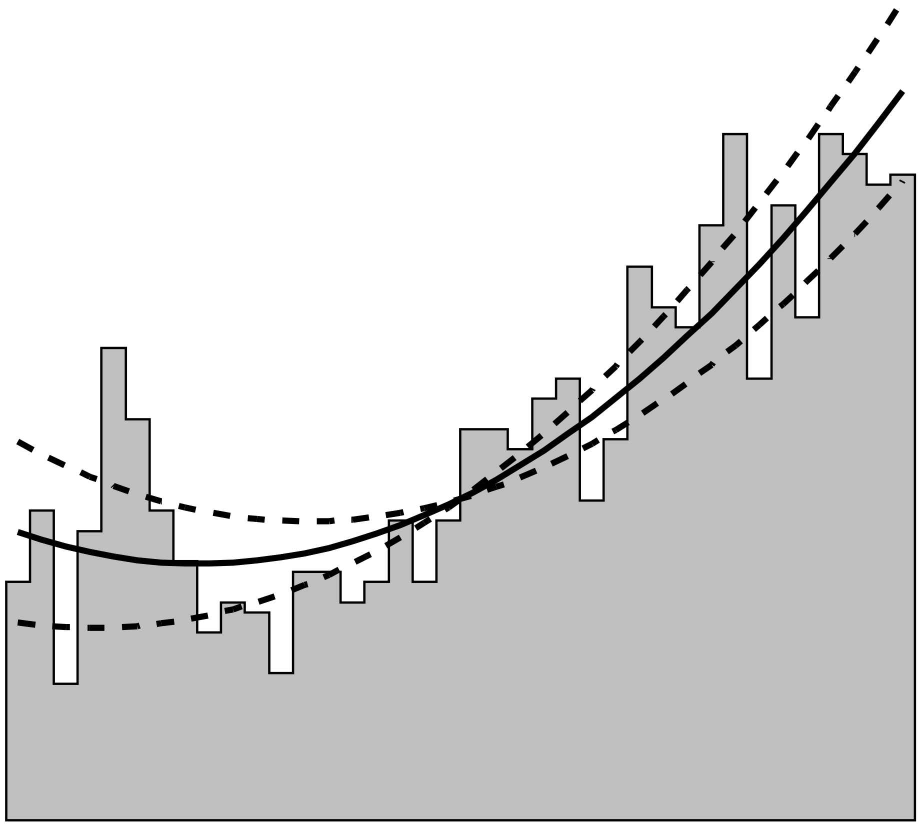
Exercises in Statistical Data Analysis

Glen D. Cowan
Physics Department
Royal Holloway, University of London



Preface

These exercises accompany G. Cowan, *Statistical Data Analysis*, Clarendon Press, Oxford, 1998, referred to in the following as SDA. The exercises and related information can be obtained from the SDA World Wide Web site, which is currently located at

`http://www.pp.rhul.ac.uk/~cowan/sda`

In case of an address change, the site can be found by following the catalogue link of Oxford University Press at

`http://www.oup.co.uk/`

The exercises, like SDA, are targeted primarily at physics students. Many of the problems are more general, however, and even those that relate to physics are formulated so as to be meaningful for most science students. The level of the problems is appropriate for advanced undergraduate or beginning graduate students, for example as part of a course in experimental methods or data analysis techniques.

A number of the exercises use short computer programs and data files which can be obtained from the SDA web site. Some of these require in addition routines from the CERN Program Library. The exercises and accompanying software are in a state of developement, and changes are to be expected. It would be greatly appreciated if corrections, suggestions and comments could be communicated to `glen.cowan@cern.ch`.

Geneva
January 1998 G.D.C.

Chapter 1

Fundamental Concepts

Exercise 1.1: Consider a sample space S and assume for a given subset B that $P(B) > 0$. Show that the conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.1}$$

satisfies the axioms of probability.

Exercise 1.2: Show that

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

(Express $A \cup B$ as the union of three disjoint sets.)

Exercise 1.3: 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 a particle detected in one layer only to be a photon?

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

Exercise 1.4: Suppose a random variable x has the p.d.f. $f(x)$. Show that the p.d.f. for $y = x^2$ is

$$g(y) \;=\; \frac{1}{2\sqrt{y}} f(\sqrt{y}) \;+\; \frac{1}{2\sqrt{y}} f(-\sqrt{y})\,. \tag{1.2}$$

Exercise 1.5: Suppose two independent random variables x and y are both uniformly distributed between zero and one, i.e. the p.d.f. $g(x)$ is given by

$$g(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise}\,, \end{cases} \tag{1.3}$$

and similarly for the p.d.f. $h(y)$.

(a) Using SDA equation (1.35), show that the p.d.f. $f(z)$ for $z = xy$ is

$$f(z) = \begin{cases} -\log z & 0 < z < 1 \\ 0 & \text{otherwise}. \end{cases} \tag{1.4}$$

(b) Find the same result using SDA equations (1.37) and (1.38) by defining an additional function, $u = x$. First, find the joint p.d.f. of z and u. Integrate this over u to find the p.d.f. for z.

(c) Show that the cumulative distribution of z is

$$F(z) = z(1 - \log z)\,. \tag{1.5}$$

Exercise 1.6: Consider a random variable x and constants α and β. Show that

$$\begin{aligned} E[\alpha x + \beta] &= \alpha E[x] + \beta\,, \\ V[\alpha x + \beta] &= \alpha^2 V[x]\,. \end{aligned} \tag{1.6}$$

Exercise 1.7: Consider two random variables x and y.

(a) Show that the variance of $\alpha x + y$ is given by

$$\begin{aligned} V[\alpha x + y] &= \alpha^2 V[x] + V[y] + 2\alpha \mathrm{cov}[x, y] \\ &= \alpha^2 V[x] + V[y] + 2\alpha\rho\sigma_x\sigma_y\,, \end{aligned} \tag{1.7}$$

where α is any constant value, $\sigma_x^2 = V[x]$, $\sigma_y^2 = V[y]$, and the correlation coefficient is $\rho = \mathrm{cov}[x, y]/\sigma_x\sigma_y$.

(b) Using the result of (a), show that the correlation coefficient always lies in the range $-1 \leq \rho \leq 1$. (Use the fact that the variance $V[\alpha x + y]$ is always greater than or equal to zero and consider the cases $\alpha = \pm\sigma_y/\sigma_x$.)

Exercise 1.8: Suppose $\mathbf{x} = (x_1, \ldots, x_n)$ is described by the joint p.d.f. $f(\mathbf{x})$, and the variables $\mathbf{y} = (y_1, \ldots, y_n)$ are defined by means of a linear transformation,

$$y_i = \sum_{j=1}^{n} A_{ij} x_j. \tag{1.8}$$

Assume that the inverse transformation $\mathbf{x} = A^{-1}\mathbf{y}$ exists.

(a) Show that the joint p.d.f. for $\mathbf{y}$ is given by

$$g(\mathbf{y}) = f(A^{-1}\mathbf{y})\,|\det(A^{-1})|. \tag{1.9}$$

(b) Find $g(\mathbf{y})$ for the case where A is orthogonal, i.e. $A^{-1} = A^T$.

Chapter 2

Examples of Probability Functions

Exercise 2.1: Consider N multinomially distributed random variables $\mathbf{n} = (n_1, \ldots, n_N)$ with probabilities $\mathbf{p} = (p_1, \ldots, p_N)$ and a total number of trials $n_{\rm tot} = \sum_{i=1}^{N} n_i$. Suppose the variable k is defined as the sum of the first M of the n_i,

$$k = \sum_{i=1}^{M} n_i, \quad M \leq N. \tag{2.1}$$

Use error propagation and the multinomial covariance,

$$\mathrm{cov}[n_i, n_j] = \delta_{ij} n_{\rm tot} p_i (1 - p_i) \; + \; (\delta_{ij} - 1) p_i p_j n_{tot}, \tag{2.2}$$

to find the variance of k. Show that this is equal to the variance of a binomial variable with $p = \sum_{i=1}^{M} p_i$ and $n_{\rm tot}$ trials.

Exercise 2.2: Suppose the random variable x is uniformly distributed in the interval $[\alpha, \beta]$, with $\alpha, \beta > 0$. Find the expectation value of $1/x$, and compare the answer to $1/E[x]$ using $\alpha = 1$, $\beta = 2$.

