Analysis of Exponential distribution

Jose Gonzalez

6 novembre 2016

Introduction

In this report we will compare the means and variances of samples taken from a exponential random variable against the theoretical mean and variance of such a distribution. We will also compare the distribution of the means of the samples to a normal distribution.

Parameters

Sampled mean vs theoretical mean

Theoretical mean

Theoretical mean of an exponential random variable is 1/lambda:

```
1/lambda
```

```
## [1] 5
```

The observed mean of 1,000 drawn from an exponential distribution with lambda 0.2:

```
mean(rexp(simulations,lambda))
```

```
## [1] 5.389714
```

Sampled mean

Following code will create a distribution of 1000 based on the mean of 40 values drawn from an exponential distribution:

```
mns = NULL
for (i in 1 : simulations) mns = c(mns, mean(rexp(n=samples,lambda)))
```

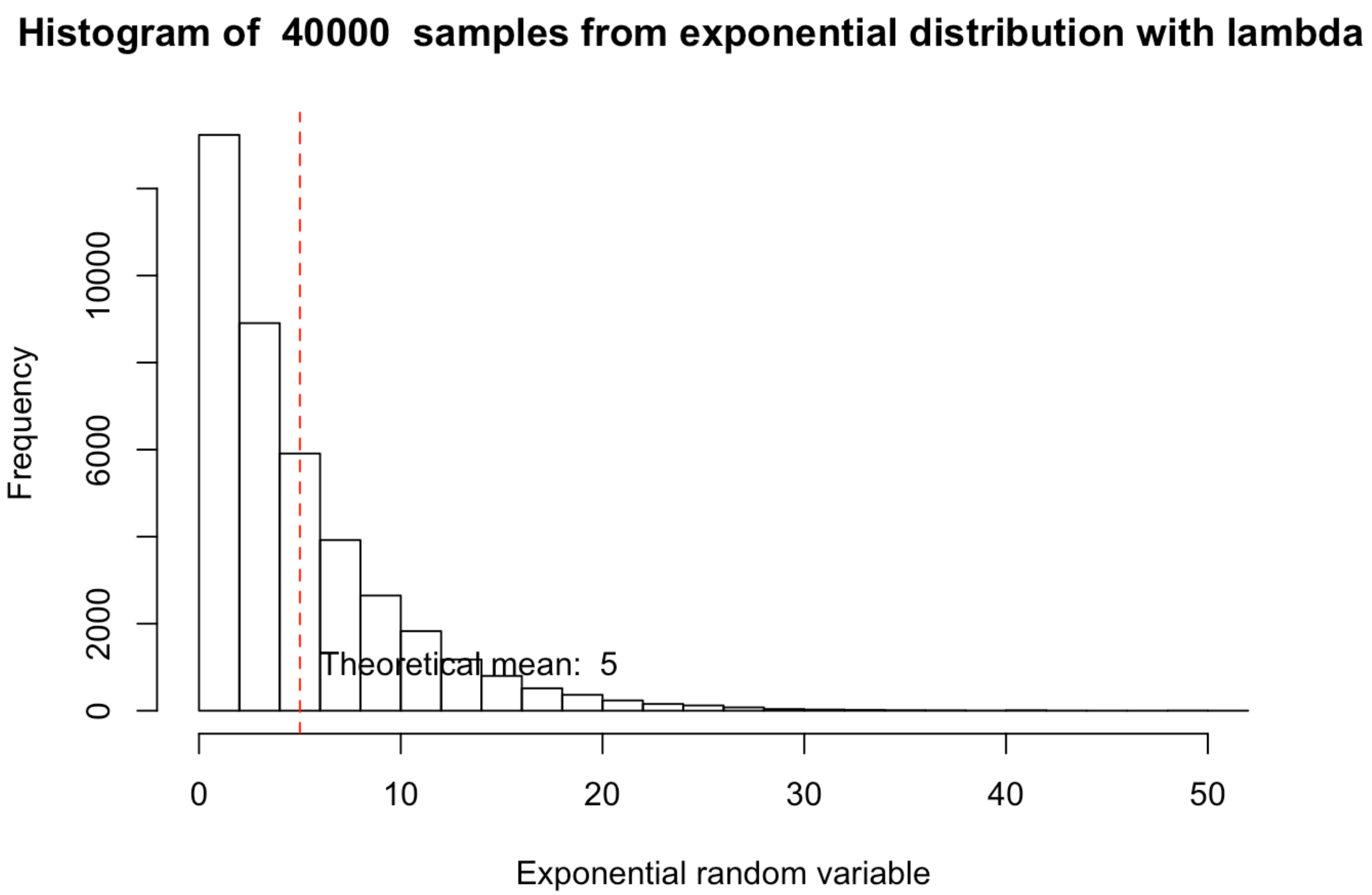
The mean of the 1000 of 40 samples is:

```
mean(mns)
```

```
## [1] 4.9809
```

