STA 032 R Homework 6

Hardy Jones
999397426
Professor Melcon
Winter 2015

1. Using the scheme given in the assignment,
$$\begin{pmatrix} \overline{X} - Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} & \overline{X} + Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \\ \overline{X} - t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}} & \overline{X} + t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \end{pmatrix}$$
   (a) $\begin{pmatrix} 3.72 & 7.52 \\ 3.43 & 7.82 \end{pmatrix}$
   (b) $\begin{pmatrix} 2.34 & 8.13 \\ 1.89 & 8.58 \end{pmatrix}$
   (c) $\begin{pmatrix} 3.47 & 5.79 \\ 3.43 & 5.84 \end{pmatrix}$
   (d) $\begin{pmatrix} 2.70 & 4.76 \\ 2.66 & 4.80 \end{pmatrix}$

2. (a) The proportion of `N` confidence intervals that cover the true mean for $Z_{\frac{\alpha}{2}}$ is 0.930. The proportion of `N` confidence intervals that cover the true mean for $t_{n-1;\frac{\alpha}{2}}$ is 0.951.
   (b) The proportion of `N` confidence intervals that cover the true mean for $Z_{\frac{\alpha}{2}}$ is 0.893. The proportion of `N` confidence intervals that cover the true mean for $t_{n-1;\frac{\alpha}{2}}$ is 0.913.
   (c) The proportion of `N` confidence intervals that cover the true mean for $Z_{\frac{\alpha}{2}}$ is 0.944. The proportion of `N` confidence intervals that cover the true mean for $t_{n-1;\frac{\alpha}{2}}$ is 0.950.
   (d) The proportion of `N` confidence intervals that cover the true mean for $Z_{\frac{\alpha}{2}}$ is 0.931. The proportion of `N` confidence intervals that cover the true mean for $t_{n-1;\frac{\alpha}{2}}$ is 0.936.
   (e) The $t_{n-1;\frac{\alpha}{2}}$ coverage is better than the $Z_{\frac{\alpha}{2}}$ coverage. This is because the sample size of each simulation was smaller than 30.
   (f) The $t_{n-1;\frac{\alpha}{2}}$ coverage is better than the $Z_{\frac{\alpha}{2}}$ coverage. This again is because the sample size of each simulation was smaller than 30.
   (g) The coverages reported in part (a) were higher than the coverages in part (b). This is probably due to the fact that the population in part (a) was normally distibuted, while the population in part (b) was exponentially distributed.

(h) The coverages reported in part (c) were again slightly higher than the coverages in part (d). However, the difference is less drastic between the two coverages. The Central Limit Theorem is starting to take effect, so it matters less that the population in part (d) was exponentially distributed.

Appendix A R code

Problem 1

```
confidence <- function(x, alpha) {
  n <- length(x)
  s <- sd(x)
  x.bar <- mean(x)
  z <- qnorm(alpha / 2, lower.tail = FALSE)
  t <- qt(alpha / 2, n - 1, lower.tail = FALSE)
  s.root.n <- s / sqrt(n)
  z.sd <- z * s.root.n
  t.sd <- t * s.root.n
  vals <- c(x.bar - z.sd, x.bar - t.sd, x.bar + z.sd, x.bar + t.sd)

  matrix(vals, 2, 2)
}
```

Problem 2

```
source("./prob1.R")

proportion <- function(pop, alpha, size, simulations, true.mean) {
  # This is a `size x simulations` matrix.
  sims <- replicate(simulations, sample(pop, size))

  # Construct a confidence interval for each simulation.
  # This is a `4 x simulations` matrix
  confs <- apply(sims, 2, function(sim) confidence(sim, alpha))
  # Grab row vectors of each respective low and high in the interval.
  # Each is a `1 x simulations` row vector.
  z.lows <- confs[1, ]
  z.highs <- confs[3, ]
  t.lows <- confs[2, ]
  t.highs <- confs[4, ]

  covered.mean <- covered(true.mean)

  # Find out how many intervals covered the true.mean.
  # These are both `1 x simulations` row vectors.
  z.covereds <- mapply(covered.mean, z.lows, z.highs)
  t.covereds <- mapply(covered.mean, t.lows, t.highs)

  # Using the duck typing of R, take the mean of both covered vectors.
  c(mean(z.covereds), mean(t.covereds))
}

# Helper function that curries its arguments.
covered <- function(val)
  function(low, high)
    low <= val && val <= high
```