Exercise 2.3: Consider the exponential p.d.f.,

$$f(x; \xi) = \frac{1}{\xi} \, e^{-x/\xi} \,, \quad x \geq 0 \,. \tag{2.3}$$

(a) Show that the corresponding cumulative distribution is given by

$$F(x) = 1 - e^{-x/\xi} \,, \quad x \geq 0 \,. \tag{2.4}$$

(b) Show that the conditional probability to find a value x between x_0 and $x_0 + x'$ given that $x > x_0$ is equal to the (unconditional) probability to find x less than x', i.e.

$$P(x \leq x_0 + x' | x \geq x_0) = P(x \leq x') \,. \tag{2.5}$$

(c) Cosmic ray muons produced in the upper atmosphere enter a detector at sea level, and some of them come to rest in the detector and decay. The time difference t between entry into the detector and decay follows an exponential distribution, and the mean value of t is the mean lifetime of the muon (approximately 2.2 μS). Explain why the time that the muon lived prior to entering the detector does not play a role in determining the mean lifetime.

Exercise 2.4: Suppose y follows a Gaussian distribution with mean μ and variance σ^2.

(a) Show that

$$x = \frac{y-\mu}{\sigma} \tag{2.6}$$

follows the standard Gaussian $\varphi(x)$ (i.e. having a mean of zero and unit variance).

(b) Show that the cumulative distributions $F(y)$ and $\Phi(x)$ are equal, i.e.

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right). \tag{2.7}$$

Exercise 2.5: (a) Show that if y is Gaussian distributed with mean μ and variance σ^2, then $x = e^y$ follows the log-normal p.d.f.,

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\frac{1}{x}\,\exp\left(\frac{-(\log x-\mu)^2}{2\sigma^2}\right). \tag{2.8}$$

(b) Find the expectation value and variance of x by explicitly computing the integrals

$$\begin{aligned} E[x] &= \int x\, f(x;\mu,\sigma^2)\,dx\,, \\ V[x] &= \int (x-E[x])^2\, f(x;\mu,\sigma^2)\,dx\,. \end{aligned} \tag{2.9}$$

(c) Compare the variance from (b) to the approximate result obtained by error propagation with $V[y] = \sigma^2$. Under what conditions is the approximation valid? (Recall that y and hence also σ^2 are dimensionless.)

Exercise 2.6: Show that the cumulative χ^2 distribution for n degrees of freedom can be expressed as

$$F_{\chi^2}(x;n) = P\left(\frac{x}{2},\frac{n}{2}\right), \tag{2.10}$$

where P is the incomplete gamma function,

$$P(n,x) = \frac{1}{\Gamma(n)}\int_0^x e^{-t}\,t^{n-1}\,dt\,. \tag{2.11}$$

Chapter 3

The Monte Carlo Method

For these exercises you will need a random number generator to produce random values uniformly distributed between zero and one. A simple FORTRAN example is given in the file `random.f`. This routine is mainly for pedagogical purposes and simple applications. More sophisticated routines such as `RANMAR` or `RANLUX` can be found in the CERN Program Library.

Exercise 3.1: (a) Using `random.f` or another random number generator, write a short program to generate 10000 random values uniformly distributed between zero and one, and display the result as a histogram with 100 bins.

Exercise 3.2: Modify the histogram from Exercise 3.1 to have only 5 bins and $N = 100$ entries. The generated histogram can be regarded as an observation of a multinomially distributed vector $(n_1, \ldots, n_5)$, with parameters $N = 100$ and $p_i = 0.2$ for $i = 1, \ldots, 5$.

(a) By placing the code to generate the histogram in a loop, modify the program to repeat the Monte Carlo experiment 100 times, each time with a different seed value. (As long as the program is not terminated after each experiment, a new seed will be used automatically.) Produce a histogram of the value of any bin n_i (e.g. for $i = 3$) after each experiment. This should follow a binomial distribution with mean $Np_i = 20$ and standard deviation $\sqrt{Np_i(1-p_i)} = 4$.

(b) Produce a scatter plot (two-dimensional histogram) for the values of any two bins n_i and n_j. These should have a covariance $\mathrm{cov}[n_i, n_j] = -Np_i p_j = -4$ or a correlation coefficient $\rho = -4/4^2 = -0.25$.

If you are using the HBOOK histogram package from the CERN Program Library, you can use the routine `HUNPAK` to unpack the values $(n_1, \ldots, n_5)$ after each experiment.

Exercise 3.3: Consider a sawtooth p.d.f.,

$$f(x) = \begin{cases} \frac{2x}{x_{\max}^2} & 0 < x < x_{\max} \, , \\ 0 & \text{otherwise} \, . \end{cases} \tag{3.1}$$

(a) Use the transformation method to find the function $x(r)$ to generate random numbers according $f(x)$, cf. Section 3.2. Implement the method in a short computer program and make a histogram of the results. (Use e.g. $x_{\max} = 1$.)

(b) Write a program to generate random numbers according to the sawtooth p.d.f. using the acceptance-rejection technique, cf. Section 3.3. Plot a histogram of the results.

Exercise 3.4: The purpose of this exercise is to generate random numbers according to a Gaussian p.d.f. A number of algorithms exist for this purpose, implemented, for example, in the routine `RNORMX` from the CERN library. A simple algorithm suitable for pedagogical purposes is based on the central limit theorem: a sum of random variables becomes Gaussian in the limit that the number of terms in the sum is large, as long as none of the terms make up a significant fraction of the sum (cf. SDA Chapter 10).

(a) Suppose x is uniformly distributed in $[0, 1]$ and consider the sum of n independent x values,

$$y = \sum_{i=1}^{n} x_i \, . \tag{3.2}$$

Show that the expectation value of y is $n/2$ and the variance is $n/12$. Show that, as a consequence, the variable

$$z = \frac{\sum_{i=1}^{n} x_i - \frac{n}{2}}{\sqrt{n/12}} \, . \tag{3.3}$$

has a mean of zero and unit standard deviation.

(b) Write a computer program to generate values z as defined in (a) for arbitrary values of n. Make histograms of 10000 values of z for $n = 0, \ldots, 20$. At what point does the distribution appear approximately Gaussian? A convenient choice for a simple Gaussian generator is $n = 12$. Comment on the limitations of such an algorithm. Optional: Derive the explicit form of the p.d.f. of z for $n = 2$.

Exercise 3.5: The variable t follows an exponential distribution with mean $\tau = 1$ and x is Gaussian distributed with mean $\mu = 0$ and standard deviation $\sigma = 0.5$. Write a Monte Carlo program to generate values of

$$y = t + x \, . \tag{3.4}$$

Here the value t could represent the true decay time of an unstable particle, and the value x the measurement error, so that y represents the measured decay time. Make a histogram of the values. Modify the program to investigate the cases $\tau \ll \sigma$ and $\tau \gg \sigma$.

3.6: Consider a random variable x distributed according to the Cauchy (Breit-Wigner) p.d.f.,

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \, . \tag{3.5}$$

(a) Show that if r is uniformly distributed in $[0, 1]$, then

$$x(r) = \tan[\pi(r - \tfrac{1}{2})] \tag{3.6}$$

follows the Cauchy p.d.f.

(b) Using the result from (a), write a computer program to generate Cauchy distributed random numbers. Generate 10000 values and display the result as a histogram.

(c) Modify the program in (b) to generate repeated experiments each consisting of n independent Cauchy distributed values for e.g. $n = 10$. For each sample, compute the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Compare a histogram of $\overline{x}$ with the original histogram of x. (See also Exercise 10.6.)

Exercise 3.7: A photomultiplier is a device capable of detecting individual photons as illustrated in Fig. 3.1.[1] A photon strikes the photocathode, where there is a certain probability for it to eject an electron (called a photoelectron). The photoelectron is accelerated in an electric field towards an electrode (called a dynode). In the collision with the first dynode, the photoelectron can liberate further electrons. These are accelerated towards the second dynode, where more electrons are produced. This continues through a series of stages until the electrons produced at the final dynode are collected.

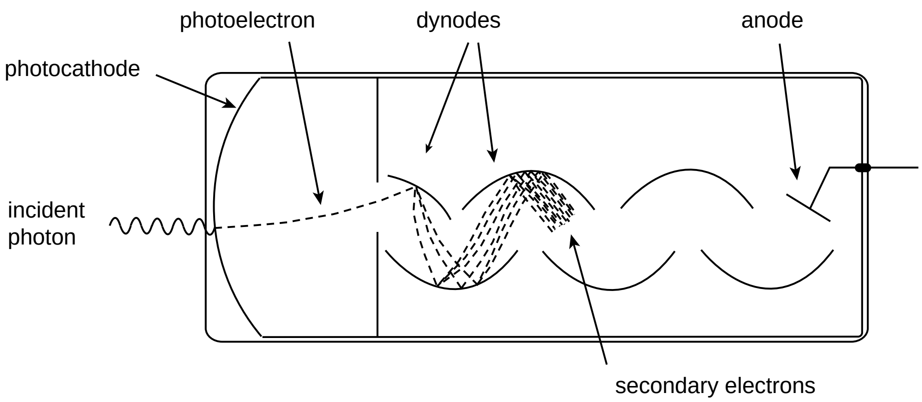


Figure 3.1: Schematic drawing of a photomultiplier tube.

