Chapter 7 Review with Applications for Math 30: Sampling Distributions and More

1. Theory Practice.

Let $X_1, \cdots, X_n$ be a random sample from an Exponential(β) distribution. Remember we thought this was a good distribution to use to model whale velocities.

a. Find unbiased estimators for $1/\beta$ - average velocity of the whales, $1/\beta^2$ - variance of the whale velocities, and $(\beta - 1)/\beta^2$. (Hint: start with $\bar{X}$ and sample variance.)

b. In general, could you conclude that $1/\bar{X}$ is unbiased for $1/\mu$? To think about this more simply, think about just a single observation, is $1/X$ unbiased for $1/\mu$ (you can/should write out the expectation to look at it) for a general distribution with pdf $f(x)$ and mean μ?

c. Find the Fisher information about $\theta = 1/\beta$ in a single observation X.

d. What would the Fisher information be about $\theta = 1/\beta$ in a random sample of n observations?

e. What is the lower bound on the variance for an estimator of $\theta = 1/\beta$? Does your unbiased estimator from a. achieve it? (I.E. is your estimator in a. for $1/\beta$ efficient?)

f. Show that $2\beta\sum_{i=1}^{n} X_i$ has a chi-squared distribution with $2n$ degrees of freedom. (Hint: Think about results we know for sums of Gamma or χ^2 RVs.)

g. Based on the CLT, we know that $\bar{X}$ is approx. normal, and we can standardize to make a standard normal random variable, $Z = \frac{\bar{X}-1/\beta}{(1/(n\beta^2))^{1/2}}$. What distribution would Y (see below) have? Reduce the expression filling in what Z would be. Does it look like this quantity might be useful for making CIs for β or $1/\beta$?

$$Y = \frac{Z}{\left(\frac{2\beta\sum X_i}{2n}\right)^{(1/2)}}$$

h. Noting that the variable in g. might be hard (or at least not very appealing) to use to make CIs, try using just Z, which is approximately standard normal. Give a formula for a 90 percent CI for β. Note the .95 quantile from the normal distribution is 1.645.

2. Application. Suppose we need to analyze lead and copper samples in drinking water to determine if the water is safe (within environmental contaminant ranges provided by the EPA). We have 10 samples from the Crystal Lake Manors subdivision in Tampa, Florida (source: Florida Water Dept. Laboratory). Looking at the samples, the sample of lead does not appear normally distributed, while the sample of copper does (see data analysis sheet), so we focus on copper. We will then assume that the copper samples are taken from a Normal distribution with unknown mean μ and unknown variance σ^2.

a. Provide a 99 percent CI for the mean, μ, amount of copper present based on the sampling distribution of the sample mean and using sample variance (i.e. frequentist).

b. What Bayesian prior in the Normal-Gamma family would you need to use in order for a similar 99 percent CI to match the frequentist CI? Provide all hyperparameters.

c. Try a different Normal-Gamma prior: Assume $\mu_0 = .5, \lambda_0 = 1, \alpha_0 = 48$, and $\beta_0 = 6$. Find the relevant posterior distribution.

d. Compute 99 percent CIs for μ based on the Bayesian prior and posterior distributions.

e. Copper is hazardous in levels above 1.3 mg/L (same units we have). This is actually called an "action" level by the EPA. Do any of the relevant CIs (not the prior CI) indicate that the copper levels in the Crystal Lake Manors subdivision are hazardous?

f. Why is it important that the prior chosen for the Bayesian CI results in a 99 percent CI which includes 1.3? If it didn't include 1.3, what would that say about your prior belief as to whether or not the copper levels were hazardous or not?