avoiding lengthy output).
  - providing short written answers explaining your work, and writing in complete sentences.

Due on Carmen Saturday, January 21 before 11:59 pm. All uploads must be .pdf. Submissions will be accepted for 24 hours past this deadline, with a deduction of 1% per hour. Absolutely no submissions will be accepted after this grace period.

Concepts & Application

In this assignment, you will

- identify the mean and variance functions of of several probability distributions.
- find expectation and variances of several probability distributions.
- simulate random numbers i R.
- learn many possible estimators for a parameter.
- find the mean and variance linear combination of random variables.

Question 1

Toss a balanced coin. Let's define a random variable X such that $X = 1$ if a heads is observed and $X = 0$ otherwise. Find the following expectations.

(a) $E(X)$.
(b) $E(-2X+1)$.
(c) $V(X)$.
(d) $V(-2X+9)$.
(e) $V(3X+10)$.

Solution to Question 1

Your answers go here.

Part a: $E(X) = 1 * 0.5 + 0 * 0.5 = 0.5$

Part b: $E(-2X+1) = -2E(X) + E(1) = -2 * 0.5 + 1 = 0$

Part c: $V(X) = 0.5(0.5) = 0.25$

Part d: $V(-2X+9) = 4V(X) + V(9) = 4 * 0.25 + 0 = 1$

Part e: $V(3X+10) = 9V(X) + V(10) = 9 * 0.25 + 0 = 2.25$

Question 2

In a day, the number of gallons of blood donation received at a local hospital, denoted by X, has the probability mass function (pmf) $f(x)$ as below.

x	0	1	2	3
$f(x)$	0.1	0.4	0.3	0.2

(a) Find the expectation of X, $E(X)$.
(b) Find the variance and standard deviation of X, $V(X)$.
(c) If the set-up cost (in hundreds of dollars) for blood donation received denoted by a random variable C is given by $C = 10 + 8X + X^2$, what is the expected (average) set-up cost for a day?

Solution to Question 2

Your answers go here.

Part a: $E(X) = 0.1 * 0 + 0.4 * 1 + 0.3 * 2 + 0.2 * 3 = 1.6$

Part b:

$$E(X^2) = 0.1*0^2+0.4*1^2+0.3*2^2+0.2*3^2 = 0.4+1.2+1.8 = 3.4V(X) = E(X^2)-E(X)^2 = 3.4-2.56 = 0.84$$

Part c: $E(10+8X+X^2) = E(10) + 8E(X) + E(X^2) = 10 + 8 * 1.6 + 3.4 = 26.2$

Question 3

A fair six-sided die is rolled. The faces are numbered 1-6. Let X represent the roll of a die.

(a) Write the PMF, $P(X)$. Please list each outcome and its corresponding probability.
(b) Compute the population parameters $\mu = E(X)$ and $\sigma^2 = V(X)$.
(c) Simulate 100 rolls of a die with the function `sample()`. Produce a histogram for the 100 samples. Plots should be presentation quality.

Hint The sample function be used as follows: `sample(c(1:6),100,replace=TRUE)`

(d) Using your sample of 100 rolls, estimate μ with the statistics

(i) $\hat{\mu_1} = \bar{x}$
(ii) $\hat{\mu_2} = \frac{x_{max}+x_{min}}{2}$
(iii) $\hat{\mu_3} = x^{0.5}$ (the median)
(iv) $\hat{\mu_4} = x_{mode}$

Which do you think is the best estimator for μ?

(e) Using your sample of 100 rolls, estimate σ^2 with s^2, the sample variance.

(f) Now simulate 1,000,000 rolls and compute the sample mean (let's assume the sample mean would be best estimator for μ) and sample variance of those rolls. How close are they to the true parameters, μ and σ^2?

Solution to Question 3

Your answers go here.