The number of electrons produced at the ith dynode for each incoming electron can be modeled as a Poisson variable n_i with mean value ν_i, which in general can be different for each stage. Suppose the photomultiplier has N dynodes. The number of electrons $n_{\rm out}$ produced at the final stage for a single incident photoelectron has an expectation value (the *gain* of the photomultiplier),

$$\nu_{\rm out} = E[n_{\rm out}] = \prod_{i=1}^{N} \nu_i \tag{3.7}$$

(a) Write a Monte Carlo program to determine the distribution of the number of electrons $n_{\rm out}$ at the end of $N = 6$ dynodes produced by a single initial photoelectron, with $\nu = 3.0$ for each dynode. (Poisson random numbers can be generated with the routine `RNPSSN` from the CERN program library. A partial solution is given in the program `pmt.f`.) Run the program to simulate $M = 1000$ initial photoelectrons and make a histogram of $n_{\rm out}$. Estimate the mean $\nu_{\rm out}$ and variance $V[n_{\rm out}] = \sigma^2_{\rm out}$ by calculating the sample mean,

$$\overline{n}_{\rm out} = \frac{1}{M}\sum_{i=1}^{M} n_{{\rm out},i} \tag{3.8}$$

[1] For a more detailed description see e.g. C. Grupen, A. Böhrer and L. Smolik, *Particle Detectors*, Cambridge University Press, Cambridge, 1996.

and sample variance

$$s_{\rm out}^2 = \frac{1}{M-1} \sum_{i=1}^{M} (n_{{\rm out},i} - \overline{n}_{\rm out})^2 \,. \tag{3.9}$$

(The sample mean and variance are described further in SDA Chapter 5.) Compare the sample mean to the value from equation (3.7). Compare the sample variance (or standard deviation) to the value that one would obtain from a Poisson variable of mean $\nu_{\rm out}$. Explain qualitatively why the standard deviation of $n_{\rm out}$ is much larger than in the Poisson case.

(b) One would like the standard deviation of $n_{\rm out}$ to be small in order to be able to determine as accurately as possible the number of initial photoelectrons (and thus estimate the number of incident photons). In some applications one would like to have the standard deviation small enough to distinguish between 1 and 2 photoelectrons; hence one tries to have a relative resolution, i.e. the ratio of standard deviation to mean, less than unity. One way of achieving this is to increase the mean number of electrons produced at the first dynode. This can be done by increasing the voltage so that the photoelectrons collide with a higher energy, and also by using a metal with a low work function, i.e. a high probability for secondary electron emission.

Modify the program from (a) so that the mean for the first dynode is larger, e.g. $\nu_1 = 6$. Run the program and estimate the ratio of standard deviation to mean of $n_{\rm out}$. Explain qualitatively why this gives a better resolution than in the case with all ν_i equal. Why does it not help much to increase the gain of the dynodes in the later stages of the photomultiplier?

(c) Try to extend the program to simulate $N = 12$ dynodes. You will quickly find out that it requires too much computing time to simulate the collision of each electron with each dynode. Instead, run the program for $N = 6$ with enough events to obtain a good estimate of the distribution of $n_{\rm out}$ (e.g. at least $M \approx 10^4$ events in a histogram with 50 bins from $0 \leq n_{\rm out} \leq 5000$). Generate numbers which follow this distribution using any method, e.g. acceptance-rejection. For each electron obtained after the first six dynodes, generate in a similar way the number of electrons that it produces in the next six. For the first six stages, use a distribution of $n_{\rm out}$ based on $\nu_1 = 6$ and the rest of the $\nu_i = 3$; for the last six, take all $\nu_i = 3$.

Chapter 4

Statistical Tests

Exercise 4.1: Charged particles traversing a gas volume produce ionization, the mean amount of which depends on the type of particle in question. Suppose a test statistic t based on ionization measurements has been constructed such that it follows a Gaussian distribution centered about 0 for electrons and about 2 for pions, with a standard deviation equal to unity for both hypotheses. A test is constructed to select electrons by requiring $t < 1$.

(a) What is the significance level of the test (i.e. the probability to accept an electron).

(b) What is the power of the test against the hypothesis that the particle is a pion. What is the probability that a pion will be accepted as an electron?

(c) Suppose a sample of particles is known to consist of 99% pions and 1% electrons. What is the purity of the electron sample selected by $t < 1$?

(d) Suppose one requires a sample of electrons with a purity of at least 95%. What should the critical region (i.e. the cut value) of the test be? What is the efficiency for accepting electrons with this cut value? Equivalently, what is the significance level of the test?

Exercise 4.2: Consider a test statistic t based on a linear combination of input variables $\mathbf{x} = (x_1, \ldots, x_n)$ with coefficients $\mathbf{a} = (a_1, \ldots, a_n)$,

$$t(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i \; = \; \mathbf{a}^T \mathbf{x} \, . \tag{4.1}$$

Suppose that under two hypotheses H_0 and H_1, the mean values of $\mathbf{x}$ are given by $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, the covariance matrices are V_0 and V_1, the means of the statistic t are τ_0 and τ_1, and the variances of t are Σ_0^2 and Σ_1^2 (see SDA Section 4.4.1).

(a) Show that the values of the coefficients $\mathbf{a}$ that maximize the separation

$$J(\mathbf{a}) = \frac{(\tau_0 - \tau_1)^2}{\Sigma_0^2 + \Sigma_1^2} \tag{4.2}$$

are given by

$$\mathbf{a} \propto W^{-1} \left(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 \right) , \tag{4.3}$$

where $W = V_0 + V_1$. This defines Fisher's linear discriminant function.

(b) Suppose that $V_0 = V_1 = V$ and the p.d.f.s for the input variables $f(\mathbf{x}|H_0)$ and $f(\mathbf{x}|H_1)$ are multidimensional Gaussians centered about $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ (cf. SDA equation (4.26)). Take the prior probabilities of the two hypotheses to be π_0 and π_1. Using Bayes' theorem, find the posterior probabilities $P(H_0|\mathbf{x})$ and $P(H_1|\mathbf{x})$ as a function of t.

(c) Show that by generalizing the test statistic to include an offset,

$$t(\mathbf{x}) = a_0 + \sum_{i=1}^{n} a_i x_i \,, \tag{4.4}$$

the posterior probability $P(H_0|\mathbf{x})$ can be expressed as

$$P(H_0|\mathbf{x}) = \frac{1}{1 + e^{-t}} \,, \tag{4.5}$$

where the offset a_0 is given by

$$a_0 = -\tfrac{1}{2}\boldsymbol{\mu}_0^T V^{-1} \boldsymbol{\mu}_0 + \tfrac{1}{2}\boldsymbol{\mu}_1^T V^{-1} \boldsymbol{\mu}_1 + \log \frac{\pi_0}{\pi_1} \,. \tag{4.6}$$

Exercise 4.3: The number of events having particular kinematic properties observed in electron-positron collisions can be treated as a Poisson variable. Suppose that for a certain integrated luminosity (i.e. time of data taking at a given beam intensity), 3.9 events are expected from known processes and 16 are observed. Compute the P-value for the hypothesis that no new process is contributing to the number of events. To sum Poisson probabilities, you can use the relation

$$\sum_{n=0}^{m} P(n;\nu) = 1 - F_{\chi^2}(2\nu; n_{\rm dof}) \,, \tag{4.7}$$

where $P(n:\nu)$ is the Poisson probability for n given a mean value ν, and F_{χ^2} is the cumulative χ^2 distribution for $n_{\rm dof} = 2(m+1)$ degrees of freedom. This can be computed using the routine `PROB` from the CERN Program Library or looked up in standard tables.

Exercise 4.4: The file `data_1.dat` contains a histogram with data; the first two columns are the bin boundaries, and the third column gives the numbers of entries n_i, $i = 1, \ldots, 20$, which we will treat as Poisson random variables. The files `theory_1.dat` and `theory_2.dat` give two predictions for the expectation values $\nu_i = E[n_i]$, and are shown with the data in Fig. 4.1.

(a) Write a computer program to read in the files and to determine the χ^2 statistic according to SDA equation (4.39) for each of the two theories. (A solution is given in `compute_chi2.f`.)

(b) Because many of the bins contain few or no entries, one does not expect the statistic above to follow the χ^2 distribution. Write a computer program to determine the true distribution assuming the two hypotheses `theory_2.dat` and `theory_2.dat`. What are the P-values for the two theories when the test statistic is computed with the data set from (a)? What would the P-values be if the one were to assume the usual χ^2 distribution? (A partial solution is given in `compute_chi2_dist.f`.)

