Stat 3202 Lab 2

Jane Weissberg.11

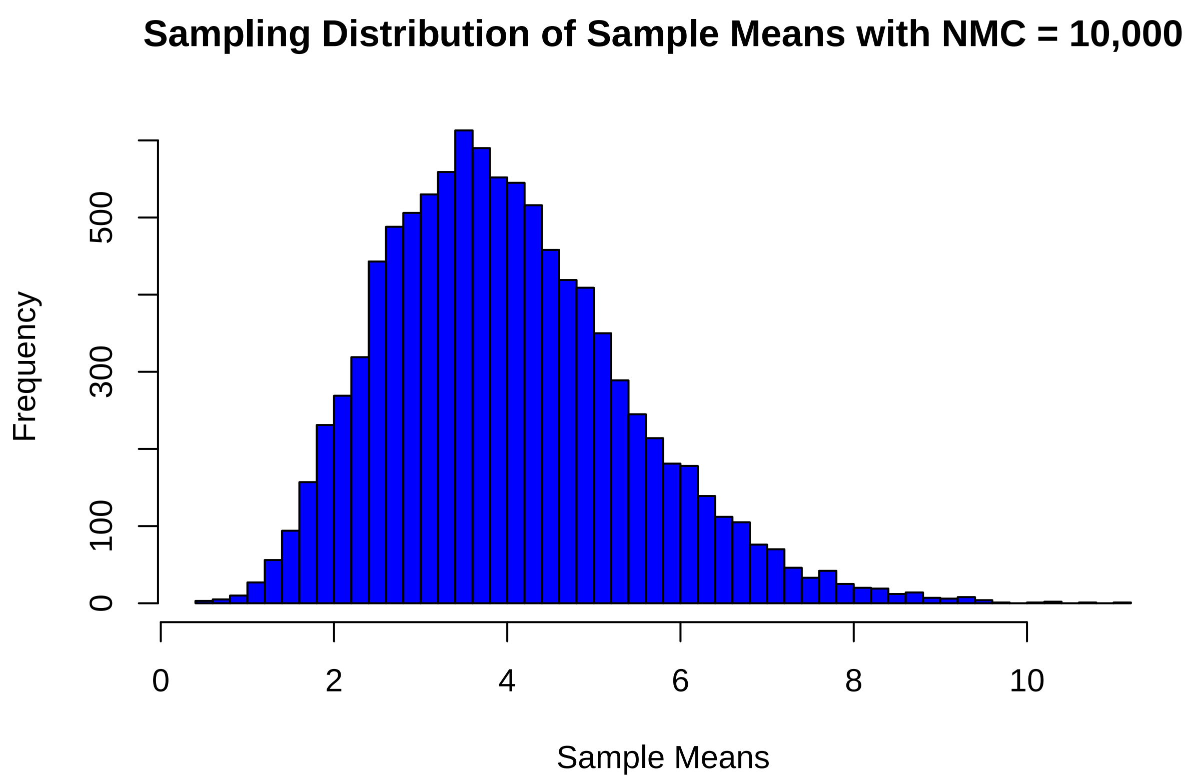
2024-05-23

1A

```
alpha <- 2
beta <- 0.5
n <- 4
NMC <- 10000

sample_means <- numeric(NMC)
for (i in 1:NMC) {
  sample <- rgamma(n, alpha, beta)
  sample_means[i] <- mean(sample)
}

hist(sample_means, breaks = 50, main = "Sampling Distribution of Sample Means with NMC = 10,000", xlab =
```



The shape of the histogram is slightly skewed right. The central limit theorem is not coming into play here because although the number of samples is large, the sample size of each individual sample is still small.

