Chapter 5
Maximum Likelihood and Other Methods to Estimate Variables

Abstract In this chapter we study the problem of estimating parameters of the distribution function of a random variable when N observations of the variable are available. We discuss methods that establish what *sample* quantities must be calculated to estimate the corresponding *parent* quantities. This establishes a firm theoretical framework that justifies the definition of the sample variance as an unbiased estimator of the parent variance, and the sample mean as an estimator of the parent mean. One of these methods, the maximum likelihood method, will later be used in more complex applications that involve the fit of two-dimensional data and the estimation of fit parameters. The concepts introduced in this chapter constitute the core of the statistical techniques for the analysis of scientific data.

5.1 The Maximum Likelihood Method for Gaussian Variables

Consider a random variable X distributed like a Gaussian. The probability of making a measurement between x_i and $x_i + dx$ is given by

$$f(x_i)dx = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} dx.$$

This probability describes the *likelihood* of collecting the data point x_i given that the distribution has a fixed value of μ and σ. Assume now that N measurements of the random variable have been made. The goal is to estimate the *most likely* values of the true—yet unknown—values of μ and σ, the two parameters that determine the distribution of the random variable. The method of analysis that follows this principle is called the *maximum likelihood method*. The method is based on the postulate that the values of the unknown parameters are those that yield a maximum probability of observing the measured data. Assuming that the measurements are made independently of one another, the quantity

$$P = \prod_{i=1}^{N} P(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{5.1}$$

M. Bonamente, *Statistics and Analysis of Scientific Data*, Graduate Texts in Physics, DOI 10.1007/978-1-4939-6572-4_5

is the probability of making N independent measurements in intervals of unit length around the values x_i, which can be viewed as the probability of measuring the dataset composed of the given N measurements.

The method of maximum likelihood consists therefore of finding the parameters of the distribution that maximize the probability in (5.1). This is simply achieved by finding the point at which the first derivative of the probability P with respect to the relevant parameter of interest vanishes, to find the extremum of the function. It can be easily proven that the second derivative with respect to the two parameters is negative at the point of extremum, and therefore this is a point of maximum for the likelihood function.

5.1.1 Estimate of the Mean

To find the maximum-likelihood estimate of the mean of the Gaussian distribution we proceed with the calculation of the first derivative of $\ln P$, instead of P, with respect to the mean μ. Given that the logarithm is a monotonic function of the argument, maximization of $\ln P$ is equivalent to that of P, and the logarithm has the advantage of ease of computation. We obtain

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2} = 0.$$

The solution is the maximum-likelihood estimator of the mean, which we define as μ_{ML}, and is given by

$$\mu_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i = \overline{x}. \tag{5.2}$$

This result was to be expected: the maximum likelihood method shows that the "best" estimate of the mean is simply the sample average of the measurements.

The quantity μ_{ML} is a quantity that, despite the Greek letter normally reserved for parent quantities, is a function of the measurements. Although it appears obvious that the sample average is the correct estimator of the true mean, it is necessary to prove this statement by calculating its expectation. It is clear that the expectation of the sample average is in fact

$$E[\bar{x}] = \frac{1}{N} E[x_1 + \ldots + x_N] = \mu,$$

This calculation is used to conclude that the sample mean is an *unbiased* estimator of the true mean.

5.1.2 Estimate of the Variance

Following the same method used to estimate the mean, we can also take the first derivative of $\ln P$ with respect to σ^2, to obtain

$$\frac{\partial}{\partial \sigma^2}\left(N \ln \frac{1}{\sqrt{2\pi\sigma^2}}\right) + \frac{\partial}{\partial \sigma^2}\left(\sum -\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right) = 0$$

from which we obtain

$$N\left(-\frac{1}{2}\frac{1}{\sigma^2}\right) + \sum -\frac{1}{2}(x_i-\mu)^2\left(-\frac{1}{\sigma^4}\right) = 0$$

and finally the result that the maximum likelihood estimator of the variance is

$$\sigma^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2. \tag{5.3}$$

It is necessary to notice that in the maximum likelihood estimate of the variance we have implicitly assumed that the mean μ was known, while in reality we can only estimate it as the sample mean, from the same data used also to estimate the variance. To account for the fact that μ is not known, we replace it with $\overline{x}$ in (5.3), and call

$$s^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2 \tag{5.4}$$

the *maximum likelihood sample variance estimator*, which differs from the sample variance defined in (2.11) by a factor of $(N-1)/N$. The fact that $\overline{x}$ replaced μ in its definition leads to the following expectation:

$$E[s^2_{ML}] = \frac{N-1}{N}\sigma^2_{ML}. \tag{5.5}$$

Proof Calculation of the expectation is obtained as

$$E[s^2_{ML}] = E[\frac{1}{N}\sum(x_i-\bar{x})^2] = \frac{1}{N}E[\sum(x_i-\mu+\mu-\bar{x})^2]$$

$$= \frac{1}{N}E[\sum(x_i-\mu)^2 + \sum(\mu-\bar{x})^2 + 2(\mu-\bar{x})\sum(x_i-\mu)]$$

The term $E[\sum(\mu - \bar{x})^2]$ is the variance of the sample mean, which we know from Sect. 4.1 to be equal to σ^2/N. The last term in the equation is $\sum(x_i - \mu) = N(\bar{x} - \mu)$, therefore:

$$E[s^2_{ML}] = \frac{1}{N}\left(E[\sum(x_i - \mu)^2] + NE[(\mu - \bar{x})^2] + 2NE[(\mu - \bar{x})(\bar{x} - \mu)]\right)$$

$$= \frac{1}{N}(N\sigma^2 + N\sigma^2/N - 2NE[(\mu - \bar{x})^2])$$

leading to the result that

$$E[s^2_{ML}] = \frac{1}{N}\left(N\sigma^2 + N\sigma^2/N - 2N\sigma^2/N\right) = \frac{N-1}{N}\sigma^2.$$

In this proof we used the notation $\sigma^2 = \sigma^2_{ML}$ □

This result is at first somewhat surprising, since there is an extra factor $(N-1)/N$ that makes $E[s^2_{ML}]$ different from the maximum likelihood estimator of σ^2. This is actually due to the fact that, in estimating the variance, the mean needed to be estimated as well and was not known beforehand. The unbiased estimator of the variance is therefore

$$s^2 = s^2_{ML} \times \frac{N}{N-1} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 \tag{5.6}$$

for which we have shown that $E[s^2] = \sigma^2$. This is the reason for the definition of the sample variance according to (5.6), and not (5.4).