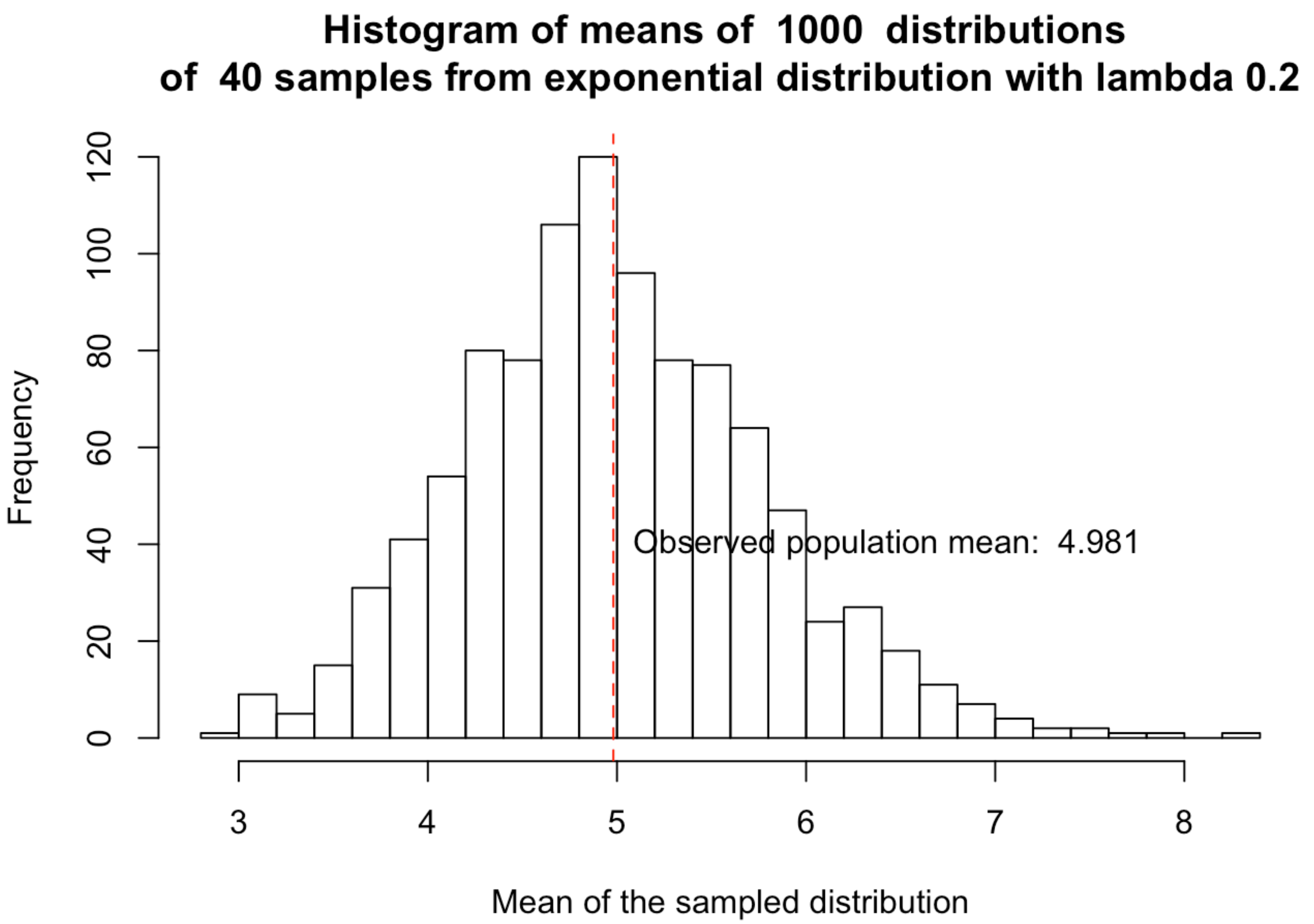
Following figure shows the distribution of 40,000 samples drawn from an exponential distribution with lambda 0.2

```
rexp_sim<-rexp(simulations*samples,lambda)

hist(rexp_sim,
     breaks=n_breaks,
     main=paste("Histogram of ",simulations*samples," samples from exponential dis
tribution with lambda",lambda),
     xlab="Exponential random variable")

abline(v=1/lambda,lty="dashed",col="red")
text(x=1/lambda,
     y=1000,
     labels=paste("Theoretical mean: ",1/lambda),
     pos=4)
```



