T-61.3040 Statistical Signal Modeling Lendasse, Miche, Eirola

Exercise 2, September 22, 2011

1. Assume that a device is used to measure the amount of water running out of a tap during 24 hours. With a probability of 50 percent, someone has mistakenly closed the tap for the whole day (and night). Otherwise, the tap is open at a random position. If the tap is open, water is running out of it with a constant rate all day long. The position of the tap and thus also the total amount of flowing water is uniformly distributed in the interval $[0, 1]$.

   (a) What is the probability distribution for the total amount of water?

   (b) What is the expected value for the total amount of water?

   (c) And the variance?

2. $x(1), ..., x(N)$ are independent observations drawn from the same Gaussian distribution. We estimate the variance of the distribution. Compute

   (a) The maximum likelihood estimator

   (b) An unbiased estimator

   (c) (Demo) A constant multiple of the unbiased estimator found in (b) that minimises the mean squared error (MSE).

   The purpose of this problem is to show that optimisation of different estimator quality criteria leads to somewhat different formulas for the estimators. None of them is generally better than the others.

3. Assume that $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of a parameter θ with variances

$$\mathrm{Var}(\hat{\theta}_1) = \sigma_1^2, \quad \mathrm{Var}(\hat{\theta}_2) = \sigma_2^2.$$

   (a) Show that for any scalar $0 \leq \alpha \leq 1$ the estimator $\hat{\theta}_3 = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2$ is unbiased.

   (b) What is the mean squared error of $\hat{\theta}_3$, when $\hat{\theta}_1$ and $\hat{\theta}_2$ are statistically independent?

   (c) What is the value of α that minimizes the error in problem (b)?

4. Consider a Poisson distributed random variable z with a density function $f(z \mid \lambda) = e^{-\lambda}\lambda^z/z!$. Determine the maximum likelihood estimate of λ using M independent samples $z_1, z_2, \ldots, z_M$.

5. (Bonus point exercise)

   Assume that the numbers $y(0), y(1), \ldots, y(M)$ have been observed. The numbers are modelled linearly using the following model of other observations $x_i(n)$:

$$\hat{y}(n) = \sum_{i=1}^{N} b(i)x_i(n) = \mathbf{x}(n)^T\mathbf{b}, \quad n = 0, 1, \ldots, M$$

The numbers $y(n)$ could be e.g. the temperatures in Kaisaniemi in December of year n and the vectors $\mathbf{x}(n)$ could be temperatures measured in July of year n in various locations in Finland. The functions and observations are assumed to be real-valued. The goal is to find coefficients $b(i)$ that minimise the mean squared error (MSE):

$$J = \frac{1}{M}\sum_{n=0}^{M} e(n)^2 = \frac{1}{M}\sum_{n=0}^{M}(y(n) - \hat{y}(n))^2$$

In other words, the goal is to predict the temperature in Kaisaniemi as accurately as possible in terms of MSE.

(a) Show that minimising the MSE (by means of setting the derivatives to be zero) leads to equations reflecting the orthogonality principle.

(b) Find the coefficients $b(i)$ using the orthogonality principle.

(Hint: gather the observations to vectors, and similarly the coefficients $b(i)$ and the errors $e(n)$.)