It is important to pay attention to the fact that (5.6) defines a statistic for which we could also find, in addition to its expectation, also its variance, similar to what was done for the sample mean. In Chap. 7 we will study how to determine the probability distribution function of certain statistics of common use, including the distribution of the sample variance s^2.

Example 5.1 We have already made use of the sample mean and the sample variance as estimators for the parent quantities in the analysis of the data from Thomson's experiment (page 23). The estimates we obtained are unbiased if the assumptions of the maximum likelihood method are satisfied, namely that I and W/Q are Gaussian distributed. ◇

5.1.3 Estimate of Mean for Non-uniform Uncertainties

In the previous sections we assumed a set of measurements x_i of the same random variable, i.e., the parent mean μ and variance σ^2 were the same. It is often the case with real datasets that observations are made from variables with the same mean, but with different variance. This could be the case when certain measurements are more precise than others, and therefore they feature the same mean (since they are drawn from the same process), but the standard error varies with the precision of the instrument, or because some measurements were performed for a longer period of time. In this case, each measurement x_i is assigned a different standard deviation σ_i, which represents the precision with which that measurement was made.

Example 5.2 A detector is used to measure the rate of arrival of a certain species of particles. One measurement consists of 100 counts in 10 s, another of 180 particles in 20 s, and one of 33 particles in 3 s. The measured count rates would be reported as, respectively, 10.0, 9.0, and 11.0 counts per second. Given that this is a counting experiment, the Poisson distribution applies to each of the measurements. Moreover, since the number of counts is sufficiently large, it is reasonable to approximate the Poisson distribution with a Gaussian, with variance equal to the mean. Therefore the variance of the counts is 100, 180, and 33, and the variance of the count rate can be calculated by the property that $Var[X/t] = Var[X]/t^2$, where t is the known time of each measurement. It follows that the standard deviation σ of the count rates is, respectively, 1.0, 0.67, and 1.91 for the three measurements. The three measurements would be reported as 10.0 ± 1.0, 9.0 ± 0.67, and 11.0 ± 1.91, with the last measurement being clearly of lower precision because of the shorter period of observation. ◇

Our goal is therefore now focused on the maximum likelihood estimate of the parent mean μ, which is the same for all measurements. This is achieved by using (5.1) in which the parent mean of each measurement is μ, and the parent variance of each measurement is σ_i^2. Following the same procedure as in the case of equal standard deviations for each measurement, we start with the probability P of making the N measurements,

$$P = \prod_{i=1}^{N} P(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma_i^2}}. \tag{5.7}$$

Setting the derivative of $\ln P$ with respect to μ—the common mean to all measurements—equal to zero, we obtain that the maximum likelihood estimates is

$$\mu_{ML} = \frac{\sum_{i=1}^{N}(x_i/\sigma_i^2)}{\sum_{i=1}^{N}(1/\sigma_i^2)} \tag{5.8}$$

This is the weighted mean of the measurements, where the weights are the inverse of the variance of each measurement, $1/\sigma_i^2$.

The variance in this weighted sample mean can be calculated using the expectation of the weighted mean, by assuming that the σ_i^2 are constant numbers:

$$Var(\mu_{ML}) = \frac{1}{\left(\sum_{i=1}^N (1/\sigma_i^2)\right)^2} \times \sum_{i=1}^N \frac{Var(x_i)}{\sigma_i^4} = \frac{\sum_{i=1}^N (1/\sigma_i^2)}{\left(\sum_{i=1}^N (1/\sigma_i^2)\right)^2},$$

which results in

$$\sigma_\mu^2 = \frac{1}{\sum_{i=1}^N (1/\sigma_i^2)}. \tag{5.9}$$

The variance of the weighted mean on (5.9) becomes the usual σ^2/N if all variances σ_i^2 are identical.

Example 5.3 Continuing Example 5.2 of the count rate of particle arrivals, we use (5.8) and (5.9) to calculate a weighted mean and standard deviation of 9.44 and 0.53. Since the interest is just in the overall mean of the rate, the more direct means to obtain this number is by counting a total of 313 counts in 33 s, for an overall measurement of the count rate of 9.48 ± 0.54, which is virtually identical to that obtained using the weighted mean and its variance. ◇

It is common, as in the example above, to assume that the parent variance σ_i^2 is equal to the value estimated from the measurements themselves. This approximation is necessary, unless the actual precision of the measurement is known beforehand by some other means, for example because the apparatus used for the experiment has been calibrated by prior measurements.

5.2 The Maximum Likelihood Method for Other Distributions

The method of maximum likelihood can also be applied when the measurements do not follow a Gaussian distribution.

A typical case is that of N measurements n_i, $i = 1, \ldots, N$, from a Poisson variable N of parameter λ, applicable to all situations in which the measurements are derived from a counting experiment. In this case, the maximum likelihood method can be used to estimate λ, which is the mean of the random variable, and the only parameter of the Poisson distribution.

The Poisson distribution is discrete in nature and the probability of making N independent measurements is simply given by

$$P = \prod_{i=1}^N \frac{\lambda^{n_i}}{n_i!} e^{-\lambda}.$$

It is convenient to work with logarithms,

$$\ln P = \sum_{i=1}^{N} \ln e^{-\lambda} + \sum_{i=1}^{N} \ln \frac{1}{n_i!} + \sum_{i=1}^{N} \ln \lambda^{n_i} = A - N\lambda + \ln \lambda \sum_{i=1}^{N} n_i$$

in which A is a term that doesn't depend on λ. The condition that the probability must be maximum requires $\partial P/\partial \lambda = 0$. This condition results in

$$\frac{1}{\lambda} \sum_{i=1}^{N} x_i - N = 0,$$

and therefore we obtain that the maximum likelihood estimator of the λ parameter of the Poisson distribution is

$$\lambda_{ML} = \frac{1}{N} \sum_{i=1}^{N} n_i.$$

This result was to be expected, since λ is the mean of the Poisson distribution, and the linear average of N measurements is an unbiased estimate of the mean of a random variable, according to the Law of Large Numbers.

