Question set 3

Deep Hariya, BS2324

Contents

Question 1:

We have a random sample of size 101 from a $N(\mu, 1)$ population, where $\mu \in \mathbb{R}$ is unknown. We want to estimate μ. There are two contending estimators: the sample mean and the sample median (i.e., the central-most value after you order the random sample). We want to approximate the standard errors of these estimators. Do this using simulation for $\mu = 10$ and $\mu = 20$.

Solution:

Here we first generate a random sample of size 101 with *true mean value* $\mu = 10$ and $\mu = 20$. Then to compare `mean()` and `median()` we find the standard error of mean and median:

Case 1: When *true mean* is 10:

```
mu1 = 10
means1 = c()
medians1 = c()
for(i in 1:10000){
  v = rnorm(101, mu1, 1)
  Mn = mean(v)
  Md = median(v)
  X = c(Mn,Md)
  means1 = append(means1, X[1])
  medians1 = append(medians1, X[2])
}
mean_error1 = sd(means1)
median_error1 = sd(medians1)
mean_error1
```

```
## [1] 0.09991808
```

```
median_error1
```

```
## [1] 0.1255741
```

We get standard error of mean is 0.0999181 and the standard error of median is 0.1255741

Case 2: When *true mean* is 20:

```
mu2 <- 20
means2 <- c()
medians2 <- c()
for(i in 1:10000){
  k <- rnorm(101, mu2, 1)
  Mn <- mean(k)
  Md <- median(k)
  X <- c(Mn,Md)
  means2 <- append(means2, X[1])
  medians2 <- append(medians2, X[2])
}
mean_error2 <- sd(means2)
median_error2 <- sd(medians2)
print(mean_error2)
```

```
## [1] 0.09947579
```

```r
print(median_error2)
```

```
## [1] 0.125826
```

We get standard error of mean is 0.0994758 and the standard error of median is 0.125826

From above experiment we can see that standard error of median is always higher than the standard error of mean. Hence we can conclude that mean is always a better estimator than median.

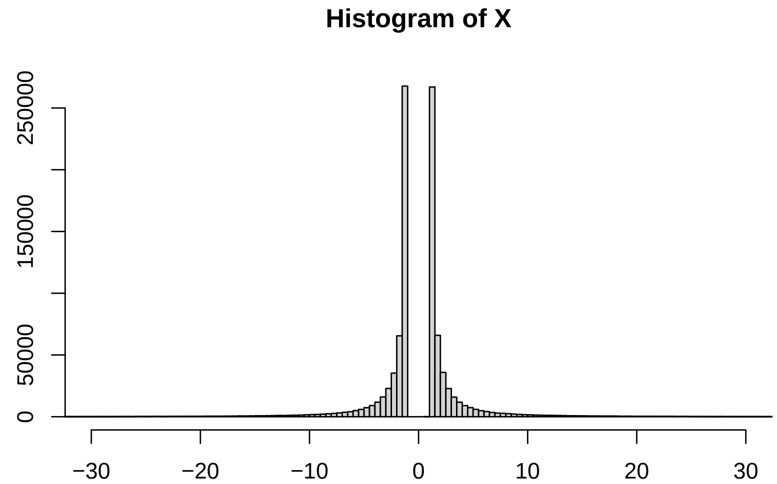
Question 2:

A point, P, is chosen at random on the circumference of the unit circle centered at the origin. All points are equally likely. Let $(X, 0)$ be the point where the tangent hits the x-axis. Take $X = 0$ if the P is at $(0, -1)$ or $(0, 1)$. Use simulation to form an idea about the distribution of X. Is the distribution normal? Answer this question by overlaying the best normal PDF on the histogram, and then visually ascertaining the fit.

Solution:

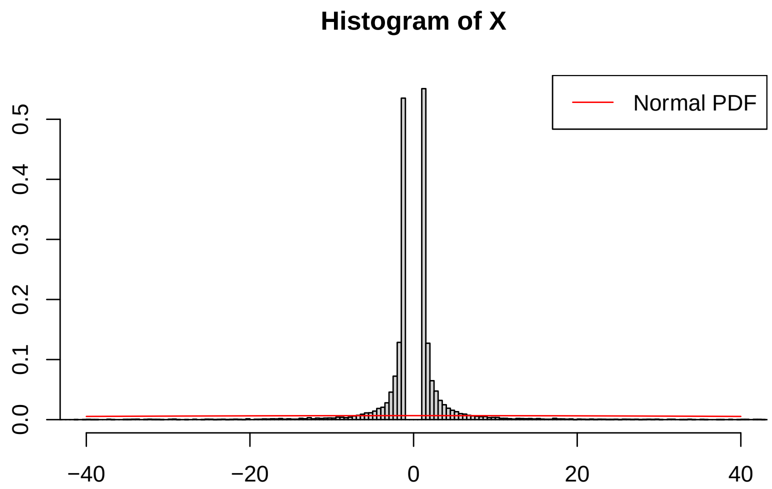
Here we first generate a uniform random sample of θ from $(0, 2\pi)$ and the X is nothing but $\sec(\theta)$, so we make histogrm of X.

```
set.seed(34857)
X <- 1/cos(runif(1000000, 0, 2*pi))
hist(X,xlim = c(-30,30), n = 1000000,xlab="",ylab='')
```



This is clearly not a normal distribution, to demonstrate this we can try to fit in a normal distribution.

```
library(MASS)
X <- 1/cos(runif(10000, 0, 2*pi))
hist(X, n = 10000,prob = T,xlim=c(-40,40),xlab="",ylab='')
fit <- fitdistr(X, "normal")
para <- fit$estimate
curve(dnorm(x, para[1], para[2]), col = 'red', add = TRUE)
legend("topright", legend = "Normal PDF", col = 'red',lwd = 1,lty  = 1)
```



As we can clearly see, our histogram doesnt fit the normal distribution curve, hence this is not a normal distribution

Question 3:

The same set up as above. Find (using simulation) two numbers L and U such that X lies between them with 90% probability. The smaller is $U - L$, the happier I would be. Also find (using simulation) the probability that X exceeds 5.

Solution:

Part 1: Same setup as before, We first get X as $\sec(\theta)$, where θ is chosen uniformly at random. Here we use `quantile()` function to get values for U and L

```
T <- runif(1000000, 0, 2*pi)
X <- 1/cos(T)
U <- quantile(X,0.05)
L <- quantile(X,0.95)
print(c(U,L))
```

```
##        5%      95%
## -6.40357  6.40647
```

And hence you get U = -6.4035695 and L = 6.4064698

Part 2: And to get probability $P(X > 5)$ we can do:

```
T <- runif(1000000, 0, 2*pi)
X <- 1/cos(T)
P <- 100*mean(X>5)
print(P)
```

```
## [1] 6.4317
```

$\therefore$ we get probability $P = 6.4317\%$