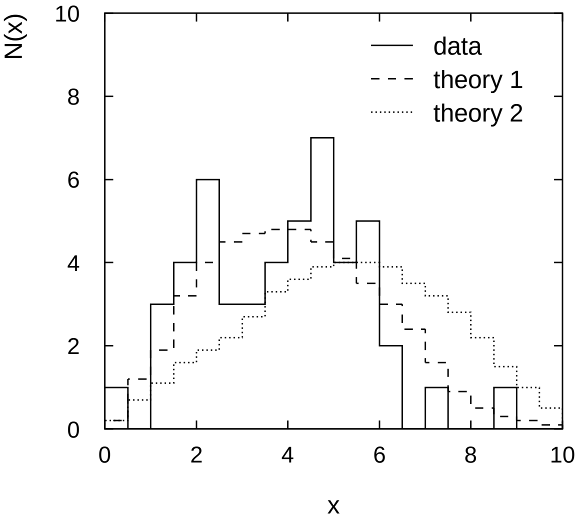


Figure 4.1: Data from the file `data_1.dat` and hypotheses from `theory_1.dat` and `theory_2.dat`.

Exercise 4.5: In an experiment on radioactivity, Rutherford and Geiger counted the number of alpha decays occurring in fixed time intervals.[1] The data are shown in Table 4.1. Assuming that the source consists of a large number of radioactive atoms and that the probability for any one of them to emit an alpha particle in a short interval is small, one would expect the number of decays m in a time interval Δt to follow a Poisson distribution. Deviations from this hypothesis would indicate that the decays were not independent. One could imagine, for example, that the emission of an alpha particle might cause neighboring atoms to decay, resulting in a clustering of decays in short time periods.

Table 4.1: Data by Rutherford and Geiger on the number of times n_m that m alpha decays were observed in a time interval of $\Delta t = 7.5$ seconds.

m	n_m	m	n_m
0	57	8	45
1	203	9	27
2	383	10	10
3	525	11	4
4	532	12	0
5	408	13	1
6	273	14	1
7	139	> 14	0

(a) Using the data in Table 4.1, find the sample mean

$$\overline{m} = \frac{1}{n_{\rm tot}} \sum_m n_m m \,, \tag{4.8}$$

[1]E. Rutherford and H. Geiger, The probability variations in the distribution of α particles, *Philosophical Magazine*, ser. 6, xx (1910) 698–707.

and the sample variance,

$$s^2 = \frac{1}{n_{\rm tot} - 1} \sum_m n_m (m - \overline{m})^2 \,, \tag{4.9}$$

where is n_m the number of occurrences of m decays and $n_{\rm tot} = \sum_m n_m = 2608$ is the total number of time intervals. The sum extends from $m = 0$ up to the maximum number of decays observed in an interval (here $m = 14$). From $\overline{m}$ and s^2, find the *index of dispersion*,

$$t = \frac{s^2}{\overline{m}} \,. \tag{4.10}$$

Since $\overline{m}$ and s^2 are estimators of the mean and variance of m (cf. SDA Chapter 5), and since these are equal if m is a Poisson variable, one would expect to find t around 1. One can show that for Poisson distributed m and large $n_{\rm tot}$, $(n_{\rm tot} - 1)t$ follows a χ^2 distribution for $n_{\rm tot} - 1$ degrees of freedom. Furthermore, for large $n_{\rm tot}$ this becomes a Gaussian distribution with mean $n_{\rm tot} - 1$ and variance $2(n_{\rm tot} - 1)$.

(b) What is the P-value for the hypothesis that m follows a Poisson distribution? What set of t values should be chosen as representing equal or less agreement with the Poisson hypothesis than the observed value of t?

(c) Write a Monte Carlo program to generate a large number of data sets each consisting of $n_{\rm tot} = 2608$ values of m according to a Poisson distribution. (Poisson random numbers can be generated with the routine `RNPSSN` from the CERN library.) For the mean value of m, take $\overline{m}$ obtained from the data in Table 4.1. For each data set, determine t and enter its value in a histogram. From the histogram and the value of t obtained from Rutherford's data, Determine the P-value for the Poisson hypothesis. Compare the result to that obtained in (a). (Optional: Record $(n_{\rm tot} - 1)t$ in a histogram and compare the result with the Gaussian distribution with mean $n_{\rm tot} - 1$ and variance $2(n_{\rm tot} - 1)$.)

Chapter 5

General Concepts of Parameter Estimation

Exercise 5.1: Consider a random variable x with expectation value μ and variance σ^2, and suppose we have a sample of n observations, $x_1, \ldots, x_n$. The purpose of this exercise is to show that the sample mean,

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{5.1}$$

is a consistent estimator for the expectation value μ.

(a) The first step is to prove the Chebyshev inequality,

$$P(|x - \mu| \geq a) \; \leq \; \frac{\sigma^2}{a^2}, \tag{5.2}$$

which holds for any positive a as long as the variance of x exists. Do this by recalling the definition of the variance,

$$\sigma^2 \; = \; \int_{-\infty}^{\infty} (x - \mu)^2 \, f(x) \, dx, \tag{5.3}$$

where $f(x)$ is the p.d.f. of x. Use the fact that the integral (5.3) would be less if the region of integration were restricted to $|x - \mu| \geq a$, and would be even less if in that region, $(x - \mu)^2$ were to be replaced by a^2.

(b) Use the Chebyshev inequality to prove the weak law of large numbers, i.e. for any $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} x_i - \mu \right| \geq \epsilon \right) = 0. \tag{5.4}$$

This is equivalent to the statement that $\overline{x}$ is a consistent estimator for μ, and holds as long as the variance of x exists.

Exercise 5.2: Consider a random variable x of mean μ and variance σ^2, for which one has obtained sample of values $x_1, \ldots, x_n$.

(a) Suppose the mean μ has been estimated using the sample mean, $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. Show that the sample variance,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{n}{n-1} \left(\overline{x^2} - \overline{x}^2 \right), \tag{5.5}$$

is an unbiased estimator of the variance σ^2. (Use the fact that $E[x_i x_j] = \mu^2$ for $i \neq j$ and $E[x_i^2] = \mu^2 + \sigma^2$ for all i.)

(b) Suppose that the mean μ is known. Show that

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \overline{x^2} - \mu^2 \tag{5.6}$$

is an unbiased estimator for σ^2.

Exercise 5.3: (a) Show that the variance of s^2 (5.5) is

$$V[s^2] = E[s^4] - (E[s^2])^2 = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \mu_2^2 \right), \tag{5.7}$$

where $\mu_k = E[(x - \mu)^k]$ is the kth central moment of x. To do this, first show that s^2 can be written as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)} \sum_{i,j=1}^{n} x_i x_j. \tag{5.8}$$

Then show that the expectation value of s^4 is

$$E[s^4] = \frac{1}{(n-1)^2} \sum_{i,j=1}^{n} E[x_i^2 x_j^2] - \frac{2}{n(n-1)^2} \sum_{i,j,k=1}^{n} E[x_i x_j x_k^2] + \frac{1}{n^2(n-1)^2} \sum_{i,j,k,l=1}^{n} E[x_i x_j x_k x_l]. \tag{5.9}$$

Count how many terms in each sum give the algebraic moments μ'_4 or ${\mu'_2}^2$. Note that the rest of the terms all contain at least one power of μ. Express the result in terms of central moments μ_2 and μ_4 by setting the terms with μ equal to zero. Subtract the value of $(E[s^2])^2$ from Exercise (5.2) to obtain the final result.

(b) Find the variance of s^2 for the case where x follows a Gaussian distribution. Use the fact that the fourth central moment of a Gaussian is $\mu_4 = 3\sigma^4$.

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

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

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 $t_1,\dots,t_n$ 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 $N_{\rm taxi}$. N observations of taxi licenses are made yielding numbers $n_1,\dots n_N$.

(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

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

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.

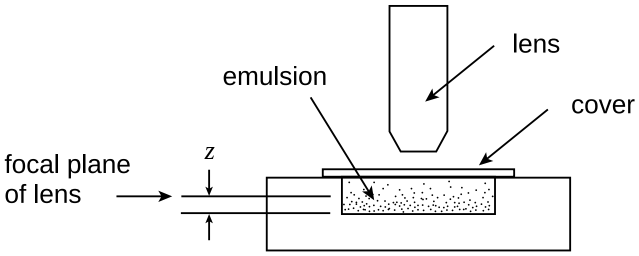


Figure 6.1: Experimental set-up of Jean Perrin for observing the number of particles suspended in water as a function of height.