The maximum likelihood method can in general be used for any type of distribution, although often the calculations can be mathematically challenging if the distribution is not a Gaussian.

5.3 Method of Moments

The method of moments takes a more practical approach to the estimate of the parameters of a distribution function. Consider a random variable X for which we have N measurements and whose probability distribution function f(x) depends on M unknown parameters, for example $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ for a Gaussian ($M = 2$), or $\theta_1 = \lambda$ for an exponential ($M = 1$), etc. The idea is to develop a method that yields as many equations as there are free parameters and solve for the parameters of the distribution. The method starts with the determination of arbitrary functions $a_j(x), j = 1, \ldots M$, that make the distribution function integrable:

$$E[a_j(X)] = \int_{-\infty}^{\infty} a_j(x) f(x) dx = g_j(\theta) \tag{5.10}$$

where $g_j(\boldsymbol{\theta})$ is an analytic function of the parameters of the distribution. Although we have assumed that the random variable is continuous, the method can also be applied to discrete distributions. According to the law of large numbers, the left-hand side of (5.10) can be approximated by the sample mean of the function of the N measurements, and therefore we obtain a linear system of M equations:

$$\frac{1}{N}(a_j(x_1) + \ldots + a_j(x_N)) = g_j(\theta) \tag{5.11}$$

which can be solved for the parameters $\boldsymbol{\theta}$ as function of the N measurements x_i.

As an illustration of the method, consider the case in which the parent distribution is a Gaussian of parameters μ, σ^2. First, we need to decide which functions $a_1(x)$ and $a_2(x)$ to choose. A simple and logical choice is to use $a_1(x) = x$ and $a_2(x) = x^2$; this choice is what gives the name of "moments," since the right-hand side of (5.10) will be, respectively, the first and second order moment. Therefore we obtain the two equations

$$\begin{cases} E[a_1(X)] = \dfrac{1}{N}(X_1 + \ldots + X_N) = \mu \\ E[a_2(X)] = \dfrac{1}{N}(X_1^2 + \ldots + X_N^2) = \sigma^2 + \mu^2. \end{cases} \tag{5.12}$$

The estimator for mean and variance are therefore

$$\begin{cases} \mu_{MM} = \dfrac{1}{N}(X_1 + \ldots + X_N) \\ \sigma^2_{MM} = \dfrac{1}{N}(X_1^2 + \ldots + X_N^2) - \left(\dfrac{1}{N}(X_1 + \ldots + X_N)\right)^2 = \dfrac{1}{N}\sum(x_i - \mu_{MM})^2 \end{cases} \tag{5.13}$$

which, in this case, are identical to the estimates obtained from the likelihood method. This method is often easier computationally than the method of maximum likelihood, since it does not require the maximization of a function, but just a careful choice of the integrating functions $a_j(x)$. Also, notice that in this application we did not make explicit use of the assumption that the distribution is a Gaussian, since the same results will apply to *any* distribution function with mean μ and variance σ^2. Equation (5.13) can therefore be used in a variety of situations in which the distribution function has parameters that are related to the mean and variance, even if they are not identical to them, as in the case of the Gaussian. The method of moments therefore returns unbiased estimates for the mean and variance of every distribution in the case of a large number of measurements.

Example 5.4 Consider the five measurements presented in Example 4.9: 0.1, 0.3, 0.5, 0.7, and 0.9, and assume that they are known to be drawn from a uniform distribution between 0 and a. The method of moments can be used to estimate the parameter a of the distribution from the measurements. The probability distribution function is $f(x) = 1/a$ between 0 and a, and null otherwise. Using the integrating function $a_1(x) = x$, the method of moments proceeds with the calculation of the first moment of the distribution,

$$E[X] = \int_0^a xf(x)dx = \frac{a}{2}$$

Therefore, using (5.12), we can estimate the only parameter a of the distribution function as

$$a = 2 \times \frac{1}{N}\sum_{i=1}^{N} x_i$$

where $N = 5$, for the result of $a = 1$. The result confirms that the five measurements are compatible with a parent mean of $1/2$. ◇

5.4 Quantiles and Confidence Intervals

The parameters of the distribution function can be used to determine the range of values that include a given probability, for example, 68.3 %, or 90 %, or 99 %, etc. This range, called *confidence interval*, can be conveniently described by the cumulative distribution function $F(x)$.

Define the α-quantile x_α, where α is a number between 0 and 1, as the value of the variable such that $x \leq x_\alpha$ with probability α:

$$\alpha \text{ quantile } x_\alpha : P(x \leq x_\alpha = \alpha) \text{ or } F(x_\alpha) = \alpha. \tag{5.14}$$

For example, consider the cumulative distribution shown in Fig. 5.1 (right panel): the $\alpha = 0.05$ quantile is the number of the variable x where the lower horizontal dashed line intersects the cumulative distribution $F(x)$, $x_\alpha \simeq 0.2$, and the $\beta = 0.95$ quantile is the number of the variable x where the upper dashed line intersects $F(x)$, $x_\beta \simeq 6$. Therefore the range x_α to x_β, or 0.2–6, corresponds to the $(\beta - \alpha) = 90\,\%$ confidence interval, i.e., there is 90 % probability that a measurement of the variable falls in that range. These confidence intervals are called *central* because they are centered at the mean (or median) of the distribution, and are the most commonly used type of confidence intervals.

Confidence intervals can be constructed at any confidence level desired, depending on applications and on the value of probability that the analyzer wishes to include in that interval. It is common to use 68 % confidence intervals because this is the probability between $\pm\sigma$ of the mean for a Gaussian variable (see Sect. 5.4.1). Normally a confidence interval or limit at a significance lower than 68 % is not considered interesting, since there is a significant probability that the random variable will be outside of this range.

