MSMS 206 : Practical 03

Ananda Biswas

March 11, 2025

➜ Objective

$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} Exp(\text{mean} = \theta)$.

(i) Using CLT we have to show that $\bar{X}$ is a CAN estimator for θ.

(ii) We also have to obtain a CAN estimator for $P[X > t] = e^{-t/\theta}$ and its asymptotic variance.

➜ Theory, R Program, Plot and Interpretation

☞ First we shall show that $\bar{X}$ is a consistent estimator for θ *i.e.* $P[|\bar{X}_n - \theta| < \epsilon]$ tends to 1 as sample size n increases.

We generate 10000 samples of size n and calculate the relative frequency of $[|\bar{X}_n - \theta| < \epsilon]$, this is the probability obtained by empirical approach. As sample size n increases, the empirical probability converges to 1.

We consider $\theta = 2$.

```r
theta <- 2
```

```r
prob <- function(size, epsilon){

  sample_list <- list()

  for(i in 1:10000){
    sample_list[[i]] <- rexp(size, rate = 1 / theta)
  }

  x_bar <- sapply(sample_list, mean)

  m <- length(which(abs(x_bar - theta) < epsilon))

  return(m/10000)
}
```

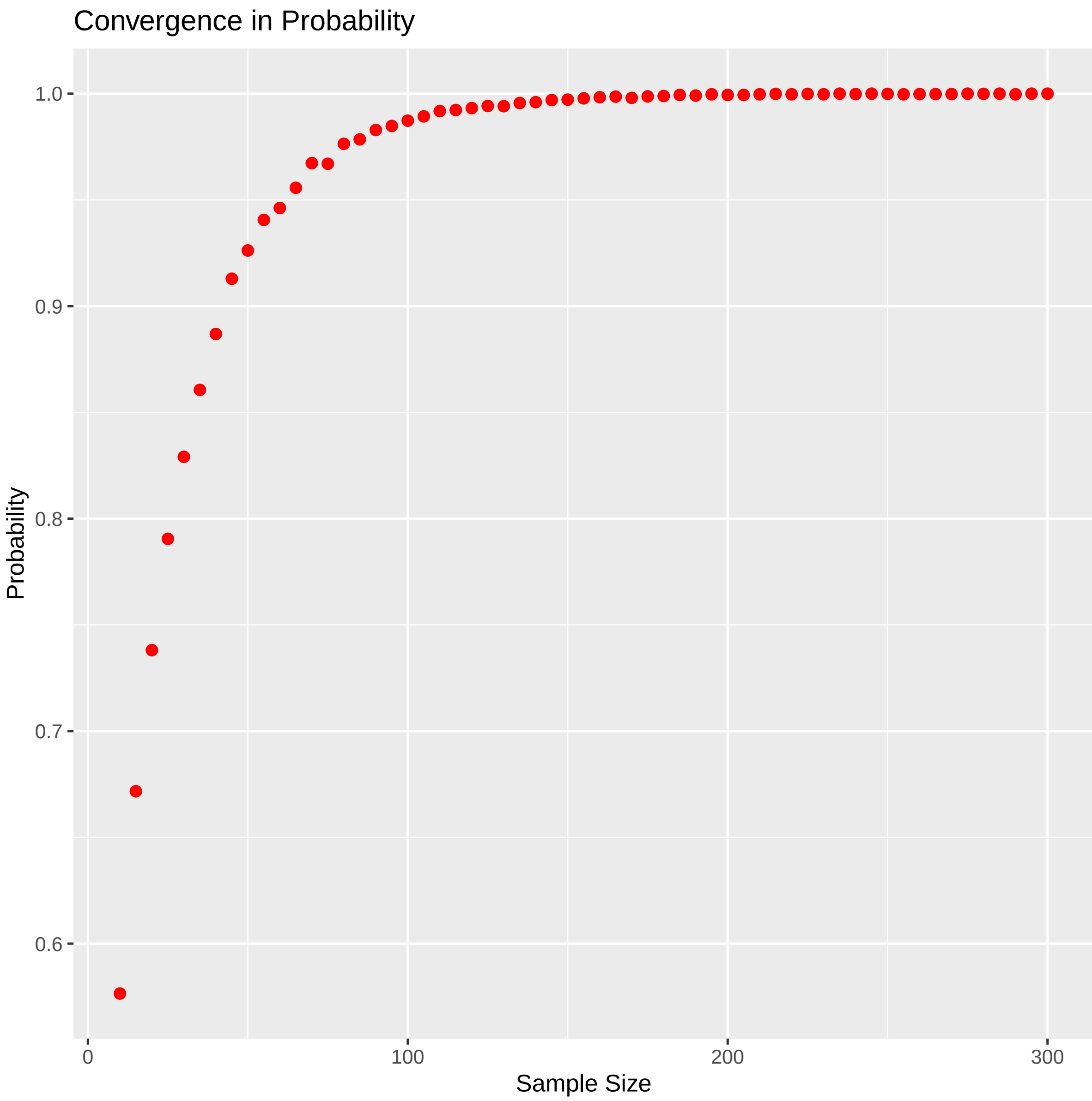
Here we take $\epsilon = 0.5$.

```r
probs <- c()

for (i in seq(10, 300, 5)) {
  probs <- append(probs, prob(size = i, epsilon = 0.5))
}
```

```r
df1 <- data.frame(sample_size = seq(10, 300, 5), probability = probs)

df1 %>%
  ggplot(aes(x = sample_size, y = probability)) +
  geom_point(size = 2, col = "red") +
  labs(x = "Sample Size", y = "Probability",
       title = "Convergence in Probability")
```