The following figure shows the distribution of the mean of 1000 distributions of 40 drawn from an exponential distribution with lambda 0.2

```
hist(mns,
      breaks=n_breaks,
     main=paste("Histogram of means of ",simulations," distributions \nof ",sample
s, "samples from exponential distribution with lambda",lambda),
     xlab="Mean of the sampled distribution")

abline(v=mean(mns),lty="dashed",col="red")
text(x=mean(mns),
     y=40,
     labels=paste("Observed population mean: ",prettyNum(mean(mns),digits=4)),
     pos=4)
```



Mean comparion conclusion

The mean of the mean of the sampled distribution 4.981 is close to the theoretical mean 1/lambda: 5

Theoretical vs sampled variance

Theoretical variance:

```
1/lambda^2
```

```
## [1] 25
```

Variance of 1,000 samples:

```
var(rexp(simulations,lambda))
```

```
## [1] 23.63786
```

Variance of sampled population

Draw a 1000 distributions of 40 drawn from an exponential distribution with lambda 0.2 and take the variance of each of the distributions. Calculate the mean of the variances.

```
var = NULL
for (i in 1 : simulations) var = c(var, var(rexp(n=samples,lambda)))
mean(var)
```

```
## [1] 25.39892
```

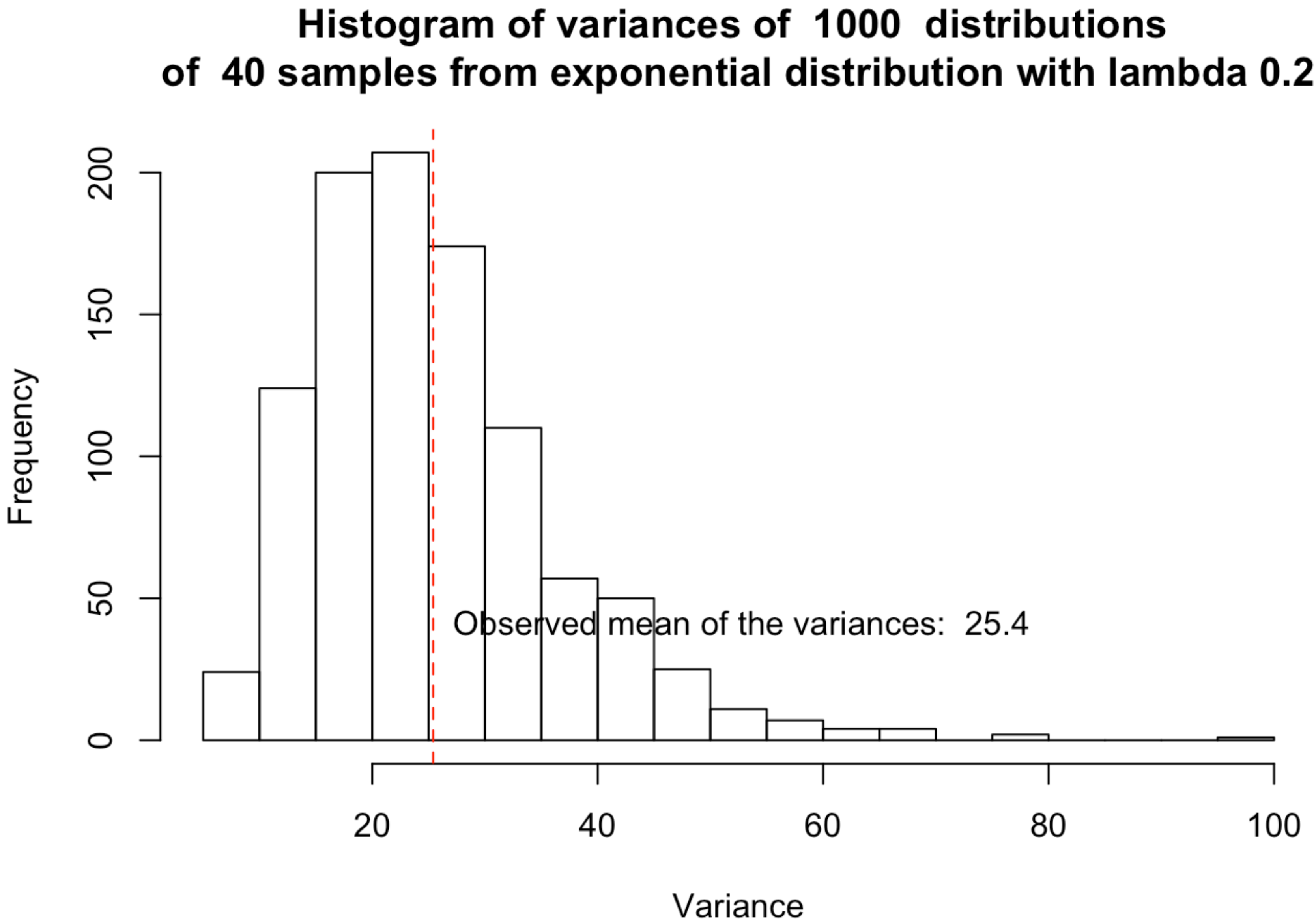
Estimation of population variance based on the mean of the variances of the samples:

```
pvar=(mean(var))*samples/(samples-1)
pvar
```

```
## [1] 26.05017
```