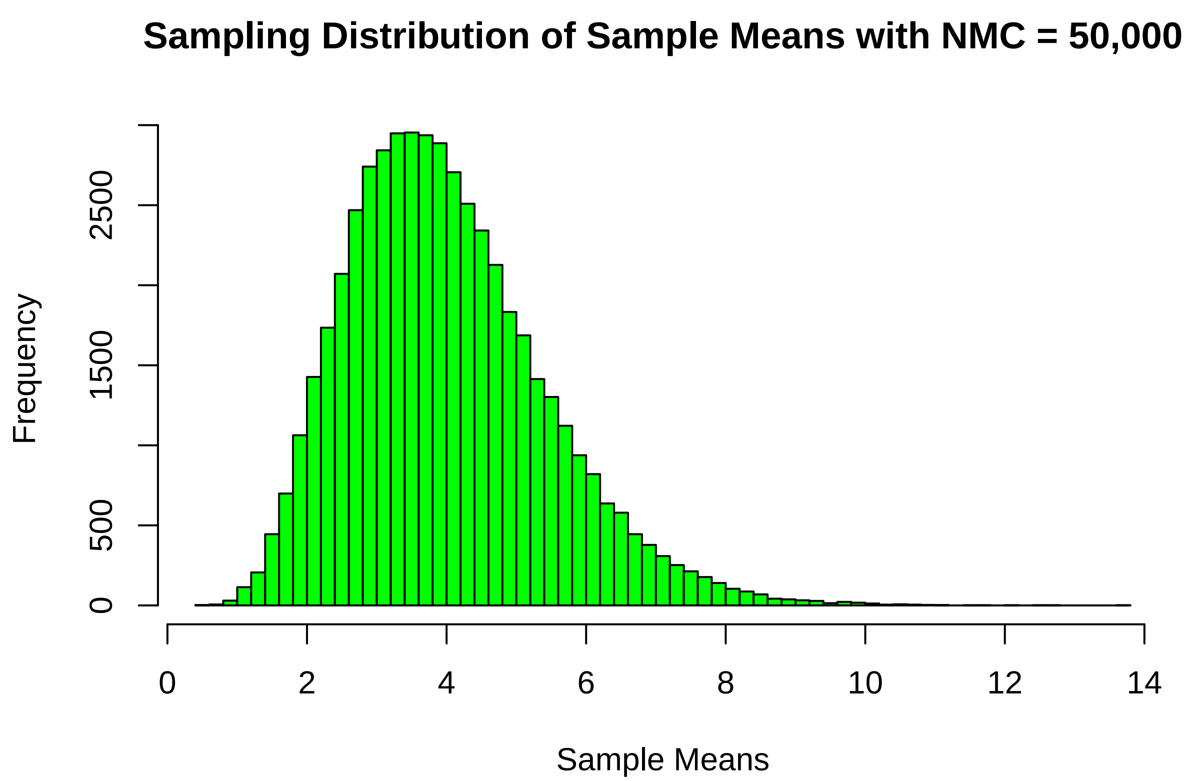
n is the sample size of each individual sample from the Gamma distribution, and Nmc is the number of samples.

1B

```
NMC <- 50000

sample_means <- numeric(NMC)
for (i in 1:NMC) {
  sample <- rgamma(n, alpha, beta)
  sample_means[i] <- mean(sample)
}

hist(sample_means, breaks = 50, main = "Sampling Distribution of Sample Means with NMC = 50,000",
     xlab = "Sample Means", col = "green")
```



Increasing NMC makes the histogram smoother, but it is still skewed right. The central limit theorem still does not apply.

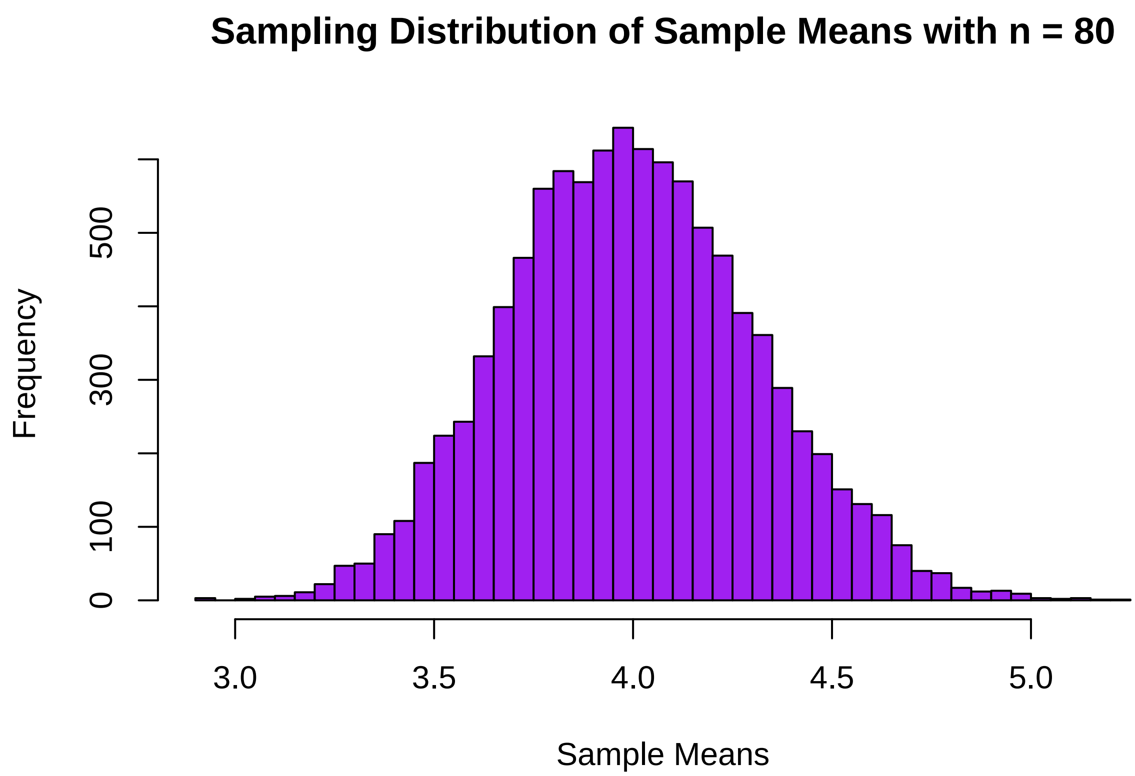
1C

```
n <- 80
NMC <- 10000

sample_means <- numeric(NMC)
```

```r
for (i in 1:NMC) {
  sample <- rgamma(n, alpha, beta)
  sample_means[i] <- mean(sample)
}

hist(sample_means, breaks = 50, main = "Sampling Distribution of Sample Means with n = 80",
     xlab = "Sample Means", col = "purple")
```