✎ As sample size increases, the probability converges to 1. This implies that the sample mean is a consistent estimator of θ.

☞ Now we shall show that $\bar{X}_n$ has an asymptotic normal distribution.

We generate 10000 samples of size n, from there we get 10000 $\bar{X}_n$. We plot their histogram along with density curve. As sample size n increases, the density curve resembles that of a normal distribution.

First we take $n = 2$.

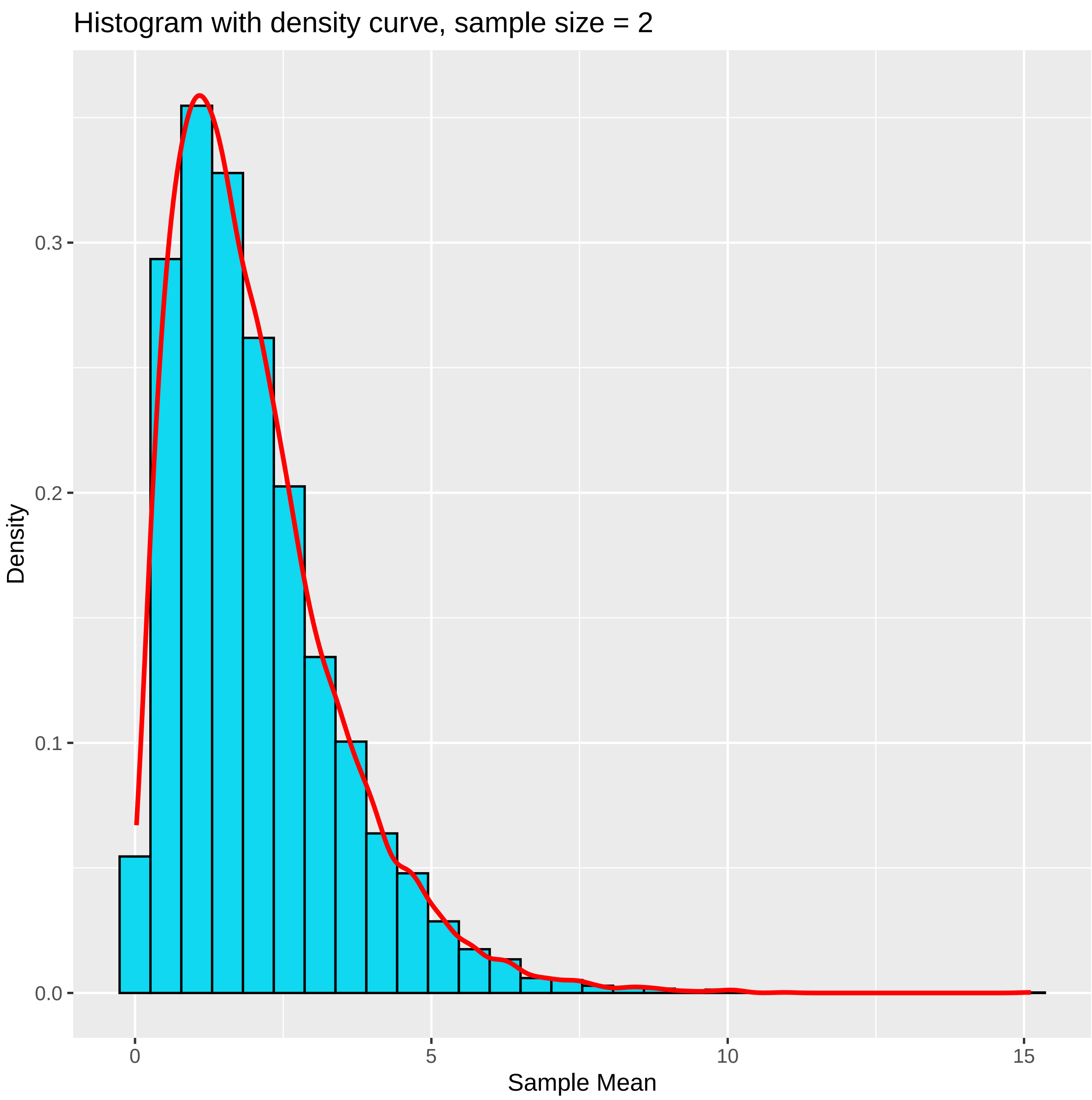
```
sample_list <- list(); size <- 2

for(i in 1:10000){
  sample_list[[i]] <- rexp(size, rate = 1 / theta)
}

x_bar <- sapply(sample_list, mean)
```

```
df2 <- data.frame(means = x_bar)

df2 %>%
  ggplot(aes(x = means)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "#0FD8F0", col = "black") +
  geom_density(col = "red", linewidth = 1) +
  labs(x = "Sample Mean", y = "Density",
       title = "Histogram with density curve, sample size = 2")
```



