STA511 Homework #5

Abbas Rizvi

November 30, 2015

1. (a) A simulation was conducted that estimated the $Pr(Reject H_0 | X_1, X_2, ..., X_{20} \sim N(0.5, 1))$ under the conditions of rejecting H_0 when $\left| \frac{\bar{X}}{1/\sqrt{n}} \right| > 1.96$ where $\bar{X} = \sum X_i / n$ and $n = 10$. The probability of rejecting H_0 was 0.447 when the seed is set to `set.seed(333)`. The simulation was conducted in R and the code can be seen below.

```
> rm(list=ls())
> set.seed(333)
> n <- 20
> x <- rnorm(n, mean = 0.5, sd = 1)
>
> xbar <- c()
> for (i in 1:1000){
+         isample <- sample(x, 20, replace=T)
+         xbar[i] <- sum(isample)/n
+ }
>
> values <- abs(xbar/(1/sqrt(20)))
> test <- subset(values, values > 1.96, drop=FALSE)
> length(test)/1000
[1] 0.447
```

(b) Part (a) can be seen as a Monte-Carlo integration problem. Monte-Carlo integration methods are sampling methods that are based on probability theory. Intuitively, they rely on the law of large numbers and central limit theorem. That is, for part (a), we reject H_0 when $\left| \frac{\bar{X}}{1/\sqrt{(n)}} \right| > 1.96$, colloquially stating that we are rejecting that population mean of the model is equal to 0 ($\mu = 0$) with 95% confidence. This can be sampled by Monte-Carlo integration because we have a distribution with a known parameter (σ^2) and an unknown parameter μ. This confidence interval can be transformed into a function of a known cumulative density function (CDF) by using probability theory. Herein, we can estimate the CDF by computing the sample mean of our data's distribution over many iterations. The more iterations that are conducted of the sample mean, the more likely that the true value of the model's mean will converge as n approaches infinity.

2. Let $X_1, X_2, ..., X_n \sim Bin(10, \theta)$. We assume the data follows a Binomial model of size 10 and probability of success θ.

(a) Find the MLE for θ.
The likelihood function of a binomial is:

$$L_n(\theta; x) = \frac{n!}{x!(n-x)!} \theta^x (1-\theta)^{n-x}$$

$$L_n(\theta) = \prod_{i=1}^{n} \frac{n!}{x_i!(n-x)!} \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$$

The likelihood function is regarded as a function of only parameter θ, such that, the factor $\frac{n!}{x!(n-x)!}$ is a fixed constant and does not affect the MLE.

$$logL(\theta) = \sum_{i=1}^{n} x_i log(\theta) + \sum_{i=1}^{n} (10 - x_i) \cdot log(1-\theta)$$

We take the derivative of this function with respect to θ and subsequently set $= 0$.

$$\frac{\delta logL(\theta)}{\delta \theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{10 - \sum_{i=1}^{n} x_i}{1 - \theta} = 0$$

1

$$(1 - \theta) \sum_{i=1}^{n} x_i - \theta\left(10 - \sum_{i=1}^{n} x_i\right) = 0$$

$$\sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} x_i - 10\theta + \theta \sum_{i=1}^{n} x_i = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{n} = 10\theta$$

$$\frac{\sum_{i=1}^{n} x_i}{10n} = \hat{\theta}_{MLE}$$

(b) Find the Method of Moments (MOM) estimator for θ.

$$\hat{\theta}_{MOM}$$

Model: $f(x|\theta)$

Left hand side is the expected value of the binomial model.

$$E[X] = n\theta$$

Now plugging in 10 for n.

$$E[X] = 10\theta$$

Right hand side is the sample mean $\bar{x}$. And $\bar{x}$ can be expanded as $\frac{\sum_{i=1}^{n} x_i}{n}$.

When both sides of the equation are set to each other:

$$10\theta = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{\theta}_{MOM} = \frac{\sum_{i=1}^{n} x_i}{10n}$$

3. We wanted to determine whether our skewness functional was a good estimator for the skewness of our data. Our data was represented as $X_i = e^{Y_i}, i = 1, ..., n$), where $Y_{i]}$ was drawn from $Y_1, ..., Y_n \sim N(0, 1)$. Let $n = 25$.

Monte Carlo methods for evaluating integrals were used to determine the point estimator of skewness $\hat{S}$:

$$\theta F = \int \frac{(x - \mu)^3 f(x)}{\sigma^3} dx$$

$$E(g(Y)) = \int (gY))(f(Y))dY$$

$$g(Y) = \frac{(x - \mu)^3}{\sigma^3}$$

$$\hat{S} = E(g(Y)) = \frac{1}{n} \cdot \sum_{i=1}^{n} g(Y_i)$$

```
> set.seed(333)
> n <- 25
> x <- exp(rnorm(n,mean=0,sd=1))
> mu <- mean(x)
> sigma <- sqrt(var(x))
>
> estimate.skew <- sum((x-mu)^3/sigma^3)/n
> estimate.skew
[1] 2.272668
```

The point estimate of $\hat{S}$ was determined to be `2.273`

Subsequently, bootstrapping methods were used to compute the standard error (denoted as v_{boot}) of $\hat{S}$ with 95% confidence intervals using $\hat{S} \pm 1.96 v_{boot}$. Bootstrapping was conducted in order to simulate the standard error 1000 times and applied to compute the confidence intervals of our functional distribution. The bootstrapping method over 1000 simulations determined that our computed confidence interval contains the true value of skewness 59.7% times relative to total number of simulations.

```
>
> vboot <- c()
> Tboot <- c()
> for (j in 1:1000){
+          for (i in 1:1000){
+                    boot.obs <- sample(x, n, replace=T)
+                    skew <- (boot.obs-mu)^3/sigma^3
+                    Tboot[i] <- mean(skew)
+          }
+          vboot[j] <- sqrt(var(Tboot))
+ }
>
> upperbound <- estimate.skew + 1.96*vboot
> lowerbound <- estimate.skew - 1.96*vboot
>
> CIs<- cbind(lowerbound,upperbound)
> true.skew <- (exp(1) + 2)*(sqrt(exp(1)-1))
> true.skew
[1] 6.184877
> true.values <- subset(CIs, CIs[,1] <= true.skew & CIs[,2] >= true.skew, drop=FALSE)
>
> length(true.values)/length(CIs) * 100
[1] 59.7
```