The histogram now appears to be approaching approximately normal because n has increased. The central limit theorem now applies.

1D

```r
empirical_mean <- mean(sample_means)
empirical_variance <- var(sample_means)

theoretical_mean <- alpha / beta
theoretical_variance <- (alpha / (beta^2)) / n

empirical_mean
```

```
## [1] 3.996453
```

```r
empirical_variance
```

```
## [1] 0.09871951
```

```r
theoretical_mean
```

```
## [1] 4
```

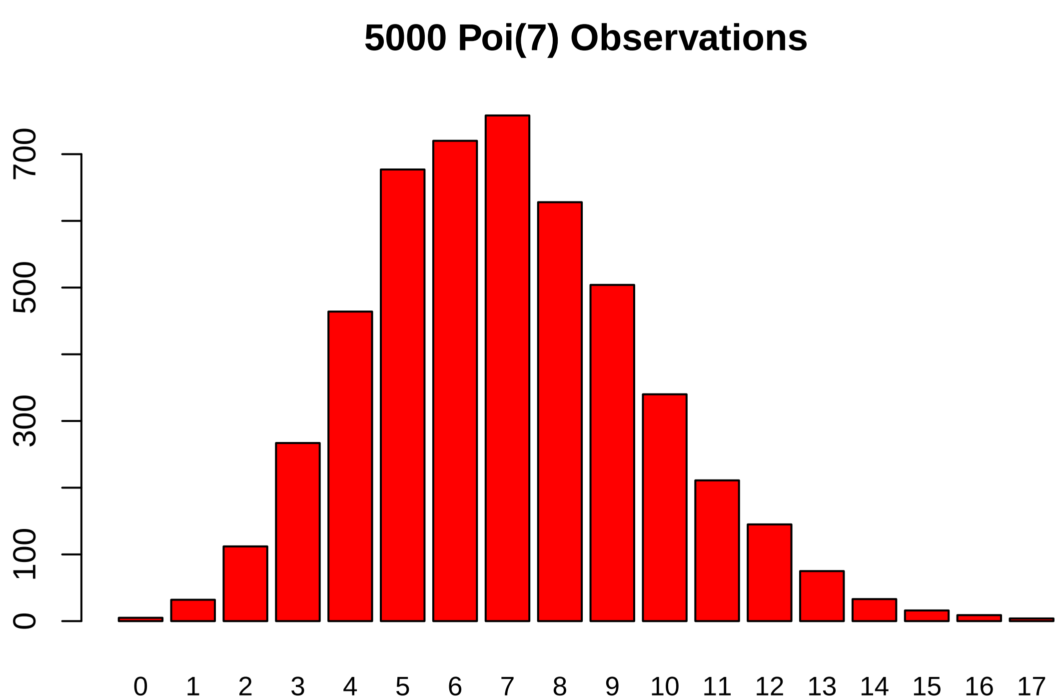
```r
theoretical_variance
```

```
## [1] 0.1
```

The empirical mean and empirical variance are both approaching the expectation of the original population and the variance of the original population divided by the sample size, respectively. According to the Central Limit Theorem, the sample means will follow a normal distribution with mean μ and variance σ^2/n.

2A

```r
x <- rpois(5000, lambda=7)
barplot(table(x), main="5000 Poi(7) Observations", col="red", cex.names=.8)
```



2B

```r
mean(x)
```

```
## [1] 6.9668
```

```r
var(x)
```

```
## [1] 7.049908
```

The sample mean is 0.016 away from the population mean of 7, and the sample variance is 0.118 away from the population variance. The sample mean is a better approximation of the truth.