Next we take $n = 5$.

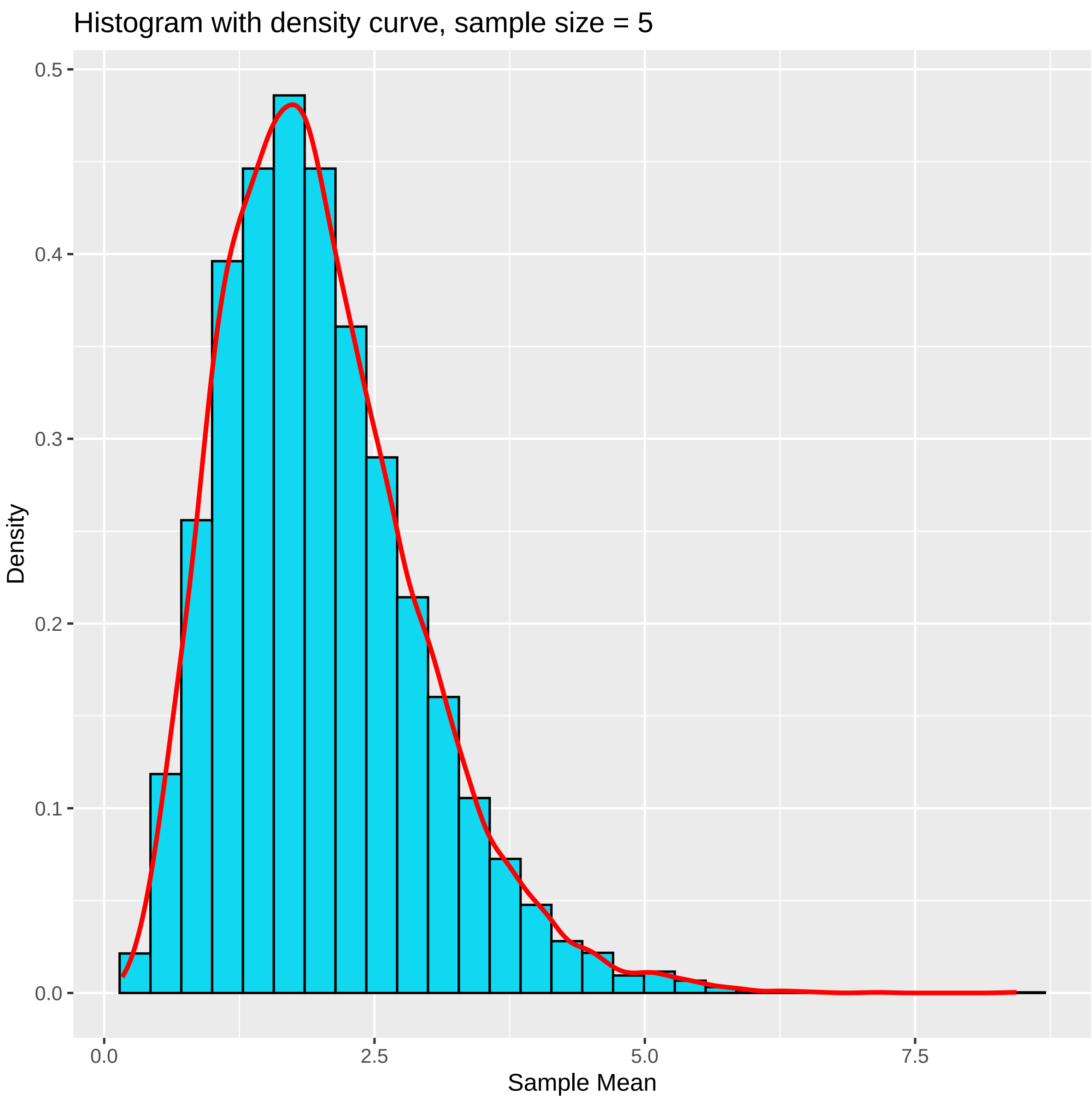
```
sample_list <- list(); size <- 5

for(i in 1:10000){
  sample_list[[i]] <- rexp(size, rate = 1 / theta)
}

x_bar <- sapply(sample_list, mean)
```

```
df2 <- data.frame(means = x_bar)

df2 %>%
  ggplot(aes(x = means)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "#0FD8F0", col = "black") +
  geom_density(col = "red", linewidth = 1) +
  labs(x = "Sample Mean", y = "Density",
       title = "Histogram with density curve, sample size = 5")
```