Part a:

x	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Part b: $\mu = E(X) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3.5$

$E(X^2) = \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} = 15.167$

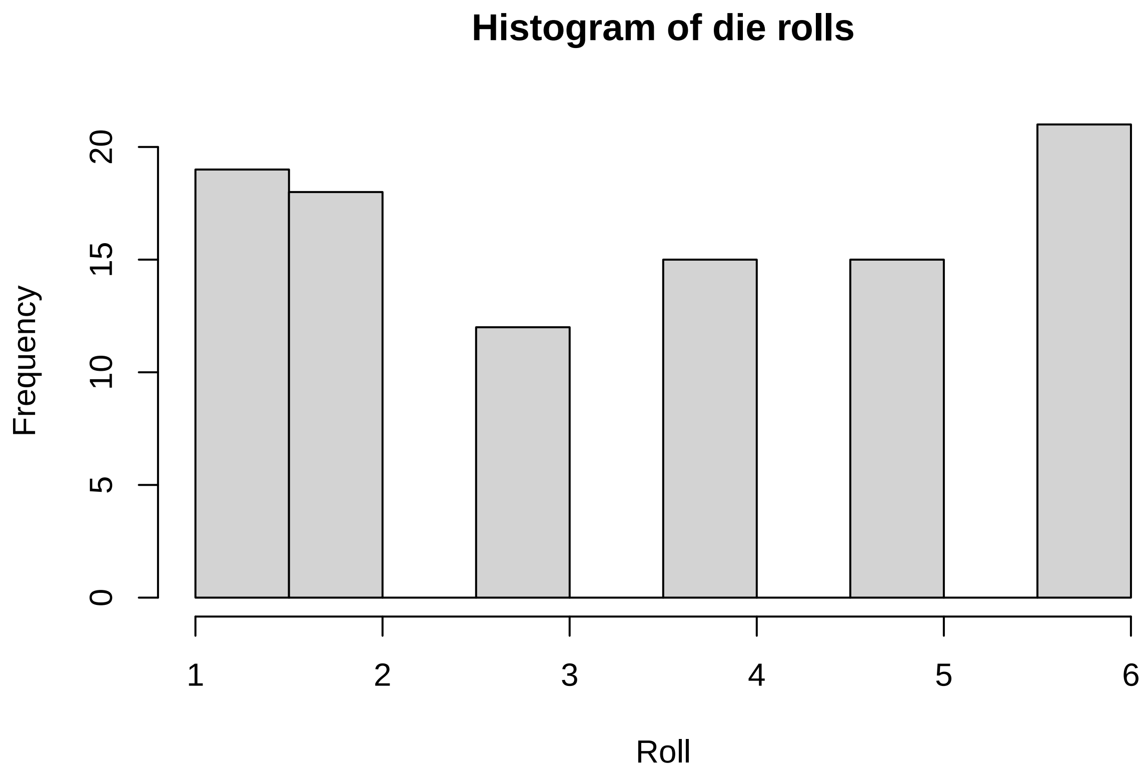
$\sigma^2 = V(X) = E(X^2) - E(X)^2 = 15.167 - 12.25 = 2.9167$

Part c:

```
set.seed(1)
dist = sample(c(1:6),100,replace=TRUE)
dist
```

```
##   [1] 1 4 1 2 5 3 6 2 3 3 1 5 5 2 6 6 2 1 5 5 1 1 6 5 5 2 2 6 1 4 1 4 3 6 2 2 6
##  [38] 4 4 4 2 4 1 6 1 4 1 6 2 3 2 6 6 2 5 2 6 6 6 1 3 3 6 4 6 3 1 4 5 1 1 6 4 5
##  [75] 5 4 6 5 4 4 1 5 5 6 1 1 3 6 2 2 3 6 2 4 3 5 2 2 1 3
```

```
hist(dist, main = "Histogram of die rolls", xlab = "Roll")
```



Part d: $(ii.)\hat{\mu_2} = \frac{x_{max}+xmin}{2}$ is the best estimator for μ. This is because a fair six-sided die is represented by a uniform distribution and ii. is the expectation of X of a uniform distribution.

