Chapter 6

The Method of Maximum Likelihood

Exercise 6.1: (a) Find the maximum-likelihood estimators for the mean μ and variance σ^2 of a Gaussian p.d.f. based on a sample of n observations, $x_1, \ldots, x_n$.

(b) Find the expectation values and variances of the estimators by relating $\hat{\mu}$ and $\widehat{\sigma^2}$ to the estimators $\overline{x}$ and s^2 given in SDA Chapter 2.

(c) Find the approximate inverse covariance matrix (valid for large samples) by computing

$$(V^{-1})_{ij} = -E\left[\frac{\partial^2 \log L}{\partial\theta_i\theta_j}\right], \tag{6.1}$$

where θ_i and θ_j $(i, j = 1, 2)$ represent μ and σ^2. Invert V^{-1} to find the covariance matrix, and compare the diagonal elements (i.e. the variances) to the exact values found in (b). Note that the answers from (b) and (c) agree in the large sample limit.

Exercise 6.2: Consider a binomially distributed variable n, the number of successes observed in N trials, where the probability of success in a single trial is p. What is the maximum-likelihood estimator for p given a single observation of n? Show that $\hat{p}$ is unbiased and find its variance. Show that the variance of $\hat{p}$ is equal to the minimum variance bound (see SDA equation (6.16)).

Exercise 6.3: (a) Consider again a binomial variable with probabilities p and $q = 1 - p$ for the outcomes of each trial. Using the estimator for p from Exercise 6.2, construct the ML estimator $\hat{\alpha}$ for the asymmetry

$$\alpha = p - q = 2p - 1, \tag{6.2}$$

and find its standard deviation $\sigma_{\hat{\alpha}}$.

(b) Suppose that one is trying to measure a very small asymmetry, expected to be at the level of $\alpha \approx 10^{-3}$. How many trials is it necessary to observe in order to have the standard deviation $\sigma_{\hat{\alpha}}$ at least a factor of three smaller than this?

Exercise 6.4: Consider a single observation of a Poisson distributed variable n. What is the maximum-likelihood estimator of the mean ν? Show that the estimator is unbiased and find its variance. Show that the variance of $\hat{\nu}$ is equal to the minimum variance bound.

Exercise 6.5: Early evidence supporting the Standard Model of particle physics was provided by the observation of a difference in the cross sections $\sigma_{\rm R}$ and $\sigma_{\rm L}$ for inelastic scattering of right (R) or left (L) hand polarized electrons on a deuterium target. For a given integrated luminosity L (proportional to the electron beam intensity and time of data taking), the numbers of scattering events of each type are Poisson variables, $n_{\rm R}$ and $n_{\rm L}$, with means $\nu_{\rm R}$ and $\nu_{\rm L}$. The means are related to the cross sections by $\nu_{\rm R} = \sigma_{\rm R} L$ and $\nu_{\rm L} = \sigma_{\rm L} L$, and the experiment is set up such that the luminosity L is equal for both cases. Using the result from Exercise 6.4, construct an estimator $\hat{\alpha}$ for the polarization asymmetry,

$$\alpha = \frac{\sigma_{\rm R} - \sigma_{\rm L}}{\sigma_{\rm R} + \sigma_{\rm L}} \, . \tag{6.3}$$

Using error propagation, find the standard deviation $\sigma_{\hat{\alpha}}$ as a function of α and $\nu_{\rm tot} = \nu_{\rm R} + \nu_{\rm L}$. Compare this to the corresponding quantity from Exercise 6.3. The asymmetry was expected to be at the level of 10^{-4}. How many scattering events must be observed so that $\sigma_{\hat{\alpha}}$ is a factor of ten smaller than this? (The number of is so large that the events could not be recorded individually, but rather the output current of the detector was measured. See C.Y. Prescott et al., Parity non-conservation in inelastic electron scattering, Phys. Lett. B77 (1978) 347.)

Exercise 6.6: A random variable x follows a p.d.f. $f(x;\theta)$ where θ is an unknown parameter. Consider a sample $\mathbf{x} = (x_1, \ldots, x_n)$ used to construct an estimator $\hat{\theta}(\mathbf{x})$ for θ (not necessarily the ML estimator). Prove the Rao-Cramér-Frechet (RCF) inequality,

$$V[\hat{\theta}] \, \geq \, \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]} \, , \tag{6.4}$$

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,

$$V[u]V[v] \geq (\mathrm{cov}[u,v])^2, \tag{6.5}$$

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

$$\begin{aligned} u &= \hat{\theta}, \\ v &= \frac{\partial}{\partial \theta} \log L, \end{aligned} \tag{6.6}$$

where $L = f_{\rm joint}(\mathbf{x};\theta)$ is the likelihood function, which is also the joint p.d.f. for $\mathbf{x}$. Write (6.5) so as to express a lower bound on $V[\hat{\theta}]$. Note that here we are treating the likelihood function as a function of $\mathbf{x}$, 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

$$E\left[\frac{\partial}{\partial\theta}\log L\right] \;=\; \int\ldots\int f_{\mathrm{joint}}(\mathbf{x};\theta)\frac{\partial}{\partial\theta}\log f_{\mathrm{joint}}(\mathbf{x};\theta)\,dx_1\ldots dx_n \;=\; 0. \tag{6.7}$$