The estimator with the lowest variance, bias, and mean squared error is typically the best estimator.

3A

$E(A) = 6$ $E(B) = 5$ $E(C) = 0.1$

$Var(A) = 4$ $Var(B) = 10$ $Var(C) = 0.01$

$E(X) = 15.7$ $Var(X) = 44.09$

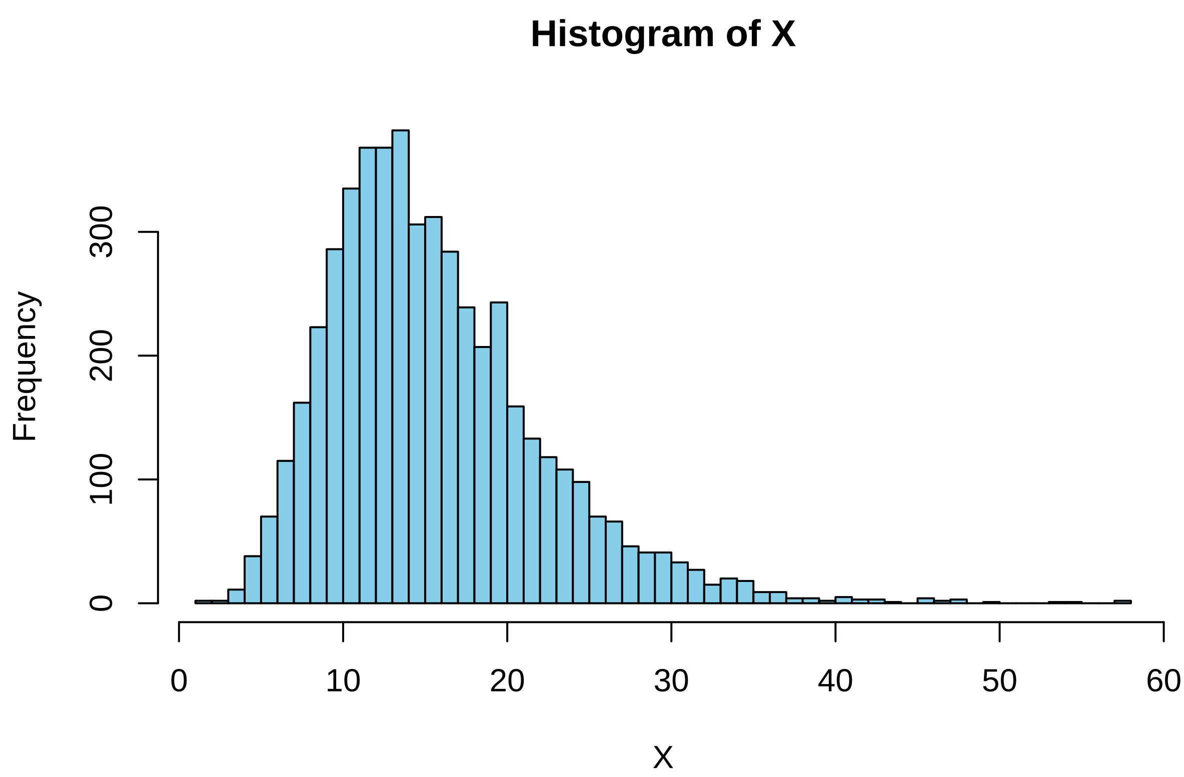
3B

```r
set.seed(321)
N <- 5000

A <- rnorm(N, 6, 2)
B <- rchisq(N, 5)
C <- rexp(N, 10)

X <- A + 2*B - 3*C

hist(X, breaks = 50, main = "Histogram of X", xlab = "X", col = "skyblue")
```



```r
sample_mean_X <- mean(X)
sample_variance_X <- var(X)

sample_mean_X
```

```
## [1] 15.63153
```

```r
sample_variance_X
```

```
## [1] 43.06341
```

```r
abs(sample_mean_X - 15.7)
```

```
## [1] 0.06846893
```

```r
abs(sample_variance_X - 44.09)
```

```
## [1] 1.026586
```

The sample mean is approximately 0.069 from the true mean and the sample variance is approximately 1.027 from the true variance.

3C

```r
pnorm(3.5, mean = 6, sd = 2)
```

```
## [1] 0.1056498
```

3D

```r
mean(A < 3.5) * 100
```

```
## [1] 10.78
```

If you consider both answers as percentages, they are 0.22% away from one another. They should be very close but not necessarily identical because of the random variation in the sample.

3E

```r
mean(X > 10) * 100
```

```
## [1] 81.82
```