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 $\Delta\rho = \rho_{\text{mastic}} - \rho_{\text{water}} = 0.063$ 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, 8^e 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

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

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

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

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

Chapter 7

The Method of Least Squares

Exercise 7.1: Galileo's studies of motion included experiments with a ball and an inclined ramp. The ball's trajectory is made horizontal before it falls over the edge, as shown in Fig. 7.1. The horizontal distance d from the edge to the point of impact is measured for different values of the initial height of the ball h. Five data points obtained by Galileo in 1608 are shown in Table 7.1.[1]

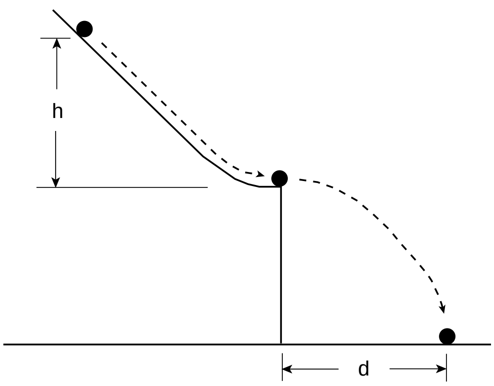


Figure 7.1: The configuration of the ball and ramp experiment performed by Galileo.

Table 7.1: Galileo's data on horizontal distance before impact d for five values of the starting height h. The units are punti (points); one punto is slightly less than 1 mm.

h	d
1000	1500
828	1340
800	1328
600	1172
300	800

Assume the heights h are known with negligible error, and that the horizontal distances d can be regarded as independent Gaussian random variables with standard deviations of $\sigma = 15$ punti.

[1]See Stillman Drake and James Maclachlan, Galileo's discovery of the parabolic trajectory, *Scientific American* **232** (March 1975) 102; Stillman Drake, *Galileo at Work*, University of Chicago Press, Chicago (1978).

(It is not actually known how Galileo estimated the measurement uncertainties, but 1–2% is plausible.) In addition, we know that if $h = 0$, then the horizontal distance d will be zero, i.e. if the ball is started at the very edge of the ramp, it will fall straight down to the floor.

(a) Consider relations between h and d of the form

$$d = \alpha h \tag{7.1}$$

and

$$d = \alpha h + \beta h^2. \tag{7.2}$$

Find the least-squares estimators for the parameters α and β. What are the values of the minimized χ^2 and the P-values for the two hypotheses?

(b) Assume a relation of the form

$$d = \alpha h^{\beta}\,. \tag{7.3}$$

Write a computer program to perform a least squares fit of α and β. Note that this is a nonlinear function of the parameters and must be solved numerically. A solution using the MINUIT minimization routines from the CERN library is given in `fit_galileo.f, fcn_galileo.f`.

(c) Galileo regarded the motion as the superposition of horizontal and vertical components, where the horizontal motion is of constant speed, and the vertical speed is zero at the lower edge of the ramp, but then increases in direct proportion to the time. Show that this leads to a relation of the form

$$d = \alpha\sqrt{h}\,. \tag{7.4}$$

Find the least squares estimate for α and the value of the minimized χ^2. What is the P-value for this hypothesis?

Exercise 7.2: Consider a least-squares fit to a histogram with y_i entries in bins $i = 1, \ldots, N$, and predicted values

$$\lambda_i(\boldsymbol{\theta}) = n \int_{x_i^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta})dx\,, \tag{7.5}$$

where $f(x;\boldsymbol{\theta})$ depends on unknown parameters $\boldsymbol{\theta}$. Suppose that one replaces the total number of entries n by a parameter ν, and that this is adjusted simultaneously with the other parameters when minimizing

$$\chi^2(\boldsymbol{\theta},\nu) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\boldsymbol{\theta},\nu))^2}{\sigma_i^2}\,. \tag{7.6}$$

(a) Show that taking $\sigma_i^2 = \lambda_i$ leads to the estimator

$$\hat{\nu}_{\mathrm{LS}} = n + \frac{\chi^2_{\mathrm{min}}}{2} \tag{7.7}$$

for the total number of entries.

(b) Show that using $\sigma_i^2 = y_i$ (modified least squares) gives the estimator

$$\hat{\nu}_{\mathrm{MLS}} = n - \chi^2_{\mathrm{min}} \,. \tag{7.8}$$

Exercise 7.3: Consider an LS fit to a histogram with y_i entries in bins $i = 1, \ldots, N$, with predicted values $\lambda_i(\boldsymbol{\theta})$. Suppose the total number of entries n is treated as a constant, so that the y_i are multinomially distributed.

(a) What is the covariance matrix $V_{ij} = \mathrm{cov}[y_i, y_j]$? Why does the inverse of this matrix not exist?

(b) Consider the fit using only the first $N-1$ bins. Find the inverse covariance matrix, and show that this is equivalent to fitting to all N bins but without consideration of the correlations.

Exercise 7.4: Suppose that a data sample of size n has resulted in measurements of N quantities $y_1, \ldots, y_N$, which are to be used in a least-squares fit of some unknown parameters. If the measurements are correlated, one requires the inverse covariance matrix V^{-1} in order to construct the χ^2. Often this is obtained by first determining the matrix of correlation coefficients, $\rho_{ij} = V_{ij}/(\sigma_i \sigma_j)$, e.g. by means of a Monte Carlo calculation.

(a) Recall that for efficient estimators, the inverse covariance matrix is proportional to the sample size n. Show that if this is the case, then the matrix of correlation coefficients is independent of the sample size.

(b) Show that the inverse covariance matrix is given by

$$(V^{-1})_{ij} = \frac{(\rho^{-1})_{ij}}{\sigma_i \sigma_j}. \tag{7.9}$$

(Start with the identity

$$\begin{aligned} \delta_{ij} &= \sum_k (V^{-1})_{ik} V_{kj} \\ &= \sum_k (V^{-1})_{ik} \rho_{kj} \sigma_k \sigma_j . \end{aligned} \tag{7.10}$$

Multiply both sides of (7.10) by ρ^{-1} and sum over the appropriate indices to obtain (7.9).)

Exercise 7.5: Consider two partially overlapping samples of a random variable x, with n and m observations, c of which are common to both. Suppose the variance of x $V[x] = \sigma^2$ is known. Consider the sample means

$$y_1 = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7.11}$$

and

$$y_2 = \frac{1}{m} \sum_{i=1}^{m} x_i \, . \tag{7.12}$$

(a) Show that the covariance is

$$\mathrm{cov}[y_1, y_2] = \frac{c\sigma^2}{nm} \, . \tag{7.13}$$

(b) Using the results of Section 7.6, find the weighted average of y_1 and y_2 and its variance (or standard deviation).

Exercise 7.6: The astronomer Claudius Ptolemy performed experiments on the refraction of light using a circular copper disk submerged to its center in water, as illustrated in Fig. 7.2. Angles of refraction θ_r for 8 values of the angle of incidence θ_i obtained by Ptolemy around 140 A.D. are shown in Table 7.2.[2]

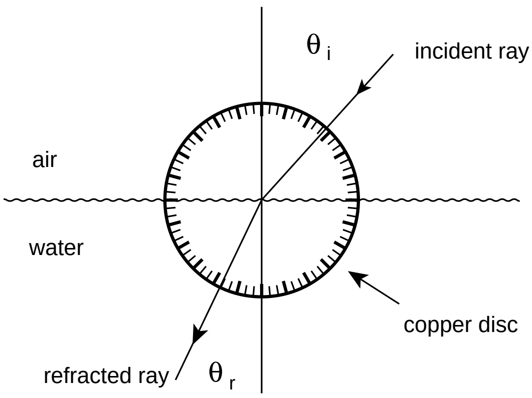


Figure 7.2: The apparatus used by Ptolemy to investigate the refraction of light.

Table 7.2: Angles of incidence and refraction (in degrees).

θ_i	θ_r
10	8
20	$15\frac{1}{2}$
30	$22\frac{1}{2}$
40	29
50	35
60	$40\frac{1}{2}$
70	$45\frac{1}{2}$
80	50

For purposes of this exercise we will take the angles of incidence to be known with negligible error and treat the angles of reflection as independent Gaussian variables with standard

[2] From Olaf Pedersen and Mogens Pihl, *Early Physics and Astronomy: A Historical Introduction*, MacDonald and Janes, London, 1974.

deviations of $\sigma = \frac{1}{2}^\circ$. (This is a reasonable guess given that the angles are reported to the nearest half degree. Note that we can absorb an error in $\theta_{\rm i}$ into an effective error in $\theta_{\rm r}$.)

