Statistical Methods Exercises 9 Autumn 2020

Return your solutions by 12.00 Finnish time on Thursday 19.11.2020 to Moodle course page: `https://moodle.helsinki.fi/course/view.php?id=30207`

1. **Combining correlated measurements with LS**. Consider two partially overlapping samples of a random variable x, with n and m observations of which c observations are common to both samples and rest are independent. Suppose the variance of x, $V[x] = \sigma^2$ is known. Consider the sample means

$$y_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad y_2 = \frac{1}{m}\sum_{j=n-c+1}^{m+n-c} x_j \, .$$

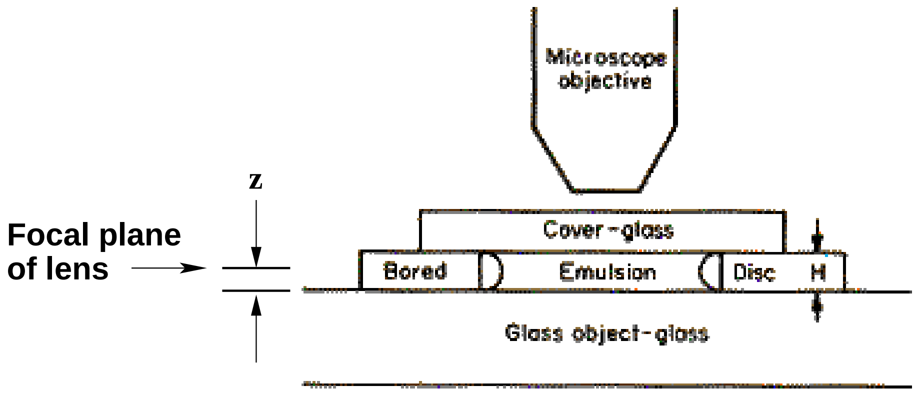
(i) Show that the covariance is

$$\mathrm{cov}\,[y_1, y_2] \quad = \quad \frac{c\sigma^2}{nm} \, .$$

(ii) Using formulas derived in lecture notes (or Cowans book), find the weighted average of y_1 and y_2 and its variance (or standard deviation).

(iii) Show that the weights w_i (i=1,2) used in computing the weighted average of y_1 and y_2 are always positive (or zero) in the case of overlapping samples. *Exercise gives max 6 points.*

2. **ML with binned data**. One of the earliest determination of Avogadro's number was based on Brownian motion. J. Perrin[1] made an experiment like the one below to observe molecular clusters of mastic ("kiviliima"), a substance used in varnish ("lakka"), suspended in water. (Perrin received the 1926 Nobel prize in physics for his work with atoms.)



The clusters can be approximated as spheres of radius $r = 0.52\ \mu$m with a density of 1.063 g/cm^3, i.e. just a bit heavier than water. By viewing the clusters through microscope, only those in a layer $\sim 1\ \mu$m thick were in focus; clusters outside this layer were not visible. By adjusting the microscope lens, the focal plane could be moved vertically. Photographs were taken at 4 different heights z (lowest height arbitrarily set to zero) and the number of clusters $n(z)$ counted. The results were:

height z (μm)	number of clusters n
0	1880
6	940
12	530
18	305

The gravitational potential energy of a spherical cluster of mastic molecules in water is given by $4\pi r^3 \Delta\rho\, gz/3$, where $\Delta\rho = \rho_{mastic} - \rho_{water}$ is the density difference between mastic and water and $g = 9.80$ m/s^2 is the gravitational acceleration. The probability for a cluster to be in an energy state E is according to statistical mechanics proportional to $e^{-E/kT}$, where k is the Boltzmann constant and T the absolute temperature. The clusters should therefore be distributed in height according to an exponential law, where the observed number n at z can be treated as a Poisson variable with mean

$$\nu\,(z) \quad = \quad \nu_0\, e^{-4\pi r^3 \Delta\rho\, gz/3kT}\ , \text{ where } \nu_0 = \nu(z=0)\,.$$

(i) Determine k with the ML method by numerically maximizing $\ln L(k)$ using the data given in table 1 (assume $\nu_0 = n\ (z=0)$). Estimate also the uncertainty on k using the graphical method. The log-likelihood function can here be constructed using the ML method for binned data:

$$\ln L\,(\nu_0, k) \quad = \quad \sum_{i=1}^{N} \left(n_i \ln \nu_i - \nu_i\right)\ ,$$

where N is number of bins ($= 4$) and T the temperature (≈ 293 K).

(ii) Use same log-likelihood formula to determine both k and ν_0, and their uncertainties with binned ML method using the same data. A two dimensional maximization (or minimization if multiplied by -1) is simplest done using minimization routines e.g. fminsearch in Matlab.

(iii) From the k value obtained either in (i) or (ii), determine Avogadro's number (with its uncertainty) using the relation $N_A = R/k$, where R is the gas constant. $R = 8.314$ J/mol·K. *Exercise gives max 12 points.*