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 (6.7) with (6.5) and (6.6) to show that

$$V[\hat{\theta}] \,\geq\, \frac{\left(E\left[\hat{\theta}\frac{\partial\log L}{\partial\theta}\right]\right)^2}{E\left[\left(\frac{\partial\log L}{\partial\theta}\right)^2\right]}. \tag{6.8}$$

(d) Show that the numerator of (6.8) can be expressed as

$$E\left[\hat{\theta}\frac{\partial\log L}{\partial\theta}\right] \;=\; 1 \,+\, \frac{\partial b}{\partial\theta}, \tag{6.9}$$

and that in a similar way the denominator is

$$E\left[\left(\frac{\partial\log L}{\partial\theta}\right)^2\right] \;=\; -E\left[\frac{\partial^2\log L}{\partial\theta^2}\right]. \tag{6.10}$$

Again assume that the order of differentiation with respect to θ and integration over $\mathbf{x}$ can be reversed. Prove (6.4) by putting together the ingredients from (c) and (d).

Exercise 6.7: Write a computer program to generate samples of n values $t_1,\ldots,t_n$ according to an exponential distribution

$$f(t;\tau) \;=\; \frac{1}{\tau}\,e^{-t/\tau}\,, \quad t\geq 0\,. \tag{6.11}$$

(a) Show that the ML estimator for τ is given by the sample mean $\hat{\tau} = \frac{1}{n}\sum_{i=1}^n t_i$. Generate 1000 samples with $\tau = 1$ and $n = 10$. Evaluate $\hat{\tau}$ for each sample, and make a histogram of the results. Compare the mean of the $\hat{\tau}$ values with the true value $\tau = 1$.

(b) Suppose the p.d.f. for t had been parametrized in terms of $\lambda = 1/\tau$, i.e.

$$f(t;\lambda) \;=\; \lambda\,e^{-\lambda t}\,, \quad t\geq 0\,. \tag{6.12}$$

Show that the ML estimator for λ is $\hat{\lambda} = 1/\sum_{i=1}^n t_i$. Modify the program in (a) to include a histogram of the estimates $\hat{\lambda}$ from the Monte Carlo experiments. Compare the mean value of $\hat{\lambda}$ to the true value $\lambda = 1$. Determine numerically the bias $b = E[\hat{\lambda}] - \lambda$ for $n = 5$, 10, 100.

Exercise 6.8: The license plates of taxis in Geneva are numbered from one up to the total number N_{taxi}. N observations of taxi licenses are made yielding numbers $n_1,\ldots n_N$.

(a) Construct the maximum-likelihood estimator for the total number of taxis. (This is a well-known example where the ML estimator is biased and not efficient. The difficulty stems from the fact that the range of possible data values depends on the parameter.)

(b) Propose a better estimator for the number of taxis. Give its expectation value and variance.

Exercise 6.9: Consider N independent Poisson variables $n_1, \ldots, n_N$, with mean values $\nu_1, \ldots, \nu_N$. Suppose the mean values are related to a controlled variable x according to relation of the form,

$$\nu(x) = \theta a(x), \tag{6.13}$$

where θ is an unknown parameter and $a(x)$ is an arbitrary known function. The N values of ν_i are thus given by $\nu(x_i) = \theta a(x_i)$, where the values $x_1, \ldots, x_N$ are assumed to be known. Show that the ML estimator for θ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{N} n_i}{\sum_{i=1}^{N} a(x_i)}. \tag{6.14}$$

Show that $\hat{\theta}$ is unbiased and that its variance is given by the minimum variance bound (cf. Exercise 6.6).

Exercise 6.10: An example of the situation described in Exercise 6.7 is provided by (anti)neutrino-nucleon scattering. According to the quark-parton model, the cross sections for the reactions $\nu N \to \mu^- X$ and $\overline{\nu} N \to \mu^+ X$ are given by

$$\begin{aligned} \sigma(\nu N \to \mu^- X) &= \frac{G^2 M E}{\pi} \left(\langle q \rangle + \frac{1}{3} \langle \overline{q} \rangle \right) \equiv \theta_\nu E \\ \sigma(\overline{\nu} N \to \mu^+ X) &= \frac{G^2 M E}{\pi} \left(\frac{1}{3} \langle q \rangle + \langle \overline{q} \rangle \right) \equiv \theta_{\overline{\nu}} E, \end{aligned} \tag{6.15}$$

where E is the energy of the incoming (anti)neutrino, $M = 0.938$ GeV is the mass of the target nucleon and $G = 1.16 \times 10^{-6}$ GeV^{-2} is the Fermi constant. Here the variable x corresponds to the energy E, and the parameters on the right-hand sides of (6.15) correspond to two different parameters, θ_ν and $\theta_{\overline{\nu}}$.