One-sided confidence intervals that extends down to $-\infty$, or to the lowest value allowed for that random variable, is called an *upper limit*, and intervals that extend

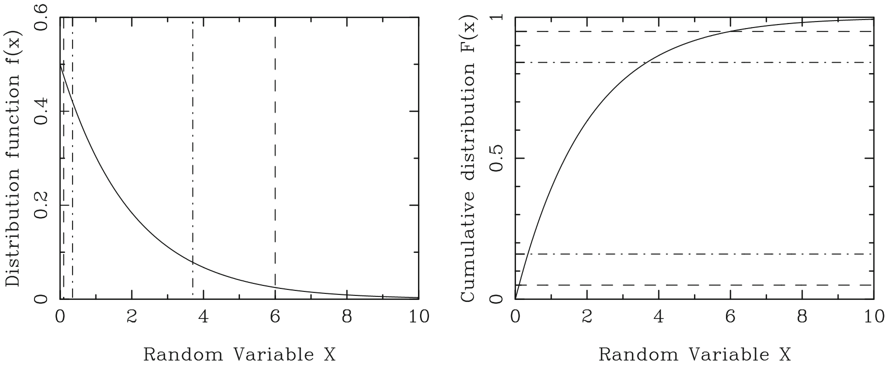

Fig. 5.1 (*Left*) Distribution function of an exponential variable with central 68 and 90 % confidence interval marked by, respectively, *dot-dashed* and *dotted lines*. (*Right*) The confidence interval are obtained as the intersection of *dot-dashed* and *dotted lines* with the cumulative distribution (*solid line*)

to $+\infty$, or to the highest allowed value, is called a *lower limit*. A lower limit describes a situation in which a large number is detected, for example counts from a Poisson experiment, and we want to describe how small the value of the variable can be, and still be consistent with the data. An upper limit is used for a situation in which a small number is detected, to describe how high can the variable be and still be consistent with the data. Lower and upper limits depend on the value of the probability that we want to use; for example, using a value for α that is closer to 0 results in a lower limit that progressively becomes $\lambda_{lo} = -\infty$ (or lowest allowed value), which is not a very interesting statement. If β is progressively closer to 1, the upper limit will tend to $\lambda_{up} = \infty$.

5.4.1 Confidence Intervals for a Gaussian Variable

When the variable is described by a Gaussian distribution function we can use integral tables (Table A.2) to determine confidence intervals that enclose a given probability. It is usually meaningful to have central confidence intervals, i.e., intervals centered at the mean of the distribution and extending by equal amounts on either side of the mean. For central confidence intervals, the relationship between the probability p enclosed by a given interval (say $p = 0.9$ or 90 % confidence interval) and the size $\Delta x = 2(z \times \sigma)$ of the interval is given by

$$p = \int_{\mu - z\sigma}^{\mu + z\sigma} f(x)dx, \tag{5.15}$$

where $f(x)$ is a Gaussian of mean μ and variance σ^2. The number z represents the number of standard deviations allowed by the interval in each direction (positive and negative relative to the mean). The most common central confidence intervals for a Gaussian distribution are reported in Table 5.1. For example, for a mean μ and variance σ^2 and the interval from $\mu - \sigma$ to $\mu + \sigma$ is a 68.3 % confidence interval, the interval from $\mu - 1.65\sigma$ to $\mu + 1.65\sigma$ is a 90 % confidence interval. In principle, one could have confidence intervals that are not centered on the mean of the distribution—such intervals would still be valid confidence intervals. It can be shown that central confidence intervals are the smallest possible, for a given confidence level.

Example 5.5 Using the data for the J.J. Thomson experiment on the measurement of the electron's mass-to-charge ratio, we can calculate the 90 % confidence interval on m/e for Tube 1 and Tube 2. For Tube 1, we estimated the mean as $\mu_1 = 0.42$ and the standard error as $\sigma_1 = 0.07$, and for Tube 2 $\mu_2 = 0.53$ and $\sigma = 0.08$. Since the random variable is assumed to be Gaussian, the 90 % confidence interval corresponds to the range between $\mu - 1.65\sigma$ and $\mu + 1.65\sigma$; therefore for the Thomson measurements of Tube 1 and Tube 2, the 90 % central confidence intervals are, respectively, 0.30–0.54 and 0.40–0.66. ◇

Upper and lower limits can be easily calculated using the estimates of μ and σ for a Gaussian variable. They are obtained numerically from the following relationships,

$$
\begin{aligned}
p &= \int_{-\infty}^{\lambda_{up}} f(x)dx = F(\lambda_{up}) && \text{upper limit } \lambda up \\
p &= \int_{\lambda_{lo}}^{\infty} f(x)dx = 1 - F(\lambda_{lo}) && \text{lower limit } \lambda_{lo}
\end{aligned}
\tag{5.16}
$$

making use of Tables A.2 and A.3. The quantites $F(\lambda_{up})$ and $F(\lambda_{lo})$ are the values of the cumulative distribution of the Gaussian, showing that λ_{up} is the p-quantile and λ_{lo} is the $(1-p)$-quantile of the distribution. Useful upper and lower limits for the Gaussian distribution are reported in Table 5.2.Upper limits are typically of interest when the measurements result in a low value of the mean of the variable. In this case we usually want to know how high the variable can be and still be

Table 5.1 Common confidence intervals for a Gaussian distribution

Interval	Range	Enclosed probability (%)
50 % confidence interval	$\mu - 0.68\sigma, \mu + 0.68\sigma$	50
1-σ interval	$\mu - \sigma, \mu + \sigma$	68.3
90 % confidence interval	$\mu - 1.65\sigma, \mu + 1.65\sigma$	90
2-σ interval	$\mu - 2\sigma, \mu + 2\sigma$	95.5
3-σ interval	$\mu - 3\sigma, \mu + 3\sigma$	99.7

Table 5.2 Common upper and lower limits for a Gaussian distribution

Upper limit	Range	Enclosed probability (%)	Lower limit	Range	Enclosed probability (%)
50 % confidence	$\leq \mu$	50	50 % confidence	$\geq \mu$	50
90 % confidence	$\leq \mu + 1.28\sigma$	90	90 % confidence	$\geq \mu - 1.28\sigma$	90
95 % confidence	$\leq \mu + 1.65\sigma$	95	95 % confidence	$\geq \mu - 1.65\sigma$	95
99 % confidence	$\leq \mu + 2.33\sigma$	99	99 % confidence	$\geq \mu - 2.33\sigma$	99
1-σ	$\leq \mu + \sigma$	84.1	1-σ	$\geq \mu - \sigma$	84.1
2-σ	$\leq \mu + 2\sigma$	97.7	2-σ	$\geq \mu - 2\sigma$	97.7
3-σ	$\leq \mu + 3\sigma$	99.9	3-σ	$\geq \mu - 3\sigma$	99.9

consistent with the measurement, at a given confidence level. For example, in the case of the measurement of Tube 1 for the Thomson experiment, the variable m/e was measured to be 0.42 ± 0.07. In this case, it is interesting to ask the question of how high can m/e be and still be consistent with the measurement at a given confidence level.