(a) The correct law of refraction was not discovered until the 17th century. Until then, a commonly used hypothesis was

$$\theta_{\rm r} = \alpha\theta_{\rm i}, \tag{7.14}$$

although it is reported that Ptolemy preferred the formula

$$\theta_{\rm r} = \alpha\theta_{\rm i} - \beta\theta_{\rm i}^2. \tag{7.15}$$

Find the LS estimates of the parameters for both hypotheses and determine the minimized χ^2. Comment on the goodness-of-fit for both hypotheses. Is it plausible that all of the numbers are based on actual measurements?[3]

(b) The law of refraction discovered by Snell in 1621 is

$$\theta_{\rm r} = \sin^{-1}\left(\frac{\sin\theta_{\rm i}}{r}\right), \tag{7.16}$$

where $r = n_{\rm r}/n_{\rm i}$ is the ratio of indices of refraction of the two media. Determine the LS estimate for r and find value of the minimized χ^2. Comment on the validity of the Gaussian assumption for $\theta_{\rm r}$ with $\sigma = \frac{1}{2}^\circ$.

Exercise 7.7: Consider again the problem of Exercise 6.5: N independent Poisson variables $\mathbf{n} = (n_1, \ldots, n_N)$ have mean values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$, where the means are related to a controlled variable x by a relation of the form

$$\nu(x) = \theta a(x). \tag{7.17}$$

(a) Consider first the LS method, where the denominators in the χ^2 use the variances $\sigma_i^2 = \nu_i$. Show that the LS estimator for θ is given by

$$\hat{\theta} = \left(\frac{\sum_{i=1}^N \frac{n_i^2}{a(x_i)}}{\sum_{i=1}^N a(x_i)}\right)^{1/2}. \tag{7.18}$$

By expanding $\hat{\theta}(\mathbf{n})$ in a Taylor series to second order about $\boldsymbol{\nu}$ and computing the expectation value, show that the bias of (7.18) is given by

$$b = \frac{N-1}{2\sum_{i=1}^N a(x_i)} + O(E[(n_i - \nu_i)^3]). \tag{7.19}$$

(Use $\mathrm{cov}[n_i, n_j] = \delta_{ij}\nu_j$ for the covariance of independent Poisson variables.)

[3] See R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics*, Vol. I, Addison-Wesley, Menlo Park, 1963, Section 26-2.

(b) Repeat (a) using the modified LS method, where the χ^2 uses variances based on the observed values: $\sigma_i^2 = n_i$. Show that the MLS estimator for θ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{N} a(x_i)}{\sum_{i=1}^{N} \frac{a(x_i)^2}{n_i}}, \tag{7.20}$$

and that its bias is given by

$$b = -\frac{N-1}{\sum_{i=1}^{N} a(x_i)} + O(E[(n_i - \nu_i)^3]). \tag{7.21}$$

Compare the biases from (a) and (b) to the results of Exercise 7.2.

(c) Estimate the variance $\hat{\theta}$ for both the LS and MLS cases using error propagation.

Note that since it was shown in Exercise 6.5 that the ML estimator for θ is both unbiased and has minimum variance, the LS and MLS estimators are not preferred here. For sufficiently large data samples, however, the three methods are very similar; cf. Exercise 7.8.

Exercise 7.8: Consider again Perrin's data on the number of mastic particles as a function of height (Exercise 6.5). Determine the LS estimates for Boltzmann's constant k (and equivalently Avogadro's number $N_{\mathrm{A}} = R/k$) and the coefficient ν_0 by minimizing

$$\chi^2(k, \nu_0) = \sum_{i=1}^{N} \frac{(n_i - \nu_i(k, \nu_0))^2}{\sigma_i^2}. \tag{7.22}$$

(a) Take the standard deviation σ_i of n_i to be $\sqrt{\nu_i}$ (the usual method of least squares).

(b) Take σ_i to be $\sqrt{n_i}$ (the modified method of least squares).

Compare the results from (a) and (b) to the estimates obtained by maximum likelihood in Exercise 6.5.

Chapter 8

The Method of Moments

Exercise 7.1: Consider a random variable x distributed according to a Gaussian p.d.f. of unknown mean μ and variance σ^2, and suppose we have a sample of values $x_1, \ldots, x_n$.

(a) Construct estimators for μ and σ^2 using the method of moments. Use the functions $a_1 = x$, $a_2 = x^2$, so that the expectation values $E[a_i(x)]$ correspond to the first and second algebraic moments of x.

(b) Compute the expectation values of the estimators $\hat{\mu}$ and $\widehat{\sigma^2}$ from (a). Are the estimators biased?

Exercise 7.2: Consider ρ^0 mesons produced in a particle reaction which decay into two charged pions ($\pi^+\pi^-$). The decay angle θ is defined as the angle of the π^+ with respect to the original direction of the ρ, measured in the $\pi^+\pi^-$ rest frame (see Fig. 8.1).

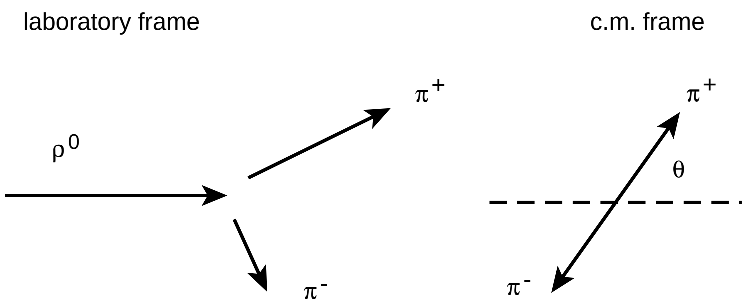


Figure 8.1: The definition of the decay angle θ in the decay $\rho^0 \to \pi^+\pi^-$.

Since the ρ^0 has spin 1 and the pions have spin 0, one can show that the distribution of $\cos\theta$ has the form

$$f(\cos\theta; \eta) = \frac{1}{2}(1-\eta) + \frac{3}{2}\eta\cos^2\theta, \tag{8.1}$$

where the spin-alignment parameter η can take on values in the range $-\frac{1}{2} \leq \eta \leq 1$.

(a) Suppose that n values of $\cos\theta$ have been measured for ρ^0 mesons produced in a certain reaction. Construct an estimator $\hat{\eta}$ for the spin alignment using the method of moments, by using the function $a = x^2$. Why is it not possible to construct an estimator using $a = x$?

(b) Determine the expectation value and variance of $\hat{\eta}$.

Chapter 9

Statistical Errors, Confidence Intervals and Limits

Exercise 9.1: Suppose an estimator $\hat{\theta}$ is Gaussian distributed about the parameter's true value θ with a standard deviation $\sigma_{\hat{\theta}}$. Assume that $\sigma_{\hat{\theta}}$ is known.

(a) Sketch the functions $u_\alpha(\theta)$ and $v_\beta(\theta)$ defining the confidence belt (cf. SDA Section 9.2).

(b) Show that the central confidence interval for θ at a confidence level $1-\gamma$ is given by

$$[\hat{\theta} - \sigma_{\hat{\theta}}\Phi^{-1}(1-\gamma/2),\, \hat{\theta} + \sigma_{\hat{\theta}}\Phi^{-1}(1-\gamma/2)], \tag{9.1}$$

where Φ^{-1} is the quantile of the standard Gaussian.

Exercise 9.2: (a) Consider n observations of an exponentially distributed variable x with mean ξ. The ML estimator for ξ (see SDA (6.6)) is given by

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{9.2}$$

and the p.d.f. for $\hat{\xi}$ (cf. SDA equation (10.25)) is

$$g(\hat{\xi};\xi) \;=\; \frac{n^n}{(n-1)!}\frac{\hat{\xi}^{n-1}}{\xi^n}e^{-n\hat{\xi}/\xi}. \tag{9.3}$$

(a) Show that the curves defining the confidence belt, $u_\alpha(\xi)$ and $v_\beta(\xi)$, are given by

$$\begin{aligned} u_\alpha(\xi) &= \frac{\xi}{2n}F^{-1}_{\chi^2}(1-\alpha;2n), \\ v_\beta(\xi) &= \frac{\xi}{2n}F^{-1}_{\chi^2}(\beta;2n), \end{aligned} \tag{9.4}$$

where $F^{-1}_{\chi^2}$ is the quantile of the χ^2 distribution. Use the fact that the cumulative χ^2 distribution can be related to the incomplete gamma function $P(n,x)$ by (cf. Exercise 2.5)

$$F_{\chi^2}(2x; 2n) = P(n, x) \equiv \int_0^x e^{-t} t^{n-1} dt. \tag{9.5}$$

Make a sketch of $u_\alpha(\xi)$ and $v_\beta(\xi)$ using $\alpha = \beta = 0.159$ and $n = 5$. (Quantiles of the χ^2 distribution can be looked up in standard tables or obtained from the routine `CHISIN` from the CERN program library.)

