Imperial College London

Problem Sheet 2

MATH50011
Statistical Modelling 1

Week 2

Lecture 3 (CRLB)

1. In the lecture notes, we sketched the proof of the Cramér-Rao lower bound (CRLB) for continuous random variables. Prove the CRLB for discrete random variables with finite support. (Recall that the *support* of X is the set of values where the pdf/pmf is greater than zero.)

2. Find the CRLB for estimating θ based on a random sample of size n from the following distributions

   (a) Exponential(θ);

   (b) Normal(θ, σ^2) with known $\sigma^2 > 0$;

   (c) Bernoulli(θ); (see Example 8)

   (d) Poisson(θ).

3. For which of the distributions in 2(a-d) can the sample mean be used to construct an unbiased estimator T with variance equal to the CRLB for estimating θ?

4. Suppose that we wish to estimate θ based on a random sample $X_1, \dots, X_n$ of Bernoulli(θ) random variables. However, we are only able to obtain a random sample $(Y_i, R_i), \dots, (Y_n, R_n)$ where the R_i's are iid Bernoulli(p_0) for known p_0 and $Y_i = R_i X_i$ for $i = 1, \dots, n$. Compare the CRLBs for estimating θ based on

   (a) The full data distribution of the X_i's;

   (b) The marginal distribution of the Y_i's;

   (c) The joint distribution of the (Y_i, R_i)'s.

Lecture 4 (Consistency)

5. Show that an asymptotically unbiased estimator sequence need not be consistent. (Hint: consider estimating μ based on a sequence of independent rv's $X_i \sim N(\mu, 2i)$ for $i = 1, 2, 3, \ldots$)

6. Show that a consistent estimator sequence T_n need not be asymptotically unbiased. (Hint: consider a sequence (T_n, Y_n) with $Y_n \sim$ Bernoulli(1/n) and $T_n | Y_n = 0 \sim N(\theta, \sigma^2/n)$ and $T_n | Y_n = 1 \sim N(n^2, 1)$.)

7. Let $X_1, X_2, \ldots$ be iid Uniform$(0, \theta)$ random variables and define $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$.

   (a) Show that $\hat{\theta}_n$ is asymptotically unbiased and consistent.

   (b) Find a sequence of constants a_n such that $a_n \hat{\theta}_n$ is unbiased and consistent.

   (c) Compare the MSE of $\hat{\theta}_n$ and $a_n \hat{\theta}_n$.

8. Let $X_1, X_2, \ldots$ be iid Bernoulli(θ) random variables and consider estimating $g(\theta) = \text{Var}(X_1) = \theta(1 - \theta)$. Define the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.

   (a) Show that $T_n = \bar{X}_n(1 - \bar{X}_n)$ is asymptotically unbiased and consistent.

   (b) Find a sequence of constants a_n such that $a_n T_n$ is unbiased and consistent.

   (c) Compare the MSE of T_n and $a_n T_n$.

R lab: Consistency of the sample median

This exercise is intended to reinforce concepts through use of the R software package.

We consider an example of a simulation study. The simulation consists of 1000 independent replications of the following process:

i. Generate $X_1, \dots, X_n$ iid $N(\mu, 1)$ with $\mu = 0$;

ii. Estimate μ by calculating the sample median m_n;

iii. Record whether the statement $|m_n - \mu| < \epsilon$ is *true* or *false*.

Lastly, the proportion of times/1000 that the statement in iii. is *true* is calculated.

9. In R, the code below implements the simulation study for $n = 10$ and $\epsilon = 0.1$.

```
set.seed(50011)
result <- logical(length = 1000)
n <- 10
epsilon <- .1
for(i in 1:1000){
   X <- rnorm(n, mean = 0)
   m <- median(X)
   result[i] <- abs(m - 0) < epsilon
}
mean(result)
```

   Note that the command `set.seed(50011)` ensures that you obtain the same results each time you run this set of commands.

   Type the above commands into your R console (or write a script) and then:

   (a) Explore how the value of `mean(result)` changes by increasing the value of `n` in this code to, e.g. $n = 30, 50, 100, 200, 500, 1000$.

   (b) Referring to the results of your experimentation, comment on whether the sample median appears to be consistent for μ in this setting.