Example 5.6 Using the data for the J.J. Thomson experiment on the measurement of the electron's mass-to-charge ratio, we can calculate the 90 % upper limits to m/e for Tube 1 and Tube 2. For Tube 1, we estimated the mean as $\mu_1 = 0.42$ and the standard error as $\sigma_1 = 0.07$, and for Tube 2 $\mu_2 = 0.53$ and $\sigma = 0.08$.

To determine the upper limit $m/e_{UL,90}$ of the ratio, we calculate the probability of occurrence of $m/e \leq m/e_{UL,90}$:

$$P(m/e \leq m/e_{UL,90}) = 0.90$$

Since the random variable is assumed to be Gaussian, the value $x \simeq \mu + 1.28\sigma$ corresponds to the 90 percentile of the distribution (see Table 5.2). The two 90 % upper limits are, respectively, 0.51 and 0.63. ◇

A common application of upper limits is when an experiment has failed to detect the variable of interest. In this case we have a *non-detection* and we want to place upper limits based on the measurements made. This problem is addressed by considering the parent distribution of the variable that we did not detect, for example a Gaussian of zero mean and given variance. We determine the upper limit as the value of the variable that exceeds the mean by 1, 2 or 3 σ, corresponding to the probability levels shown in Table 5.2. A 3-σ upper limit, for example, is the value of the variable that has only a 0.1 % chance of being observed based on the parent distribution for the non-detection, and therefore we are 99.9 % confident that the true value of the variable is lower than this upper limit.

Example 5.7 (Gaussian Upper Limit to the Non-detection of a Source) A measurement of $n = 8$ counts in a given time interval is made in the presence of a source of unknown intensity. The instrument used for the measurement has a background level with a mean of 9.8 ± 0.4 counts, as estimated from an independent experiment of long duration. Given that the measurement is below the expected background level, it is evident that there is no positive detection of the source. The hypothesis that the source has zero emission can be described by a distribution function with a mean of approximately 9.8 counts and, since this is a counting experiment, the probability distribution of counts should be Poisson. We are willing to approximate the distribution with a Gaussian of same mean, and variance equal to the mean, or $\sigma \simeq 3.1$, to describe the distribution of counts one expects from an experiment of the given duration as the one that yielded $n = 8$ counts.

A 99 % upper limit to the number of counts that can be recorded by this instrument, in the given time interval, can be calculated according to Table 5.2 as

$$\mu + 2.33\sigma = 9.8 + 2.33 \times 3.1 \simeq 17.$$

This means that we are 99 % confident that the true value of the source plus background counts is less than 17. A complementary way to interpret this number is that the experimenter can be 99 % sure that the measurement *cannot* be due to just the background if there is a detection of $\geq$17 total counts. A conservative analyst might also want to include the possibility that the Gaussian distribution has a slightly higher mean, since the level of the background is not known exactly, and conservatively assume that perhaps 18 counts are required to establish that the source does have a positive level of emission. After subtraction of the assumed background level, we can conclude that the 99 % upper limit to the source's true emission level in the time interval is 8.2 counts. This example was adapted from the analysis of an astronomical source that resulted in a non-detection [4]. ◇

5.4.2 Confidence Intervals for the Mean of a Poisson Variable

The Poisson distribution does not have the simple analytical properties of the Gaussian distribution. For this distribution it is convenient to follow a different method to determine its confidence intervals.

Consider the case of a single measurement of a Poisson variable of unknown mean λ for which n_{obs} was recorded. We want to make inferences on the parent mean based on this information. Also, we assume that the measurement includes a uniform and known background λ_B. The measurement is therefore drawn from a random variable

$$X = N_S + N_B \tag{5.17}$$

in which $N_B = \lambda_B$ is assumed to be a constant, i.e., the background is known exactly (this generalization can be bypassed by simply setting $\lambda_B = 0$). The probability distribution function of X (the total source plus background counts) is

$$f(n) = \frac{(\lambda + \lambda_B)^n}{n!} e^{-(\lambda+\lambda_B)}, \tag{5.18}$$

where n is an integer number describing possible values of X. Equation (5.18) is true even if the background is not known exactly, since the sum of two Poisson variables is also a Poisson variable with mean equal to the sum of the means. It is evident that, given the only measurement available, the estimate of the source mean is

$$\hat{\lambda} = n_{obs} - \lambda_B.$$

This estimate is the starting point to determine a confidence interval for the parent mean. We define the lower limit λ_{lo} as the value of the source mean that results in the observation of $n \geq n_{obs}$ with a probability α:

$$\alpha = \sum_{n=n_{obs}}^{\infty} \frac{(\lambda_{lo} + \lambda_B)^n}{n!} e^{-(\lambda_{lo}+\nu_B)} = 1 - \sum_{n=0}^{n_{obs}-1} \frac{(\lambda_{lo} + \lambda_B)^n}{n!} e^{-(\lambda_{lo}+\lambda_B)}. \tag{5.19}$$

The mean λ_{lo} corresponds to the situation shown in the left panel of Fig. 5.2: assuming that the actual mean is as low as λ_{lo}, there is only a small probability α (say 5 %) to make a measurement above or equal to what was actually measured. Thus, we can say that there is only a very small chance (α) that the actual mean could have been as low (or lower) than λ_{lo}. The quantity λ_{lo} is the *lower limit with confidence* $(1-\alpha)$, i.e., we are $(1-\alpha)$, say 95 %, confident that the mean is higher than this value.