(b) Find the confidence interval $[a, b]$ as a function of the estimate $\hat{\xi}$, the sample size n and the confidence levels α and β. Suppose the estimate is $\hat{\xi} = 1.0$. Sketch this on the plot of $u_\alpha(\xi)$ and $v_\beta(\xi)$. Evaluate a and b for $n = 5$, $\alpha = \beta = 0.159$. Compare the result to the interval obtained from plus or minus one standard deviation about the estimate $\hat{\xi}$.

Exercise 9.3: Show that the upper and lower limits for the parameter p of a binomial distribution are

$$\begin{aligned} p_{\rm lo} &= \frac{n F_F^{-1}[\alpha; 2n, 2(N-n+1)]}{N-n+1 \;+\; n F_F^{-1}[\alpha; 2n, 2(N-n+1)]} \\ p_{\rm up} &= \frac{(n+1) F_F^{-1}[1-\beta; 2(n+1), 2(N-n)]}{(N-n) \;+\; (n+1) F_F^{-1}[1-\beta; 2(n+1), 2(N-n)]} \,. \end{aligned} \tag{9.6}$$

Here the confidence levels for the upper and lower limits are $1-\alpha$ and $1-\beta$, respectively, n is the number of successes observed in N trials, and F_F^{-1} is the quantile of the F distribution. This is defined by the p.d.f.

$$f(x; n_1, n_2) = \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{\Gamma(\frac{1}{2}(n_1+n_2))}{\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)}\, x^{n_1/2-1} \left(1 + \frac{n_1}{n_2} x\right)^{-(n_1+n_2)/2} , \tag{9.7}$$

where $x > 0$ and n_1 and n_2 are integer parameters (degrees of freedom). Use the fact that the cumulative binomial distribution is related to the cumulative distribution $F_F(x)$ for $n_1 = 2(n+1)$ and $n_2 = 2(N-n)$ degrees of freedom by [1]

$$\sum_{k=0}^{n} \frac{N!}{k!(N-k)!}\, p^k \,(1-p)^{N-k} = 1 - F_F\left[\frac{(N-n)p}{(n+1)(1-p)}; 2(n+1), 2(N-n)\right] . \tag{9.8}$$

Quantiles of the F distribution can be obtained from standard tables or computed with the routine `ffinv`. Equations (9.6) are implemented in the routines `binomlo`, `binomup` and `binomint`.

Exercise 9.4: In an antineutrino-nucleon scattering experiment with the Gargamelle bubble chamber at CERN, events were selected having only hadrons (from the neutral-current proccess $\overline{\nu}_\mu N \to \overline{\nu}_\mu X$) or with hadrons and a muon (the charged-current procces $\overline{\nu}_\mu N \to \mu^+ X$). Out of a sample of 212 events, 64 were classified as neutral current (NC) and 148 as charged current (CC). Estimate the probability for an event to be NC and find the 68.3% central confidence

[1] Use of the F distribution for evaluating binomial confidence intervals is due to A. Hald, *Statistical Theory with Engineering Applications*, John Wiley, New York, 1952.

interval. Find the corresponding estimator and interval for the ratio of probabilities for NC and CC events.[2]

Exercise 9.5: In an experiment investigating particle collisions, 10 events are selected as being of a certain type, say, having a high value of some property x. Out of the 10 high-x events, 2 are found to contain muons.

(a) Find the 68.3% central confidence interval for the binomial parameter p for high-x events to contain muons. Express the answer as $p = \hat{p}^{+c}_{-d}$ where $\hat{p}$ is the ML estimate for p and $[\hat{p}-c, \hat{p}+d]$ is the confidence interval. (The routine `binomint` can be used.)

(b) Compare the interval from (a) to $\hat{p} \pm \hat{\sigma}_{\hat{p}}$, where $\hat{\sigma}_{\hat{p}}$ is the estimate of the standard deviation of $\hat{p}$.

(c) A common mistake is to regard the number 10 of high-x events as a random variable and to include its variance in the error for $\hat{p}$ (e.g. using error propagation). Why is this not the correct approach?

Exercise 9.6: Suppose that to produce the events in Exercise 9.5, the total amount of data collected corresponded to an integrated luminosity of $L = 1\ \text{pb}^{-1}$ (known with a negligible error). The total number of events produced of a given type can be regarded as a Poisson variable n with mean value $\nu = \sigma L$, where σ is the production cross section. (Why is the Poisson distribution appropriate?)

(a) Find the 68.3% central confidence intervals for the expected numbers of events ν_x and $\nu_{x\mu}$ for events with high x and high x with muons, given $n_x = 10$ and $n_{x\mu} = 2$ events observed. What are the corresponding confidence intervals for the production cross sections, σ_x and $\sigma_{x\mu}$?

(b) Compare the confidence intervals from (a) to intervals constructed as plus or minus one standard deviation about the estimate of the corresponding parameter.

(c) Suppose that in a separate experiment with an integrated luminosity of $L' = 100\ \text{pb}^{-1}$, $n'_x = 1173$ high-x events are observed. This experiment, however, is unable to identify muons. Construct the log-likelihood function for the parameters σ_x and $p = \sigma_{x\mu}/\sigma_x$ using the data n_x, $n_{x\mu}$ and n'_x. Show that the ML estimator for p does not depend on n'_x. Does it make sense that the result of the second experiment has no impact on the estimate of p?

(d) Suppose the original experiment had not measured the number of high-x events but had only reported the number of high-x events with muons. Using only the two results $n_{x\mu} = 2$ and $n'_x = 1173$, construct the log-likelihood function for σ_x and p. From this find the ML estimators. Use error propagation to estimate the standard deviation of $\hat{p}$, and compare the interval $\hat{p} \pm \hat{\sigma}_{\hat{p}}$ to the intervals from Exercises 9.5 (a) and (b). Optional: How would you go about constructing a confidence interval for p in this case?

Exercise 9.7: A Particle created in an interaction is emitted at a certain angle with respect to the z axis, as shown in Fig. 9.1. A detector located a distance d from the interaction point measures the particle's position x perpendicular to the z direction. The angle θ is defined as the angle between the z axis and the projection of the particle's trajectory into the (x, z) plane. Suppose the measured value x can be regarded as a Gaussian variable centered about the true value and having a standard deviation σ_x.

[2] In the actual experiment, small background corrections were included; see F.J. Hasert et al., Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment, *Phys. Lett.* **46B** (1973) 138.

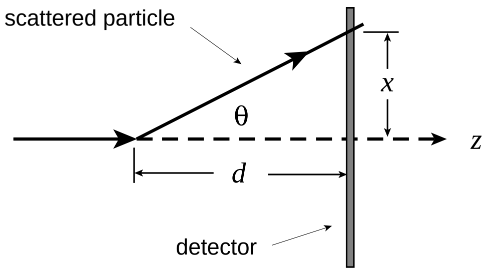


Figure 9.1: The definition of the scattering angle θ from the trajectory projected into the (x, z) plane.

(a) Find the central confidence interval at a confidence level $1 - \gamma$ for the cosine of the angle θ. Assume that the distance d is known without error.

(b) Take $d = 1$ m, $\sigma_x = 1$ mm and suppose the measured value is $x = 2.0$ mm. Find the 68.3% and 95% central confidence intervals for $\cos\theta$.

Chapter 10

Characteristic Functions

Exercise 10.1: Show that the characteristic function of the Gaussian p.d.f.,

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) , \tag{10.1}$$

is given by

$$\phi(k) = \exp(i\mu k - \tfrac{1}{2}\sigma^2 k^2) . \tag{10.2}$$

Exercise 10.2: Show that the characteristic function of the exponential p.d.f.,

$$f(x;\xi) = \frac{1}{\xi} e^{-x/\xi}, \tag{10.3}$$

is given by

$$\phi(k) = \frac{1}{1 - ik\xi}. \tag{10.4}$$

Exercise 10.3: Show that the characteristic function of the χ^2 p.d.f. for n degrees of freedom,

$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2}, \tag{10.5}$$

is given by

$$\phi(k) = (1 - 2ik)^{-n/2}. \tag{10.6}$$

For this you will need the definition of the gamma function,

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt. \tag{10.7}$$

Exercise 10.4: Suppose the random variables $x_1, \ldots, x_n$ are independent and each follow a Gaussian distribution of mean μ and variance σ^2. As seen in Chapters 5 and 6, the sample mean

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{10.8}$$

can be used as an estimator for the mean μ.

(a) Find the characteristic function for the sample mean.

(b) From this, show that the p.d.f. for $\overline{x}$ is itself Gaussian, and find its mean and variance.