Suppose data are collected N different values of E. At each energy, the expected number of events is given by

$$\nu_i = \sigma(E_i)\, \varepsilon(E_i)\, \mathcal{L}_i, \tag{6.16}$$

where $\sigma(E_i)$ is the anti(neutrino) cross section at energy E_i, $\mathcal{L}_i$ is the integrated luminosity, and $\varepsilon(E_i)$ is the probability for the detector to register the event (the efficiency), which is in general a function of the energy. For purposes of this exercises, we will assume that the energies E_i and corresponding integrated luminosities $\mathcal{L}_i$ and efficiencies $\varepsilon_i \equiv \varepsilon(E_i)$ are known without error. (Assume in addition that there are no background events.)

Determine the ML estimators for the parameters θ_ν and $\theta_{\overline{\nu}}$, and from them find estimators for $\langle q \rangle$ and $\langle \overline{q} \rangle$. In the context of the quark-parton model, these correspond to the fraction of the nucleon's momentum carried by quarks and antiquarks, respectively. Determine the fraction of the momentum carried by particles other than quarks and antiquarks (i.e. gluons), $\langle g \rangle = 1 - \langle q \rangle - \langle \overline{q} \rangle$.

Exercise 6.11: One of the earliest determinations of Avogadro's number was based on Brownian motion. The experimental set-up shown in Fig. 6.1 was used by Jean Perrin [1] to observe particles of mastic (a substance used in varnish) suspended in water.

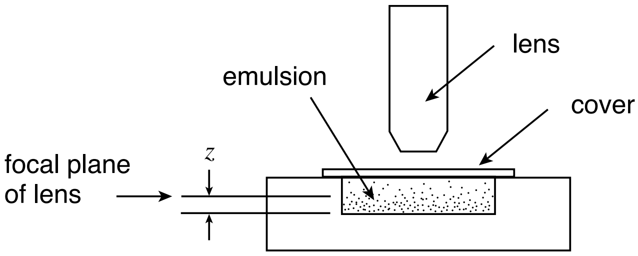

Figure 6.1: Experimental set-up of Jean Perrin for observing the number of particles suspended in water as a function of height.

The particles were spheres of radius $r = 0.52$ μm and had a density of 1.063 g/cm^3, i.e. 0.063 g/cm^3 greater than that of water. By viewing the particles through the microscope, only those in a layer approximately 1 μm thick were in focus; particles 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, (the lowest height is arbitrarily assigned a value $z = 0$) and the number of particles $n(z)$ counted. The data are shown in Table 6.1.

Table 6.1: Perrin's data on the number of mastic particles observed at different heights z in an emulsion.

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

The gravitational potential energy of a spherical particle of mastic in water is given by

$$E = \frac{4}{3}\pi r^3 \, \Delta\rho \, g z, \tag{6.17}$$

where $\Delta\rho = \rho_{\mathrm{mastic}} - \rho_{\mathrm{water}} = 0.063$ g/cm^3 is the difference in densities and $g = 980$ cm/s^2 is the acceleration of gravity. Statistical mechanics predicts that the probability for a particle to be in a state of energy E is proportional to

$$P(E) \propto e^{-E/kT}, \tag{6.18}$$

where k is Boltzmann's constant and T the absolute temperature. The particles should therefore be distributed in height according to an exponential law, where the number n observed at z can be treated as a Poisson variable with a mean $\nu(z)$. By combining (6.17) and (6.18), this is found to be

[1] Jean Perrin, Mouvement brownien et réalité moléculaire, *Ann. Chimie et Physique, 8e série*, **18** (1909) 1-114; *Les Atomes*, Flammarion, Paris, 1991 (first edition, 1913); *Brownian Movement and Molecular Reality*, in Mary-Jo Nye, ed., *The Question of the Atom*, Tomash, Los Angeles, 1984.

$$\nu(z) = \nu_0 \exp\left(-\frac{4\pi r^3\,\Delta\rho\,gz}{3kT}\right), \tag{6.19}$$

where ν_0 is the expected number of particles at $z = 0$.

(a) Write a computer program to determine the parameters k and ν_0 with the method of maximum likelihood. Use the data given in Table 6.1 to construct the log-likelihood function based on Poisson probabilities (cf. SDA Section 6.10),

$$\log L(\nu_0, k) = \sum_{i=1}^{N} (n_i \log \nu_i - \nu_i), \tag{6.20}$$

where $N = 4$ is the number of measurements. For the temperature use $T = 293$ K.

(b) From the value you obtain for k, determine Avogadro's number using the relation

$$N_{\mathrm{A}} = R/k, \tag{6.21}$$

where R is the gas constant. The value used by Perrin was $R = 8.32 \times 10^7$ erg/mol K.

(c) Instead of maximizing the log-likelihood function (6.20), estimate ν_0 and k by minimizing

$$\chi^2_{\mathrm{P}}(\nu_0, k) \;=\; 2 \sum_{i=1}^{N} \left(n_i \,\log \frac{n_i}{\nu_i} \;+\; \nu_i \;-\; n_i \right) \;, \tag{6.22}$$

where $\nu_i = \nu(z_i)$ depends on ν_0 and k through equation (6.19). Use the value of χ^2_{P} to evaluate the goodness-of-fit (cf. SDA Section 6.11). Comment on possible systematic errors in Perrin's determination of N_{A}.