IST, Digital Signal Processing, Lab. #5 – Estimation
Cláudia Soares and Pedro M. Q. Aguiar
March, 2012

In this lab., we will study the behavior of different estimators for a specific problem. In particular, we will have in mind that the performance of an unbiased estimator can be accessed by comparing its variance with the Cramér-Rao bound (CRB). This bound, which depends on the problem at hand, *i.e.*, on the model of the data, represents the smallest possible variance of an unbiased estimator.

Goal scoring in a soccer game. Recently, researchers have proposed[1] that the number of goals g a team scores against a particular opponent can be modeled as a Poisson random variable, *i.e.*, following

$$\mathrm{P}(g) = \frac{\lambda^g e^{-\lambda}}{g!}, \qquad g = 0, 1, 2, \ldots$$

where λ is a free parameter. Note that $\mathrm{E}\{g\} = \lambda$ and $\mathrm{Var}\{g\} = \mathrm{E}\left\{(g-\lambda)^2\right\} = \lambda$ (from the definition and using the expression above, your can show that these equalities hold).
Three students debate on what should be the score of team i on this week's match against team j. They are aware of the study above and try to find the expected number of goals team i scores in such a game, looking at the results of the past N matches between the same opponents. Let g_n, $n = 1, 2, \ldots, N$ be the number of goals scored by team i against team j in match n between them. Student A notes that the last games are more relevant than the old ones for estimating the forthcoming result, thus should count more; student B says he knows nothing about soccer, proposing a sample mean; student C reminds that the decreasing strength of team i in the current season should be taken into account. Accordingly, they propose the following estimators for λ:

$$\hat{\lambda}_A = \frac{2}{N(N+1)} \sum_{n=1}^{N} n g_n\,, \qquad \hat{\lambda}_B = \frac{1}{N} \sum_{n=1}^{N} g_n\,, \qquad \hat{\lambda}_C = \frac{1}{2N} \sum_{n=1}^{N} g_n\,.$$

You want to check which estimator is the most reliable. Since the matches are distant in time, we consider their results as independent events, *i.e.*, we consider the problem of estimating the parameter λ of a Poisson random variable from a set of N independent observations, $\{g_1, g_2, \ldots, g_N\}$.

Analytical performance evaluation.

1. Find the mean and variance of each of the estimators $\hat{\lambda}_A$, $\hat{\lambda}_B$, and $\hat{\lambda}_C$.

2. Write the log-likelihood function for the estimation problem and find the CRB for the estimate of λ.

3. For $\lambda = 1$, plot the CRB as a function of $N = 1, 2, \ldots 20$. Superimpose to this plot the estimator variances. Which is the best estimator? Why?

Simulation.

1. Using $\lambda = 1$ as the expected number of goals, simulate the past scores of $N = 20$ matches, by generating a 20×1 vector $\mathtt{g}$ with independent samples of a Poisson random variable. {`poissrnd`}

2. Visualize the histogram of $\mathtt{g}$. Specify the bin centers with all the values from 0 to the maximum of g. Is this data model credible, according to your knowledge of the game?

3. Compute the estimates $\hat{\lambda}_A$, $\hat{\lambda}_B$, and $\hat{\lambda}_C$ from the data in $\mathtt{g}$ and compare them with the histogram. Which seems more adequate?

4. To evaluate the performances of the estimators, use a Monte Carlo method: run 1000 times **1.** and **3.**, storing the values of the estimates; at the end, compute the mean square error (MSE) of each estimator with respect to the true value of λ. Comment and draw conclusions about which is the best estimator.

[1] A. Heuer, C. Muller, and O. Rubner, "Soccer: Is scoring goals a predictable Poissonian process?", A Letters Journal Exploring the Limits of Physics, EPL 89 (2010).