Exercise 10.5 Using the characteristic function, show that the first four algebraic moments of the Gaussian distribution are

$$\begin{aligned} E[x] &= \mu \\ E[x^2] &= \mu^2 + \sigma^2 \\ E[x^3] &= \mu^3 + 3\mu\sigma^2 \\ E[x^4] &= 3(\mu^2 + \sigma^2)^2 . \end{aligned} \tag{10.9}$$

Exercise 10.6: (a) Using the characteristic function, show that the mean and variance of the χ^2 distribution for n degrees of freedom are n and $2n$, respectively.

(b) Suppose z follows the χ^2 distribution for n degrees of freedom. Show that in the limit of large n this becomes a Gaussian distribution with mean $\mu = n$ and variance $\sigma^2 = 2n$. To do this, define the variable

$$y = \frac{z - n}{\sqrt{2n}} , \tag{10.10}$$

which has a mean of zero and standard deviation of unity. Show that the characteristic function for y is

$$\phi_y(k) = e^{-ik\sqrt{n/2}} \phi_z \left(\frac{k}{\sqrt{2n}} \right) . \tag{10.11}$$

Expand the logarithm of $\phi_y(k)$ and retain terms that do not vanish in the limit of large n. Transform back to the original variable z to obtain the final result.

Exercise 10.7: Suppose n independent random variables $x_1, \ldots, x_n$ each follow a standard Gaussian distribution, i.e.,

$$\varphi(x_i) = \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \tag{10.12}$$

for all i, and consider

$$y = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} . \tag{10.13}$$

(a) First consider only one of the x_i. By transformation of variables, show that the p.d.f. of $u = x_i^2$ is

$$f(u) = \frac{1}{\sqrt{2\pi u}}\, e^{-u/2}. \tag{10.14}$$

This is the χ^2 distribution for one degree of freedom.

(b) Show that the characteristic function of u is

$$\phi_u(k) = \frac{1}{\sqrt{1 - 2ik}}. \tag{10.15}$$

(c) Using the addition theorem, find the characteristic function for

$$v = \sum_{i=1}^{n} x_i^2. \tag{10.16}$$

(d) Using transformation of variables, show that the p.d.f. for $y = (\sum_{i=1}^{n} x_i^2)^{1/2}$ is

$$h(y) = \frac{1}{2^{n/2-1}\Gamma(n/2)}\, y^{n-1}\, e^{-y/2}. \tag{10.17}$$

This is a special case of the gamma distribution.

(e) Write down the p.d.f. for $n = 3$. This is the Maxwell-Boltzmann distribution. Suppose the velocity components of molecules in a gas v_x, v_y and v_z are independent Gaussian variables with mean values of zero, and standard deviations σ. Write down the p.d.f. for the molecular speed $v = (v_x^2 + v_y^2 + v_z^2)^{1/2}$.

(f) Write down the p.d.f. for $n = 1$. That is, if x follows a standard Gaussian, what is the p.d.f. of $y = |x|$?

Exercise 10.8: Consider a variable x distributed according to the Cauchy (Breit-Wigner) p.d.f.,

$$f(x) \;=\; \frac{1}{\pi}\,\frac{1}{1+x^2}. \tag{10.18}$$

(a) Show that the characteristic function is

$$\phi(k) \;=\; e^{-|k|}. \tag{10.19}$$

(Use the residue theorem and close the integral in the upper half plane for $k > 0$, and in the lower half plane for $k < 0$.)

(b) Consider a sample of n observations of a Cauchy distributed variable x. Using the addition theorem with the characteristic function from (a), show that the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ also follows the Cauchy p.d.f. This is a rare case where the p.d.f. of $\overline{x}$ does not change as the sample size increases, and is related to the fact that the moments of the Cauchy distribution does not exist.

Exercise 10.9: The Dirac delta function,

$$f(x;\mu) = \delta(x-\mu), \tag{10.20}$$

is defined by

$$\begin{aligned} \delta(x-\mu) &= 0 \quad \text{for } x \neq \mu\,, \\ \int_{-\infty}^{\infty} \delta(x-\mu)\,dx &= 1. \end{aligned} \tag{10.21}$$

That is, $\delta(x-\mu)$ has an infinitely sharp peak at $x = \mu$ and is zero elsewhere. Find the characteristic function of $\delta(x-\mu)$, and use this to obtain an integral representation of the delta function.

Chapter 11

Unfolding

Exercise 11.1: Consider the detector set-up shown in Fig. 9.1. Suppose the resolution function for x is Gaussian,

$$s(x|x') = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-x')^2}{2\sigma_x^2}\right]. \tag{11.1}$$

Find the resolution function for $\cos\theta = x/\sqrt{x^2+d^2}$.

Exercise 11.2: Consider the Tikhonov regularization function with $k = 1$ for a histogram $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ with equal bin widths,

$$S(\boldsymbol{\mu}) = -\sum_{i=1}^{M-1} (\mu_i - \mu_{i+1})^2. \tag{11.2}$$

Find the $M \times M$ matrix G such that $S(\boldsymbol{\mu})$ can be expressed in the form

$$S(\boldsymbol{\mu}) = -\sum_{i,j=1}^{M} G_{ij}\mu_i\mu_j = -\boldsymbol{\mu}^T G \boldsymbol{\mu}. \tag{11.3}$$

Exercise 11.3: Consider a histogram of expectation values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ and the corresponding probabilities $\mathbf{p} = \boldsymbol{\mu}/\mu_{\rm tot}$, where $\mu_{\rm tot} = \sum_{i=1}^{M} \mu_i$.

(a) Show that the Shannon entropy,

$$H(\mathbf{p}) = -\sum_{i=1}^{M} p_i \log p_i, \tag{11.4}$$

is maximum when $p_i = 1/M$ for all i. (Use a Lagrange multiplier to impose the constraint $\sum_{i=1}^{M} p_i = 1$.)

(b) Show that the cross-entropy,

$$K(\mathbf{p};\mathbf{q}) = -\sum_{i=1}^{M} p_i \log \frac{p_i}{Mq_i}, \tag{11.5}$$

is maximum when the probabilities $\mathbf{p}$ are equal to the reference distribution $\mathbf{q}$.

Exercise 11.4: Consider an observed histogram $\mathbf{n} = (n_1, \ldots, n_N)$, for which the corresponding expectation values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ are related to a true histogram of expectation values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ by $\boldsymbol{\nu} = R\boldsymbol{\mu}$. Assume that the covariance matrix V and response matrix R are known and that the histograms contain no background.

(a) Construct estimators for $\boldsymbol{\mu}$ by maximizing

$$\begin{aligned} \Phi(\boldsymbol{\mu}) &= -\frac{\alpha}{2}\chi^2(\boldsymbol{\mu}) + S(\boldsymbol{\mu}) \\ &= -\frac{\alpha}{2}(\mathbf{n} - R\boldsymbol{\mu})^T V^{-1}(\mathbf{n} - R\boldsymbol{\mu}) - \boldsymbol{\mu}^T G \boldsymbol{\mu}\,, \end{aligned} \tag{11.6}$$

where α is the regularization parameter and the $M \times M$ symmetric matrix G is given by a known set of constants (cf. SDA Section 11.5.1). Show that the estimators $\hat{\boldsymbol{\mu}}$ are given by

$$\hat{\boldsymbol{\mu}} = (\alpha R^T V^{-1} R + 2G)^{-1} \alpha R^T V^{-1} \mathbf{n}\,, \tag{11.7}$$

and find the covariance matrix $U_{ij} = \mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j]$.

(b) Now impose the constraint that $\nu_{\mathrm{tot}} = \sum_{i=1}^N \nu_i = \sum_{i=1}^N \sum_{j=1}^M R_{ij}\mu_j$ be equal to the total observed number of events $n_{\mathrm{tot}} = \sum_{i=1}^N n_i$. The solution is found by maximizing

$$\varphi(\boldsymbol{\mu}) = -\frac{\alpha}{2}(\mathbf{n} - R\boldsymbol{\mu})^T V^{-1}(\mathbf{n} - R\boldsymbol{\mu}) - \boldsymbol{\mu}^T G \boldsymbol{\mu} + \lambda(n_{\mathrm{tot}} - \nu_{\mathrm{tot}}) \tag{11.8}$$

with respect to the parameters $\boldsymbol{\mu}$ and the Lagrange multiplier λ. Find the estimators $\hat{\boldsymbol{\mu}}$ and their covariance matrix.

(c) Construct estimators $\hat{\mathbf{b}}$ for the bias $\mathbf{b} = E[\hat{\boldsymbol{\mu}}] - \boldsymbol{\mu}$ using SDA equation (11.76) for both cases (a) and (b).