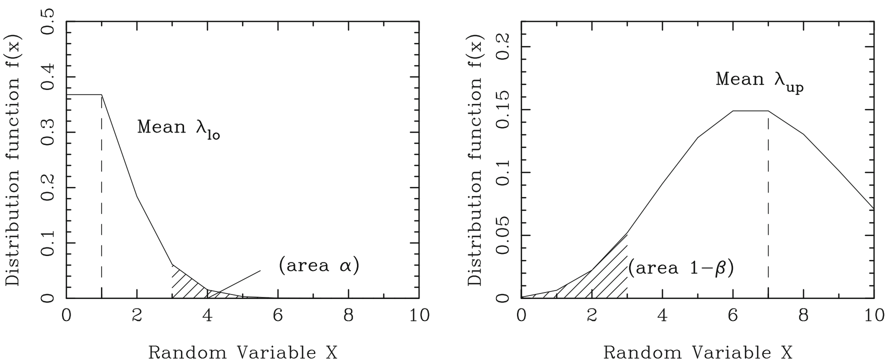

Fig. 5.2 This illustration of the upper and lower limits to the measurement of a Poisson mean assumes a measurement of $n_{obs} = 3$. On the *left*, the lower limit to the parent mean is such that there is a probability of α to measure n_{obs} or higher (*hatched area*); on the *right*, the upper limit leaves a probability of $(1-\beta)$ that a measurement is n_{obs} or lower

By the same logic we also define λ_{up} as the parent value of the source mean that results in the observation of $n \leq n_{obs}$ with a probability $(1-\beta)$, or

$$1-\beta = \sum_{n=0}^{n_{obs}} \frac{(\lambda_{up}+\lambda_B)^n}{n!} e^{-(\lambda_{up}+\lambda_B)}. \tag{5.20}$$

This is illustrated in the right panel of Fig. 5.2, where the number $(1-\beta)$ is intended as a small number, of same magnitude as α. Assuming that the mean is as high as λ_{up}, there is a small probability of $1-\beta$ to make a measurement equal or lower than the actual measurement. Therefore we say that there is only a small probability that the true mean could be as high or higher than λ_{up}. The number λ_{up} is the *upper limit with confidence* β, that is, we are β (say 95 %) confident that the mean is lower than this value.

If we combine the two limits, the probability that the true mean is above λ_{up} or below λ_{lo} is just $(1-\beta)+\alpha$, say 10 %, and therefore the interval λ_{lo} to λ_{up} includes a probability of

$$P(\lambda_{lo} \leq \lambda \leq \lambda_{up}) = 1-(1-\beta)-\alpha = \beta-\alpha,$$

i.e., this is a $(\beta-\alpha)$, say 90 %, confidence interval.

The upper and lower limits defined by (5.19) and (5.20) can be approximated analytically using a relationship that relates the Poisson sum with an analytic distribution function:

$$\sum_{x=0}^{n_{obs}-1} \frac{e^{-\lambda}\lambda^x}{x!} = 1 - P_{\chi^2}(\chi^2, \nu) \tag{5.21}$$

where $P_{\chi^2}(\chi^2, \nu)$ is the cumulative distribution of the χ^2 probability distribution function defined in Sect. 7.2, with parameters $\chi^2 = 2\lambda$ and $\nu = 2n_{obs}$,

$$P_{\chi^2}(\chi^2, \nu) = \int_{-\infty}^{\chi^2} f_{\chi^2}(x, \nu)dx.$$

The approximation is due to Gehrels [16], and makes use of mathematical relationships that can be found in the handbook of Abramowitz and Stegun [1]. The result is that the upper and lower limits can be simply approximated once we specify the number of counts n_{obs} and the probability level of the upper or lower limit. The probability level is described by the number S, which is the equivalent number of Gaussian σ that corresponds to the confidence level chosen (for example, 84 %

confidence interval corresponds to $S = 1$, etc. see Table 5.3)

$$\begin{cases} \lambda_{up} = n_{obs} + \dfrac{S^2+3}{4} + S\sqrt{n_{obs} + \dfrac{3}{4}} \\ \lambda_{lo} = n_{obs}\left(1 - \dfrac{1}{9n_{obs}} - \dfrac{S}{3\sqrt{n_{obs}}}\right)^3. \end{cases} \tag{5.22}$$

The S parameter is also a quantile of a standard Gaussian distribution, enclosing a probability as illustrated in Table 5.3.

Proof Use of (5.21) into (5.19) and (5.20) gives a relationship between the function P_{χ^2} and the probability levels α and β,

$$\begin{cases} P_{\chi^2}(2\lambda_{lo}, 2n_{obs}) = \alpha \\ P_{\chi^2}(2\lambda_{up}, 2n_{obs}+2) = \beta. \end{cases} \tag{5.23}$$

We use the simplest approximation for the function P_{χ^2} described in [16], one that is guaranteed to give limits that are accurate within 10 % of the true values. The approximation makes use of the following definitions: for any probability $a \leq 1$, y_a is the a-quantile of a standard normal distribution, or $G(y_a) = a$,

$$G(y_a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_a} e^{-t^2/2} dt.$$

If $P_{\chi^2}(\chi^2_a, \nu) = a$, then the simplest approximation between χ^2_a and y_a given by Gehrels [16] is

$$\chi^2_a \simeq \frac{1}{2}\left(y_a + \sqrt{2\nu - 1}\right)^2. \tag{5.24}$$

Consider the upper limit in (5.23). We can solve for λ_{up} by using (5.24) with $2\lambda_{up} = \chi^2_a$, $\nu = 2n_{obs} + 2$ and $S = y_a$, since y_a is the a-quantile of a standard

Table 5.3 Poisson parameters S and corresponding probabilities

Upper or lower limit	Range	Probability (%)	Poisson S parameter
90 % confidence	$\leq \mu + 1.28\sigma$	90	1.28
95 % confidence	$\leq \mu + 1.65\sigma$	95	1.65
99 % confidence	$\leq \mu + 2.33\sigma$	99	2.33
1-σ	$\leq \mu + \sigma$	84.1	1.0
2-σ	$\leq \mu + 2\sigma$	97.7	2.0
3-σ	$\leq \mu + 3\sigma$	99.9	3.0

normal distribution, thus equivalent to S. It follows that

$$2\lambda_{up} = \frac{1}{2}\left(S + \sqrt{4n_{obs} + 3}\right)^2$$

and from this the top part of (5.22) after a simple algebraic manipulation. A similar result applies for the lower limit. □

Equation (5.22) is tabulated in Tables A.5 and A.6 for several interesting values of n_{obs} and S. A few cases of common use are also shown in Table 5.4.