Following figure shows the distribution of the variances of the sampled distribution:

```
hist(var,
     breaks=n_breaks,
       main=paste("Histogram of variances of ",simulations," distributions \nof ",
samples, "samples from exponential distribution with lambda",lambda),
     xlab="Variance")

abline(v=mean(var),lty="dashed",col="red")
text(x=mean(var),
     y=40,
     labels=paste("Observed mean of the variances: ",prettyNum(mean(var),digits=4)
),
     pos=4)
```



Variance comparion conclusion

The estimated variance based on the mean of the variances of the sampled distribution 26.05 is close to the theoretical mean 1/lambda^2: prettyNum(1/lambda^2,digits=4)

Comparison to normal distribution

For this comparison, we will calculate the probability of having the 2.5% to 95% percentiles of the means of the sampled distribution inside the 2.5% to 97.5% percentiles of a draw from a normal distribution.

First let's get the quantiles of our distribution of the means

```
quantiles<-c(0.025,.5,.975)
quantile(mns,quantiles)
```

```
##      2.5%       50%     97.5%
## 3.559935 4.933784 6.677292
```

Then get the quantiles of normal a normal distribution with mean equal to observed mean and variance equal to observed variance

```
qnorm(quantiles,mean=mean(mns),sd=sqrt(pvar))
```

```
## [1] -5.022633  4.980900 14.984433
```

Probability of finding the sampled 2.5%-97.5% interval inside normal distribution:

```
prob<-pnorm(q=quantile(mns,quantiles[length(quantiles)]),mean=mean(mns),sd=sqrt(pv
ar/samples))-pnorm(q=quantile(mns,quantiles[1]),mean=mean(mns),sd=sqrt(pvar/sample
s))
prob
```

```
##     97.5%
## 0.9430904
```