Next we take $n = 10$.

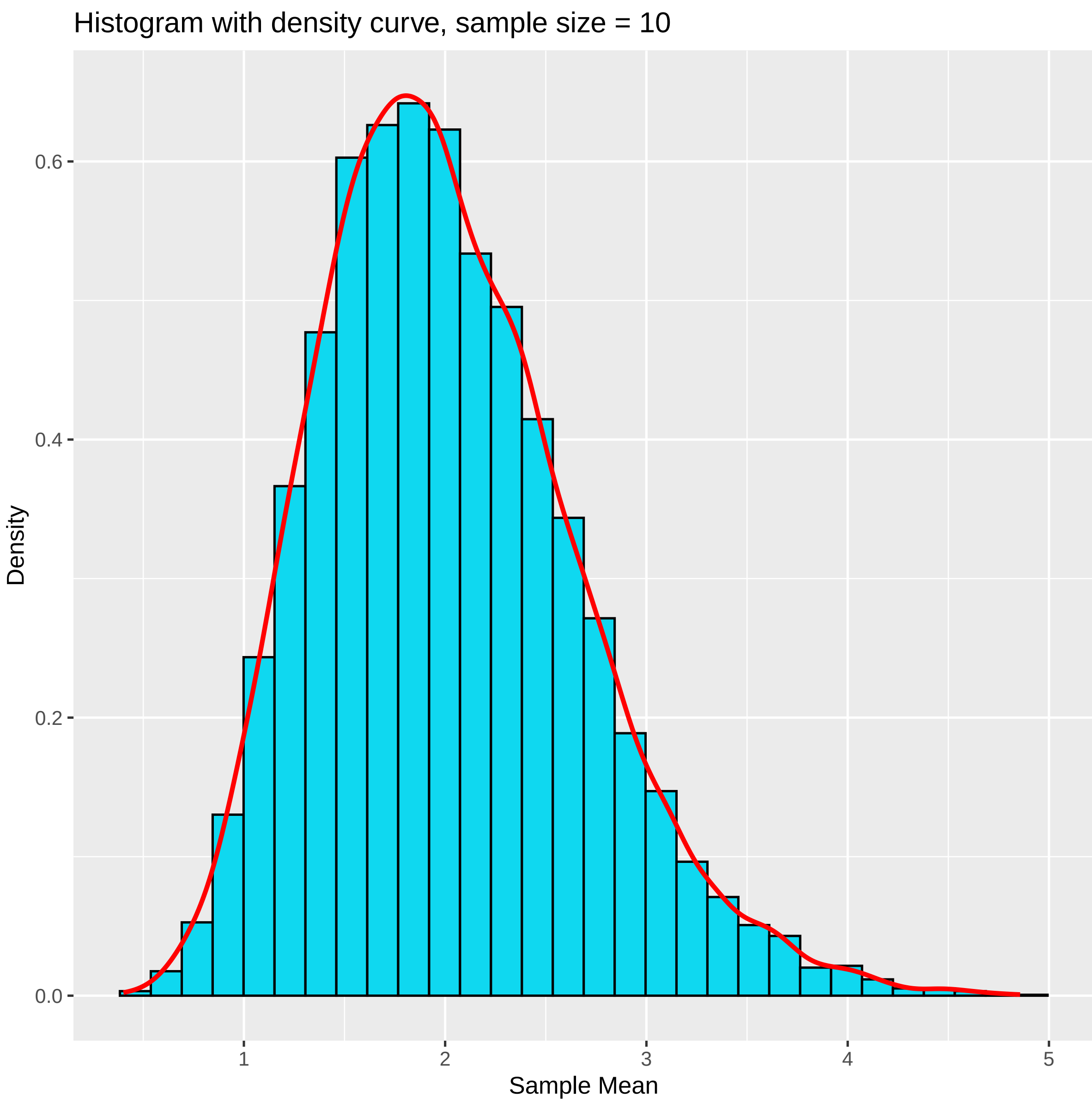
```
sample_list <- list(); size <- 10

for(i in 1:10000){
  sample_list[[i]] <- rexp(size, rate = 1 / theta)
}

x_bar <- sapply(sample_list, mean)
```

```
df2 <- data.frame(means = x_bar)

df2 %>%
  ggplot(aes(x = means)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "#0FD8F0", col = "black") +
  geom_density(col = "red", linewidth = 1) +
  labs(x = "Sample Mean", y = "Density",
       title = "Histogram with density curve, sample size = 10")
```