Example 5.8 (Poisson Upper Limit to Non-detection with No Background) An interesting situation that can be solved analytically is that corresponding to the situation in which there was a complete *non-detection* of a source, $n_{obs} = 0$. Naturally, it is not meaningful to look for a lower limit to the Poisson mean, but it is quite interesting to solve (5.20) in search for an upper limit with a given confidence β. In this case of $n = 0$ the equation simplifies to

$$1 - \beta = e^{-(\lambda_{up}+\lambda_B)} \Rightarrow \lambda_{up} = -\lambda_B - \ln\beta.$$

For $\beta = 0.84$ and zero background ($\lambda_B = 0$) this corresponds to an upper limit of $\lambda_{up} = -\ln 0.16 = 1.83$. This example can also be used to test the accuracy of the approximation given by (5.22). Using $n_{obs} = 0$, we obtain

$$\lambda_{up} = \frac{1}{4}(1 + \sqrt{3})^2 = 1.87$$

which is in fact just 2 % higher than the exact result. An example of upper limits in the presence of a non-zero background is presented in Problem 5.8. ◇

Table 5.4 Selected Upper and Lower limits for a Poisson variable using the Gehrels approximation (see Tables A.5 and A.6 for a complete list of values)

	Poisson parameter S or confidence level		
	$S = 1$	$S = 2$	$S = 3$
n_{obs}	(1-σ, or 84.1 %)	(2-σ, or 97.7 %)	(3-σ, or 99.9 %)
Upper limit			
0	1.87	3.48	5.60
1	3.32	5.40	7.97
…			
Lower limit			
…			
9	6.06	4.04	2.52
10	6.90	4.71	3.04
…			

5.5 Bayesian Methods for the Poisson Mean

The Bayesian method consists of determining the *posterior probability* $P(\lambda/obs)$, having calculated the likelihood as

$$P(n_{obs}/\lambda) = \frac{(\lambda + \lambda_B)^{n_{obs}}}{n_{obs}!} e^{-(\lambda+\lambda_B)}. \tag{5.25}$$

We use Bayes' theorem,

$$P(\lambda/obs) = \frac{P(n_{obs}/\lambda)\pi(\lambda)}{\int_0^\infty P(n_{obs}/\lambda')\pi(\lambda'), d\lambda'} \tag{5.26}$$

in which we needed to introduce a *prior probability* distribution $\pi(\lambda)$ in order to calculate the posterior probability. The use of a prior distribution is what constitutes the Bayesian approach. The simplest assumption is that of a uniform prior, $\pi(\lambda) = C$, over an arbitrarily large range of $\lambda \geq 0$, but other choices are possible according to the information available on the Poisson mean prior to the measurements. In this section we derive the Bayesian expectation of the Poisson mean and upper and lower limits and describe the differences with the classical method.

5.5.1 Bayesian Expectation of the Poisson Mean

The posterior distribution of the Poisson mean λ (5.26) can be used to calculate the Bayesian expectation for the mean can be calculated as the integral of (5.26) over the entire range allowed to the mean,

$$E[\lambda/obs] = \frac{\int_0^\infty P(\lambda/obs)\lambda d\lambda}{\int_0^\infty P(\lambda/obs) d\lambda}. \tag{5.27}$$

The answer will in general depend on the choice of the prior distribution $\pi(\lambda)$. Assuming a constant prior, the expectation becomes

$$E[\lambda/obs] = \int_0^\infty \frac{e^{-\lambda}\lambda^{n_{obs}+1}}{n!} d\lambda = n_{obs} + 1, \tag{5.28}$$

where we made use of the integral

$$\int_0^\infty e^{-\lambda}\lambda^n d\lambda = n!$$

The interesting result is therefore that a measurement of n_{obs} counts implies a Bayesian expectation of $E[\lambda] = n_{obs} + 1$, i.e., one count more than the observation. Therefore even a non-detection results in an expectation for the mean of the parent distribution of 1, and not 0. This somewhat surprising result can be understood by considering the fact that even a parent mean of 1 results in a likelihood of $1/e$ (i.e., a relatively large number) of obtaining zero counts as a result of a random fluctuations. Moreover, the Poisson distribution is skewed, with a heavier tail at large values of λ. This calculation is due to Emslie [12].

5.5.2 Bayesian Upper and Lower Limits for a Poisson Variable

Using a uniform prior, we use the Bayesian approach (5.26) to calculate the upper limit to the source mean with confidence β (say, 95 %). This is obtained by integrating (5.26) from the lower limit of 0 to the upper limit λ_{up},

$$\beta = \frac{\int_0^{\lambda_{up}} P(n_{obs}/\lambda)d\lambda}{\int_0^{\infty} P(n_{obs}/\lambda)d\lambda} = \frac{\int_0^{\lambda_{up}} (\lambda + \lambda_B)^{n_{obs}} e^{-(\lambda+\lambda_B)} d\lambda}{\int_0^{\infty} (\lambda + \lambda_B)^{n_{obs}} e^{-(\lambda+\lambda_B)} d\lambda} \tag{5.29}$$

Similarly, the lower limit can be estimated according to

$$\alpha = \frac{\int_0^{\lambda_{lo}} P(n_{obs}/\lambda)d\lambda_S}{\int_0^{\infty} P(n_{obs}/\lambda)d\lambda} = \frac{\int_0^{\lambda_{lo}} (\lambda + \lambda_B)^{n_{obs}} e^{-(\lambda+\lambda_B)} d\lambda}{\int_0^{\infty} (\lambda + \lambda_B)^{n_{obs}} e^{-(\lambda+\lambda_B)} d\lambda} \tag{5.30}$$

where α is a small probability, say 5 %. Since n_{obs} is always an integer, these integrals can be evaluated analytically.