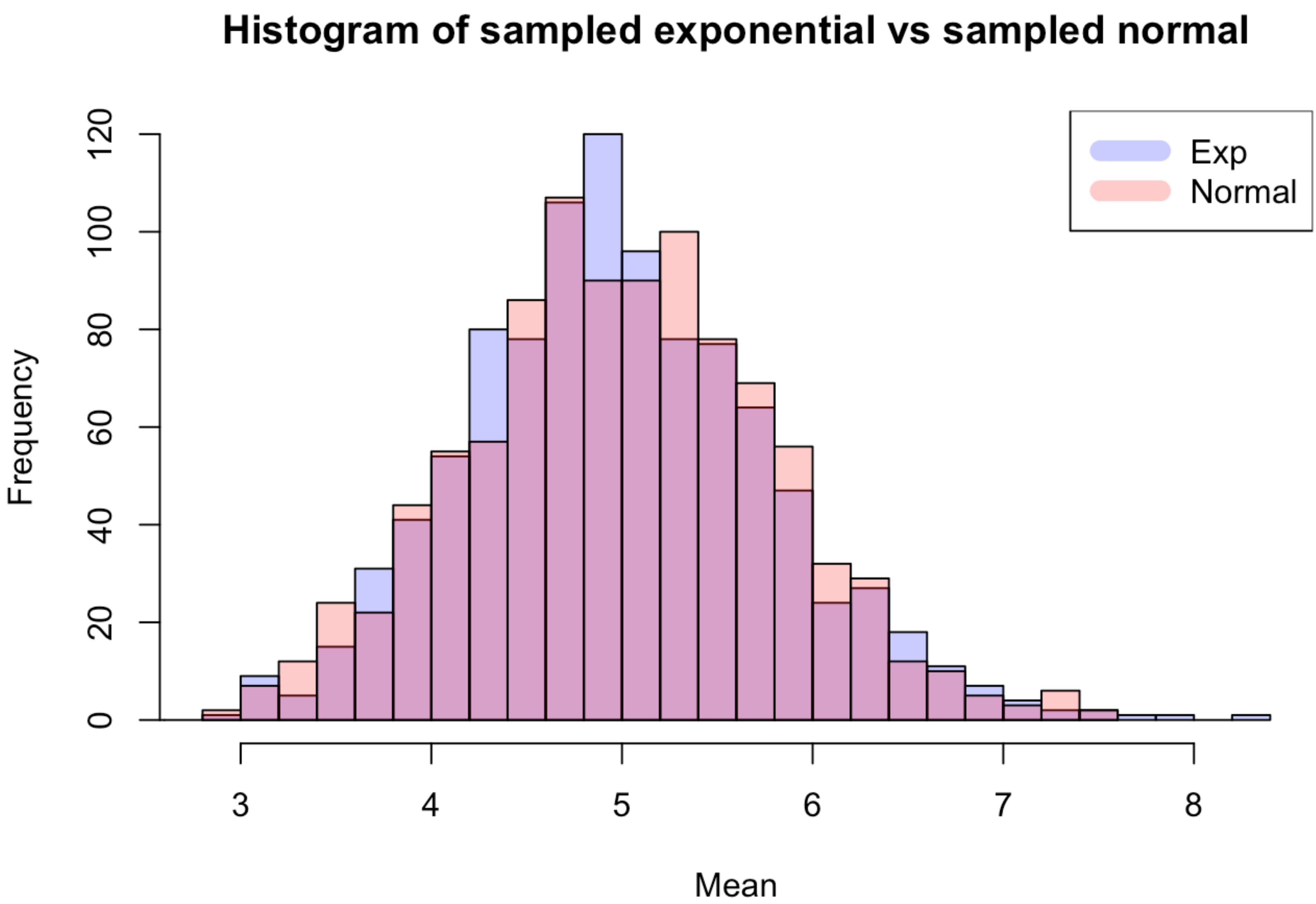
The probability of having drawn the same means from a normal random variable is 94%

Following histogram compares the means drawn from the exponential random variable vs the means drawn from a normal distribution

```
mns_norm = NULL
for (i in 1 : simulations) mns_norm = c(mns_norm, mean(rnorm(n=samples,mean=mean(m
ns),sd=sqrt(pvar))))
mean(var)
```

```
## [1] 25.39892
```

```
hist(mns, col=rgb(0,0,1,1/4),breaks=n_breaks,
     main="Histogram of sampled exponential vs sampled normal",
     xlab="Mean")  # first histogram
hist(mns_norm, col=rgb(1,0,0,1/4), add=T,breaks=n_breaks)  # second
legend("topright", c("Exp", "Normal"), col=c(rgb(0,0,1,1/4), rgb(1,0,0,1/4)), lwd=
10)
```



Comparion to normal distribution conclusion

Both the probability of having the 2.5% to 97.5% percentiles inside the equivalent percentiles of a normal distribuion and the histogram show that the distribution of the means of the samples is normal like.

Conclusion

This illustrated the Central Limit Theorem. Eventhough the original random variable was exponential, drawing the means from a significant number of samples “normalizes” the distribution. This is of course only applicable as the number of samples in each distribution and the number of distribuions taken is sufficiently high.