Part e:

```
s2 = var(dist)
s2
```

```
## [1] 3.34303
```

$\hat{\sigma} = 3.34303$

Part f:

```
set.seed(1)
dist2 = sample(c(1:6),1000000,replace=TRUE)
m2 = mean(dist2)
m2
```

```
## [1] 3.497758
```

```
v2 = var(dist2)
v2
```

```
## [1] 2.916044
```

$\hat{\mu} = 3.497758$

$\hat{\sigma} = 2.916044$

The computed $\hat{\mu}$ is -0.002242 off of μ.

The computed $\hat{\sigma}$ is -0.00062266 off of σ.

The result of the sample being so close to the real mean and variance is expected since they have huge sample sizes.

Question 4

Let's consider a continuous uniform distribution as defined below.

Definition: Let X be a random variable with a, b, real values such that $a < b$. A random variable X can be gives as a continuous uniform distribution on the interval (a, b), denoted by $X \sim U(a, b)$, if and only if its probability density function is $f(x) = \frac{1}{(b-a)}$,for any $x \in (a, b)$.

Suppose a random variable X for a particular situation is distributed as $X \sim U(\lambda, \lambda + 10)$, where λ is parameter to be estimated later.

(i) Using calculus (as opposed the known properties of the uniform distribution), compute the first moment $E(X)$, the second moment $E(X^2)$ and the variance $V(X)$.
(ii) Using known properties of the uniform distribution (that is, by using the formulas for expectation and variance of a uniform distribution), check that your answers to part (i)) for the expectation and variance are correct.

Solution to Question 4

$E(X) = \int_{\lambda}^{\lambda+10} \frac{x}{\lambda+10-\lambda} dx = \lambda + 5$

$E(X^2) = \int_{\lambda}^{\lambda+10} \frac{x^2}{\lambda+10-\lambda} dx = \lambda^2 + 10\lambda + \frac{100}{3}$

$V(X) = E(X^2) - E(X)^2 = \lambda^2 + 10\lambda + \frac{100}{3} - (\lambda^2 + 10\lambda + 25) = 8.33$

Question 5

Child abduction occurring in a particular city can be attributed by a Poisson process. Suppose the average number of child abduction per week is 2. Let X represent number of children abducted in a week, and let us assume that $X \sim Poisson(\lambda = 2)$. A sample over 10 weeks collected the following abductions: $\{1,1,2,3,5,1,3,3,2,0\}$. Let us assume that abducted children are independent.

a) Using known properties of the Poisson distribution, what are the population mean and population variance of this distribution? No need for calculus, just use what you know about a Poisson random variable.

b) Compute the sample mean $\bar{x}$ and the sample variance s^2. You may use the functions `mean()` and `var()` in `R` to claculate thoses values.

c) Do $\bar{x}$ and s^2 seem to be good close estimates of the true parameter λ? Which is better, and what criterion did you use to define "better"?

Solution to Question 5

a. $\mu = 2$

$\sigma^2 = 2$

b.

```
sample = c(1,1,2,3,5,1,3,3,2,0)
m = mean(sample)
m
```

```
## [1] 2.1
```

```
var = var(c(1,1,2,3,5,1,3,3,2,0))
var
```

```
## [1] 2.1
```

$\hat{\mu} = 2.1$

$\hat{\sigma} = 2.1$ c. X and s^2 are 'good' close estimates, but μ and σ^2 are better since they are equal to the true parameter.

Question 6

Consider a population with density $f(y \mid \beta) = \beta y^{\beta-1}$ for $0 < y < 1$ and $\beta > -1$. Compute the expectation and variance of this population.

Solution to Question 6

$E(X) = \int_0^1 y\beta y^{\beta-1} dy = \frac{\beta}{\beta+1}$

$E(X^2) = \int_0^1 y^2 \beta y^{\beta-1} dy = \frac{\beta}{\beta+2}$

$V(X) = E(X^2) - E(X)^2 = \frac{\beta}{\beta+2} - \frac{\beta^2}{\beta^2+2\beta+1}$