Lecture March, 22, 2013.

1. When we take a sample of n independent observations $X_1, \ldots, X_n$ from a $N(m, \sigma^2)$ population, the sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ can be shown to verify:

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}.$$

   Using this fact (and remembering that a χ^2_n distribution has mean n and variance $2n$),

   (a) Obtain an unbiased estimator $\hat{\sigma}^2$ of σ^2.

   (b) Compute its variance.

   (c) Compute the Cramer-Rao lower bound for the variance of unbiased estimators of σ^2.

   (d) Is $\hat{\sigma}^2$ efficient?

2. Write the expression of the Cramer-Rao lower bound for the variance of unbiased estimators of a parameter based on n observations, under regularity conditions. Then answer the following questions:

   (a) If an unbiased estimator attains the Cramer-Rao lower bound, can we assert that its variance decreases to zero as $n \to \infty$? At which rate?

   (b) Is such an estimator consistent?

3. Let X be distributed in $(0, \theta)$ with density $f_X(x) = 3x^2/\theta^3$. Consider the estimator $\hat{\theta} = \frac{4}{3}\overline{X}$. Is it unbiased? What is its variance? Can we use the Cramer-Rao lower bound to check its efficiency?

4. According to a well known model (the Hardy-Weinberg equilibrium), the offspring of a generation in which two characters have frequencies p and $q = 1 - p$ are distributed in three classes, with probabilities p^2, $2pq$ and q^2. If we have a sample in which those three classes have absolute frequencies 158, 475 and 362, write the equation giving the MLE of p.

5. X is distributed as binary, with parameter p. We know that $\frac{1}{3} \leq p \leq \frac{2}{3}$.

   (a) What is the MLE of p based on one single observation?

   (b) Is it unbiased?

   (c) What is its mean squared error if the true p is 0.5?

(d) Is the MLE optimal? (HINT: Consider the —rather silly— estimator $\hat{p} = 0.5$, irrespective of what is the value of the single observation: compute its MSE. Compute now the MSE of the MLE estimator and plot it for different values of p (using the `curve` command of R may be of help). See how this compares with the MSE of the "silly" estimator.)

This simple example is meant to show that the MLE need not be very good if the sample is small.)

6. A plausible estimator of m in a $N(m, \sigma^2)$ distribution would be the sample median, since we know in that distribution the mean m and the median take the same value. It can be shown that, asymptotically, the sample median Me verifies:

$$\sqrt{n}(Me - m) \to N(0, \sigma^2 = (0.5/f_X(m))^2$$

What is the efficiency of Me?

You may find similar and additional problems in many books, such as [3], [2], [5], [4] and [1] (the last, at a higher level than this course).

References

[1] D. R. Cox and D. V. Hinkley. *Problems and Solutions in Theoretical Statistics*. Chapman and Hall, London, 1980 edition, 1980.

[2] A. Garín and F. Tusell. *Problemas de Probabilidad e Inferencia Estadística*. Ed. Tébar-Flores, Madrid, 1991. In the reserved collection, signature AL-519.2(076) an in the general collection, signature 519.2(076.1) TUS.

[3] F.J. Martin-Pliego, J.M. Montero, and L. Ruiz-Maya. *Problemas de Inferencia Estadística*. Editorial AC, 2000. Signatura: AL-519.23 MAR.

[4] L. Ruíz-Maya and F.J. Martín-Pliego. *Fundamentos de Inferencia Estadística*. Thomson - Paraninfo, 2005. Signatura: AL-519.23 RUI.

[5] J.M. Sarabia. *Curso Práctico de Estadística*. Ed. Cívitas, 1993. Signature 591.2(076).