STA 250: Statistics
Lab 7

In this lab, we will look at Bayesian and classical interpretations of reported parameter range.

The setting

Consider our food study where weekly food expenditures $X_1, \cdots, X_n$ are recorded for n Duke students. We model the data as $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$, both unknown. We will look at producing 95% intervals for μ.

As we saw in class, a 95% ML confidence interval for μ is

$$\bar{x} \mp z_{n-1}(\alpha)\frac{s_x}{\sqrt{n}},$$

which is most efficient (i.e., has shortest average length) among a wide class of 95% confidence interval rules.

Earlier we discussed a conjugate Bayesian analysis of the model with a (contextual) prior distribution $\xi(\mu, \sigma^2) = \mathsf{N}\chi^{-2}(m, k, r, s^2)$. We had elicited $m = 200$, $k = 1$, $r = 1.81$ and $s = 42.44$. Recall that the posterior is $\mathsf{N}\chi^{-2}(\tilde{m}, \tilde{k}, \tilde{r}, \tilde{s}^2)$ with

$$\tilde{m} = \frac{km + n\bar{x}}{k+n}, \quad \tilde{k} = k + n, \quad \tilde{r} = r + n$$

$$\tilde{s}^2 = \frac{1}{r+n}\left\{rs^2 + \frac{kn}{k+n}(\bar{x} - m)^2 + (n-1)s_x^2\right\}$$

and under the posterior $\sqrt{\tilde{k}}(\mu - \tilde{m})/\tilde{s} \sim t(\tilde{r})$.

Intervals based on Fall 2012 data

In Fall 2012, $n = 21$ students enrolled in this course reported the following weekly expenditures:

```
297, 220, 150, 150, 120, 159, 130, 150,  45, 235, 203,
115, 120, 112, 105, 123, 220,  73, 150, 160, 100
```

Task 1. Evaluate the 95% ML confidence interval for μ based on above data.

Task 2. What posterior probability would you assign to μ being in the above interval under the Bayesian analysis with our chosen prior?

Task 3. What is the 95% posterior range for μ under the Bayesian analysis with our chosen prior?

Coverage of the Bayesian 95% posterior range

We could view the 95% posterior range under our chosen prior as another interval rule and subject it to classical guarantee calculations. In particular we could evaluate its coverage of μ at different values of μ and σ. The code snippet below shows how to numerically calculate the coverage of the 95% ML confidence interval at a given parameter value.

```
is.included <- function(a, interval) return(prod(a - interval) <= 0)

expt <- function(n, mu, sigma, int.rule, ...){
  x <- rnorm(n, mu, sigma)
  interval <- int.rule(x, ...)
  return(is.included(mu, interval))
}

ml.int <- function(x, conf = 0.95){
  alpha <- 1 - conf
  n <- length(x)
  x.bar <- mean(x)
  s.x <- sd(x)
  return(x.bar + c(-1,1) * qt(1 - alpha/2, n - 1) * s.x / sqrt(n))
}

coverage.ml <- mean(replicate(1e4, expt(21, 200, 40, ml.int)))
cat("ML coverage =", round(100 * coverage.ml), "%\n")
```

TASK 4. Write down a function `bayes.int()` that will produce the 95% posterior range for μ under our choice of the prior for any plausible data x.

TASK 5. Numerically approximate the coverage of the interval rule that uses the Bayes 95% posterior interval for $\mu = 200$ and $\sigma = 40$.

TASK 6. Repeat task 5 but with $\mu = 150$ and $\sigma = 30$.