Statistical Methods Exercises 7 Autumn 2020

Return your solutions by 12.00 Finnish time on Thursday 5.11.2020 to Moodle course page: `https://moodle.helsinki.fi/course/view.php?id=30207`

1. **A ML estimator**. Assume a Poisson distributed variable n, e.g. number of nuclear decays in a radioactive source per a given time interval, with an expectation value ν.

   (i) What is the maximum-likelihood (ML) estimator for ν, $\hat{\nu}$, given only a single observation of n?

   (ii) Show that the ML estimator $\hat{\nu}$ is unbiased and find its variance.

   (iii) Show further that the ML estimator $\hat{\nu}$ is efficient, i.e. variance of $\hat{\nu}$ equal to the RCF (Rao-Cramér-Fréchet or Cramér-Rao) bound.

   (iv) Determine also the ML estimator for ν, $\hat{\nu}_m$, in the case of m observations of n (hint: derive the ML estimator in the case of 2 observations: n_1 and n_2, and make the generalization to the case of m observations).

2. **ML estimates**. Assume a study of some physics phenomena having a sinusoidal behavior (like a harmonic oscillator), whose occurrence one is able to measure depends on the (absolute) probability density of the oscillation versus time, t (in s). To determine the frequency, f (in Hz), of the oscillation two measurements were made: collecting 20 and 100 occurrences (e.g. times t_i). The results are found in the Moodle course page as `ml_sample_1.txt` and `ml_sample_2.txt`, respectively. Assume pdf $f(t_i)$ for data t_i to be $|sin\,(2\pi \cdot f \cdot t_i)\,|$ and $t_i \in [0, 1]$.

   (i) Determine the ML estimate, $\hat{f}_{20}$, for the data sample with 20 occurrences using the ML method by determining the lnL-value for different f-values. Allow f to vary in the range 0.1 - 10 Hz. Hint: try sufficient many f-values to be able to follow the rapid changes in lnL with f. Don't exclude apriori f values near the beginning or end of the f range. What is the $lnL_{max,20}$-value and its corresponding f-value ($= \hat{f}_{20}$)?

   (ii) Determine the uncertainty on $\hat{f}_{20}$ using the graphical method described in the lecture notes.

   (iii) Determine $\hat{f}_{100}$ and the $lnL_{max,100}$-value for the data sample with 100 occurrences using the same ML method and plot lnL vs $log_{10}f$.

   (iv) Determine the uncertainty on $\hat{f}_{100}$ using the graphical method.

   (v) What is the ratio between the uncertainty on $\hat{f}_{20}$ and the uncertainty on $\hat{f}_{100}$? Is the ratio consistent with the difference of the sizes of the data samples? *Exercise gives max 9 points instead of usual 6.*