The difference between the classical upper limits described by (5.19) and (5.20) and the Bayesian limits of (5.29) and (5.30) is summarized by the different variable of integration (or summation) in the relevant equations. For the classical limits we use the Poisson probability to make n_{obs} measurements for a true mean of λ. We then estimate the upper or lower limits as the values of the mean that gives a probability of, respectively, $1-\beta$ and α, to observe $n \leq n_{obs}$ events. In this case, the probability is evaluated as a sum over the number of counts, for a fixed value of the parent mean.

In the case of the Bayesian limits, on the other hand, we first calculate the posterior distribution of λ, and then require that the range between 0 and the limits λ_{up} and λ_{lo} includes, respectively, a $1-\beta$ and α probability, evaluated as an integral over the mean for a fixed value of the detected counts. In general, the two methods will give different results.

Example 5.9 (Bayesian Upper Limit to a Non-detection) The case of non-detection, $n_{obs} = 0$, is especially simple and interesting, since the background drops out of the equation, resulting in $\beta = 1 - e^{-\lambda_{up}}$, which gives

$$\lambda_{up} = -\ln(1-\beta) \text{ (case of } n_{obs} = 0 \text{, Bayesian upper limit)} \tag{5.31}$$

The Bayesian upper limit is therefore equal to the classical limit, when there is no background. When there is background, the two estimate will differ.[1] ◇

Summary of Key Concepts for this Chapter

- ☐ *Maximum Likelihood (ML) method*: A method to estimate parameters of a distribution under the assumption that the best-fit parameters maximize the likelihood of the measurements.
- ☐ *ML estimates of mean and variance*: For a Gaussian variable, the unbiased ML estimates are

$$\begin{cases} \mu_{ML} = \overline{x} \\ s^2 = \dfrac{1}{N-1} \sum (x_i - \overline{x})^2 \end{cases}$$

- ☐ *Estimates of mean with non-uniform uncertainties*: They are given by

$$\begin{cases} \mu_{ML} = \dfrac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2} \\ \sigma_\mu^2 = \dfrac{1}{\sum 1/\sigma_i^2} \end{cases}$$

- ☐ *Confidence intervals*: Range of the variable that contains a given probability of occurrence (e.g., $\pm 1\sigma$ range contains 68 % of probability for a Gaussian variable).
- ☐ *Upper and lower limits*: An upper (lower) limit is the value below (above) which there is a given probability (e.g., 90 %) to observe the variable.

Problems

5.1 Using the definition of weighted sample mean as in (5.8), derive its variance and show that it is given by (5.9).

5.2 Using the data from Mendel's experiment (Table 1.1), calculate the standard deviation in the measurement of each of the seven fractions of dominants, and the weighted mean and standard deviation of the seven fractions.

Compare your result from a direct calculation of the overall fraction of dominants, obtained by grouping all dominants from the seven experiments together.

[1] Additional considerations on the measurements of the mean of a Poisson variable, and the case of upper and lower limits, can be found in [10].

5.3 The Mendel experiment of Table 1.1 can be described as n number n of measurements of n_i, the number of plants that display the dominant character, out of a total of N_i plants. The experiment is described by a binomial distribution with probability $p = 0.75$ for the plant to display the dominant character.

Using the properties of the binomial distribution, show analytically that the weighted average of the measurements of the fraction $f_i = n_i/N_i$ is equal to the value calculated directly as

$$\mu = \frac{\sum_{i=1}^{n} n_i}{\sum_{i=1}^{n} N_i}$$

5.4 Consider a decaying radioactive source observed in a time interval of duration $T = 15$ s; N is the number of total counts, and B is the number of background counts (assumed to be measured independently of the total counts):

$$\begin{cases} N = 19 & \text{counts} \\ B = 14 & \text{counts} \end{cases}.$$

The goal is to determine the probability of detection of source counts $S = N - B$ in the time interval T.

(a) Calculate this probability directly via:

$$\text{Prob(detection)} = \text{Prob}(S > 0/\text{data})$$

in which S is treated as a random variable, with Gaussian distribution of mean and variance calculated according to the error propagation formulas. Justify why the Gaussian approximation *may* be appropriate for the variable S.

(b) Use the same method as in (a), but assuming that the background B is known without error (e.g., as if it was observed for such along time interval that its error becomes negligible).

(c) Assume that the background is a variable with mean of 14 counts in a 15 s interval, and that it can be observed for an interval of time $T \gg 15$ s. Find what interval of time T makes the error σ_{B15} of the background over a time interval of 15-s have a value $\sigma_{B15}/B_{15} = 0.01$, e.g., negligible.

5.5 For the Thomson experiment of Table 2.1 (tube 1) and Table 2.2 (tube 2), calculate:

(a) The 90 % central confidence intervals for the variable v;

(b) The 90 % upper and lower limits, assuming that the variable is Gaussian.

5.6 Consider a Poisson variable X of mean μ.

(a) We want to set 90 % confidence upper limits to the value of the parent mean λ, assuming that one measurement of the variable yielded the result of $N = 1$. Following the classical approach, find the equation that determines the exact

90 % upper limit to the mean λ_{up}. Recall that the classical 90 % confidence upper limit is defined as the value of the Poisson mean that yields a $P(X \leq N) = \beta$, where $1 - \beta = 0.9$.

(b) Using the Bayesian approach, which consists of defining the $1 - \beta = 0.9$ upper limit via

$$\beta = \frac{\int_0^{\lambda_{up}} P(n_{obs}/\mu)d\mu}{\int_0^{\infty} P(n_{obs}/\mu)d\mu} \tag{5.32}$$

where $n_{obs} = N$; find the equation that determines the 90 % upper limit to the mean λ_{up}.

5.7 The data provided in Table 2.3 from Pearson's experiment on biometric data describes the cumulative distribution function of heights from a sample of 1,079 couples. Calculate the 2σ upper limit to the fraction of couples in which both mother and father are taller than 68 in.

5.8 Use the data presented in Example 5.7, in which there is a non-detection of a source in the presence of a background of $\lambda_B \simeq 9.8$. Determine the Poisson upper limit to the source count at the 99 % confidence level and compare this upper limit with that obtained in the case of a zero background level.