Finally we take $n = 20$.

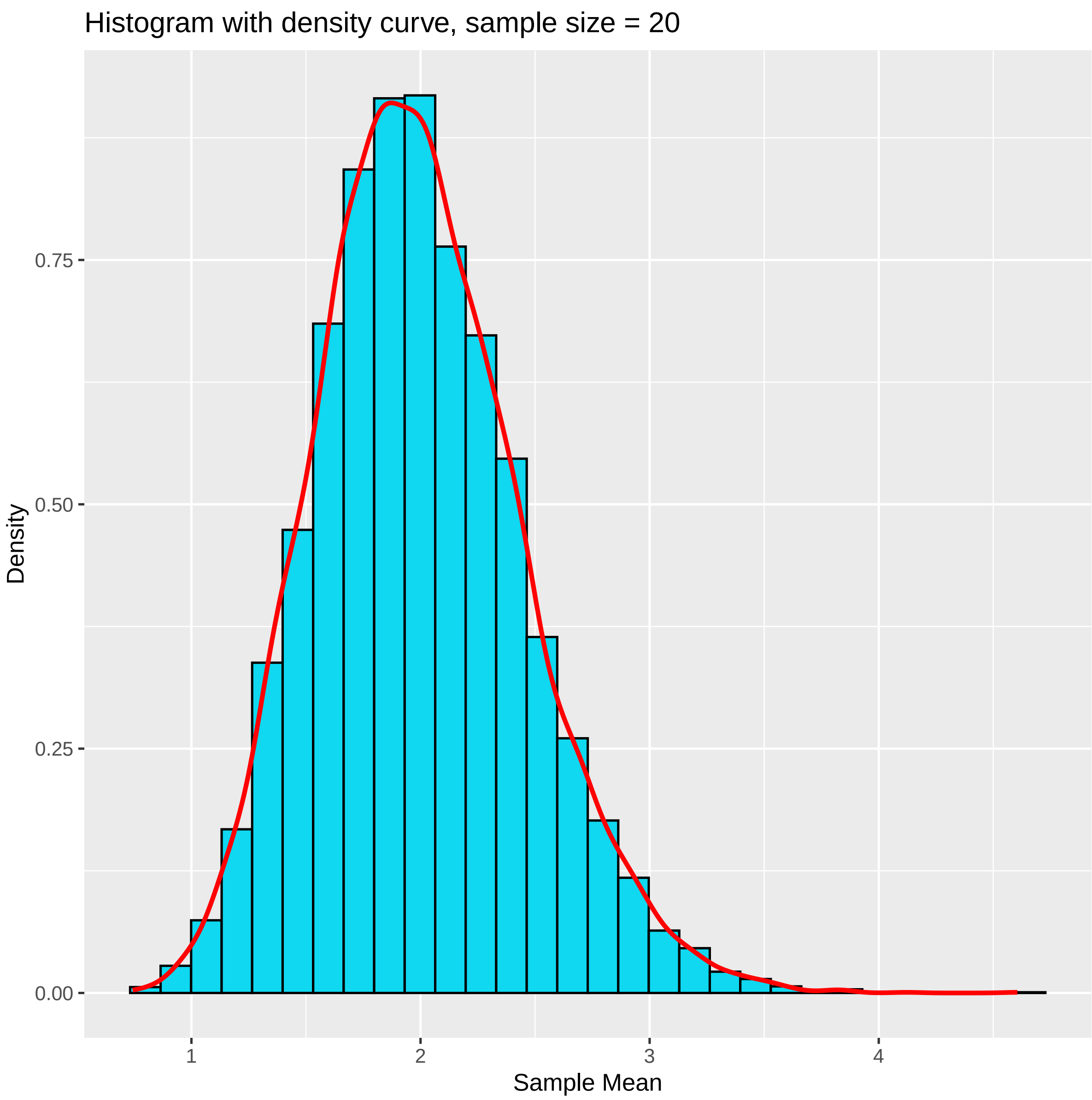
```
sample_list <- list(); size <- 20

for(i in 1:10000){
  sample_list[[i]] <- rexp(size, rate = 1 / theta)
}

x_bar <- sapply(sample_list, mean)
```

```
df2 <- data.frame(means = x_bar)

df2 %>%
  ggplot(aes(x = means)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "#0FD8F0", col = "black") +
  geom_density(col = "red", linewidth = 1) +
  labs(x = "Sample Mean", y = "Density",
       title = "Histogram with density curve, sample size = 20")
```



✎ The density curve has pretty much resemblance with that of a normal density curve, thus implying convergence in distribution of the sample mean.

Hence it is established that the sample mean is a CAN estimator for θ.

☞ Now we have to obtain a CAN estimator for $\psi(\theta) = e^{-t/\theta}$. $\psi(\theta)$ is differentiable function of θ, $\dfrac{d\psi}{d\theta}$ is non-vanishing and continuous. We already have $\bar{X}$ is a CAN estimator for θ. Thus by invariance property of a CAN estimator, $\psi(\bar{X})$ is a CAN estimator of $\psi(\theta)$ and

$$\psi(\bar{X}) \sim AN\left(\psi(\theta), \frac{\theta^2}{n}\left(\frac{d\psi}{d\theta}\right)^2\right).$$

So asymptotic variance of $\psi(\bar{X})$ is $\dfrac{t^2}{n\theta^2}e^{-2t/\theta}$.

Here we take $t = 1$ and $\theta = 2$.

So $\psi(\theta) = 0.6065307$.

```
estimate <- function(n) return(exp(-1/mean(rexp(n, rate = 1/2))))

asv <- function(n) return(exp(-1) / (n * 4))
```

```
df3 <- data.frame(sample_size = seq(5, 50, 5),
                  estimated_psi_theta = sapply(seq(5, 50, 5), FUN = estimate),
                  variance = asv(seq(5, 50, 5)))
df3

##    sample_size estimated_psi_theta    variance
## 1            5           0.3734812 0.018393972
## 2           10           0.5506846 0.009196986
## 3           15           0.5726035 0.006131324
## 4           20           0.6453451 0.004598493
## 5           25           0.6252023 0.003678794
## 6           30           0.6857392 0.003065662
## 7           35           0.4983047 0.002627710
## 8           40           0.5112385 0.002299247
## 9           45           0.5741369 0.002043775
## 10          50           0.5627382 0.001839397
```

✎ Thus we get estimates of the parametric function for different sample sizes. The variance of the estimate decreases as sample size increases.