Problem Set 5–MATH391

Theodore Dounias

3/1/2018

6.39

I will use the definition of consistency:

$$\lim_{n\to\infty} P(|X_{max} - b| < \epsilon) = \lim_{n\to\infty} P(-\epsilon < X_{max} - b < \epsilon)$$

However, since X_{max} takes values from zero to b, $X_{max} - b < 0 < \epsilon$ in all cases, so the right side of the inequality is always true, and so redundant.

$$\lim_{n\to\infty} P(X_{max} > b-\epsilon) = \lim_{n\to\infty}(1-P(X_{max} < b-\epsilon)) = \lim_{n\to\infty}(1-\prod_{i=1}^{n} P(X_i < b-\epsilon)) = \lim_{n\to\infty}(1-(F_X(b-\epsilon))^n) = 1$$

The last step is true since the cdf is a quantity always between 0 and 1–and here it cannot be 1 because $b - \epsilon < b$ and the cdf is strictly increasing–which means the limit of the cdfs is zero, and the total limit is equal to one.

6.40

By theorem 6.2, we know that:

$$E(\hat{\sigma}_n^2) = \sigma^2 \frac{n-1}{n} \Rightarrow \lim_{n\to\infty}(E(\hat{\sigma}_n^2)) = \sigma^2$$

Using a result from excercise 6.27 of the previous problem set we have:

$$Var(\hat{\sigma}_n^2) = \frac{\sigma^4}{n^2} 2(n-1) \Rightarrow \lim_{n\to\infty} Var(\hat{\sigma}_n^2) = 2 \lim_{n\to\infty} \frac{\sigma^4}{n} - \frac{\sigma^4}{n^2} = 0$$

Therefore, byt Proposition 6.6, the estimator is consistent.

7.3

A.

```r
sd <- 50/sqrt(100)
sim_mean <- 210

CI_1 <- c(sim_mean - qnorm(.95)*sd, sim_mean - qnorm(0.05)*sd)
CI_1
```

```
## [1] 201.7757 218.2243
```

B. The margin of error can be compued as $(q_2 - q_1)\frac{\sigma}{\sqrt{n}}$. By using R for the quantiles we obtain:

```r
(qnorm(.975) - qnorm(0.025))*(50/2)
```

```
## [1] 97.9982
```

So we have: $\sqrt{n} = 9.8 \Rightarrow n = 96.04$, so we need 97 samples.

C. Repeating the process for a 99% CI:

```r
(qnorm(.995) - qnorm(0.005))*(50/2)
```

```
## [1] 128.7915
```

So we have: $\sqrt{n} = 12.9 \Rightarrow n = 166.41$, so we need 166 samples. There is a difference between being X% confident and having a margin of error be at most something, which is reflected here in the decision to round up or down.

7.8

```r
miss_function<- function(n){
  set.seed(251981)
  tooLow <- 0
  tooHigh <- 0
  q <- qt(.975, n-1)
  N <- 10^4
  for(i in 1:N){
    x <- rgamma(n, shape = 5, rate = 2)
    xbar <- mean(x)
    s <- sd(x)
    L <- xbar - q*s/sqrt(n)
    U <- xbar + q*s/sqrt(n)
    if(U < 5/2) tooLow <- tooLow + 1
    if(L > 5/2) tooHigh <- tooHigh +1

  }
  (tooLow + tooHigh)/N
}

miss_function(30)
```

```
## [1] 0.0556
```

```r
miss_function(60)
```

```
## [1] 0.0529
```

```r
miss_function(100)
```

```
## [1] 0.0493
```

```r
miss_function(250)
```

```
## [1] 0.0507
```

The way that the percentage of misses falls and then rises again might be suggesting some sort of quadratic term in the relationship, but it seems like sample size is relativelly independent of the percentage of times we miss the true mean in the CI.

7.34

As is heavily sugested by the excercise, I will use $2\lambda X$ as a pivot to construct a confidence interval:

$$P(q_1 < 2\lambda X < q_2) = P(\frac{q_1}{2X} < \lambda < \frac{q_2}{2X})$$

We can use R to derive the quantiles:

```r
qchisq(c(.025, .975), 4)
```

```
## [1]  0.4844186 11.1432868
```

And so the confidence interval is: $(\frac{.484}{2X}, \frac{11.143}{2X})$, where X is our observation from the data.