GGI 2017
Problems on Statistics (1)
G. Cowan / RHUL

Exercise 1: Using the Kolmogorov axioms, show that

$$P(A \cup B) \;=\; P(A) + P(B) - P(A \cap B)\,.$$

(Hint: express $A \cup B$ as the union of disjoint sets and use Kolmogorov's third axiom.)

Exercise 2: A beam of particles consists of a fraction 10^{-4} electrons and the rest photons. The particles pass through a double-layered detector which gives signals in either zero, one or both layers. The probabilities of these outcomes for electrons (e) and photons (γ) are

$$\begin{array}{lcl} P(0\,|\,\mathrm{e}) = 0.001 & \quad\text{and}\quad & P(0\,|\,\gamma) = 0.99899 \\ P(1\,|\,\mathrm{e}) = 0.01 & & P(1\,|\,\gamma) = 0.001 \\ P(2\,|\,\mathrm{e}) = 0.989 & & P(2\,|\,\gamma) = 10^{-5}\,. \end{array}$$

(a) What is the probability for the particle to be a photon given a detected signal in one layer only?

(b) What is the probability for a particle to be an electron given a detected signal in both layers?

Exercise 3: Consider N independent observations $n_1, \ldots, n_N$ of a Poisson random variable with the same unknown mean value ν.

(a) Write down the likelihood function for the parameter ν. (Since the Poisson distribution is not a pdf but rather a probability, here the likelihood is found directly from the joint probability for the data.) Find the maximum-likelihood estimator for ν.

(b) Show that the estimator is unbiased and find its variance in closed form (use the known mean and variance of a Poisson variable).

(c) Show that the variance of $\hat{\nu}$ is equal to the minimum variance bound (the right-hand side of the information inequality).

Exercise 4: An experiment yields n time values $t_1, \ldots, t_n$, and a calibration value y, all of which are independent. The time measurements are all exponentially distributed with a mean of $\tau + \lambda$ and the calibration measurement, y, follows a Gaussian distribution with a mean λ and a standard deviation σ. Suppose that σ is known and we want to estimate τ and λ.

4(a) Write down the likelihood function for τ and λ, and show that the Maximum Likelihood (ML) estimators for these parameters are

$$\begin{aligned} \hat{\tau} &= \frac{1}{n}\sum_{i=1}^{n} t_i - y\,, \\ \hat{\lambda} &= y\,. \end{aligned}$$

4(b) Find the variances of $\hat{\tau}$ and $\hat{\lambda}$, and the covariance $\mathrm{cov}[\hat{\tau}, \hat{\lambda}]$. Use the fact that the variance of an exponentially distributed variable is equal to the square of its mean. (It may also be useful to note that for any random variables x, y and z, $\mathrm{cov}[x + y, z] = \mathrm{cov}[x, z] + \mathrm{cov}[y, z]$.)

4(c) Show using a sketch how a contour of constant log-likelihood can be used to determine the standard deviations of $\hat{\tau}$ and $\hat{\lambda}$. Explain qualitatively how you would expect the variance of $\hat{\tau}$ to be different if the parameter λ were to be known exactly.

4(d) Show that the (co)variances of $\hat{\tau}$ and $\hat{\lambda}$ obtained from the matrix of second derivatives of the log-likelihood are the same as those found in (b).

Exercise 5: The binomial distribution is given by

$$f(n; N, \theta) = \frac{N!}{n!(N-n)!}\theta^n(1-\theta)^{N-n} \ ,$$

where n is the number of 'successes' in N independent trials, with a success probability of θ for each trial. Recall that the expectation value and variance of n are $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$, respectively. Suppose we have a single observation of n and using this we want to estimate the parameter θ.

5(a) Find the maximum likelihood estimator $\hat{\theta}$.

5(b) Show that $\hat{\theta}$ has zero bias and find its variance. Find the minimum variance bound of $\hat{\theta}$.

5(c) Suppose we observe $n = 0$ for $N = 10$ trials. Find the upper limit for θ at a confidence level of CL $= 95\%$ and evaluate numerically.

5(d) Suppose we treat the problem with the Bayesian approach using the Jeffreys prior, $\pi(\theta) \propto \sqrt{I(\theta)}$, where

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

is the expected Fisher information. Find the Jeffreys prior $\pi(\theta)$ and the posterior pdf $p(\theta|n)$ as proportionalities.

5(e) Explain how in the Bayesian approach how one would determine an upper limit on θ using the result from (d). (You do not actually have to calculate the upper limit.)

Explain briefly the differences in the interpretation between frequentist and Bayesian upper limits.

Exercise 6: A random variable x follows a p.d.f. $f(x;\theta)$ where θ is an unknown parameter. Consider a sample $\vec{x} = (x_1, \ldots, x_n)$ used to construct an estimator $\hat{\theta}(\vec{x})$ for θ (not necessarily the ML estimator). Prove the Rao-Cramér-Frechet (RCF) inequality,

$$V[\hat{\theta}] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]} \ , \tag{1}$$

where $b = E[\hat{\theta}] - \theta$ is the bias of the estimator. This will require several steps:

(a) First, prove the Cauchy–Schwarz inequality, which states that for any two random variables u and v,

$$V[u]V[v] \geq (\mathrm{cov}[u, v])^2, \tag{2}$$

where $V[u]$ and $V[v]$ are the variances and $\mathrm{cov}[u, v]$ the covariance. Use that fact that the variance of $\alpha u + v$ must be greater than or equal to zero for any value of α. Then consider the special case $\alpha = (V[v]/V[u])^{1/2}$.

(b) Use the Cauchy–Schwarz inequality with

$$\begin{aligned} u &= \hat{\theta}, \\ v &= \frac{\partial}{\partial \theta} \log L, \end{aligned} \tag{3}$$

where $L = f_{\mathrm{joint}}(\vec{x}; \theta)$ is the likelihood function, which is also the joint p.d.f. for $\vec{x}$. Write (2) so as to express a lower bound on $V[\hat{\theta}]$. Note that here we are treating the likelihood function as a function of $\vec{x}$, i.e. it is regarded as a random variable.

(c) Assume that differentiation with respect to θ can be brought outside the integral to show that

$$E\left[\frac{\partial}{\partial \theta} \log L\right] = \int \dots \int f_{\mathrm{joint}}(\vec{x}; \theta) \frac{\partial}{\partial \theta} \log f_{\mathrm{joint}}(\vec{x}; \theta)\, dx_1 \dots dx_n = 0. \tag{4}$$

The form of the RCF inequality that we will derive depends on this assumption, which is true in most cases of interest. (It is fulfilled as long as the limits of integration do not depend on θ.) Use (4) with (2) and (3) to show that

$$V[\hat{\theta}] \geq \frac{\left(E\left[\hat{\theta} \frac{\partial \log L}{\partial \theta}\right]\right)^2}{E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right]}. \tag{5}$$

(d) Show that the numerator of (5) can be expressed as

$$E\left[\hat{\theta} \frac{\partial \log L}{\partial \theta}\right] = 1 + \frac{\partial b}{\partial \theta}, \tag{6}$$

and that in a similar way the denominator is

$$E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]. \tag{7}$$

Again assume that the order of differentiation with respect to θ and integration over $\vec{x}$ can be reversed. Prove (1) by putting together the ingredients from (c) and (d).