Chapter 4
Functions of Random Variables and Error Propagation

Abstract Sometimes experiments do not directly measure the quantity of interest, but rather associated variables that can be related to the one of interest by an analytic function. It is therefore necessary to establish how we can infer properties of the interesting variable based on properties of the variables that have been measured directly. This chapter explains how to determine the probability distribution function of a variable that is function of other variables of known distribution, and how to measure its mean and variance, the latter usually referred to as *error propagation* formulas. We also establish two fundamental results of the theory of probability, the central limit theorem and the law of large numbers.

4.1 Linear Combination of Random Variables

Experimental variables are often related by a simple linear relationship. The linear combination of N random variables X_i is a variable Y defined by

$$Y = \sum_{i=1}^{N} a_i X_i \tag{4.1}$$

where a_i are constant coefficients. A typical example of a variable that is a linear combination of two variables is the signal detected by an instrument, which can be thought of as the sum of the intrinsic signal from the source plus the background. The distributions of the background and the source signals will influence the properties of the total signal detected, and it is therefore important to understand the statistical properties of this relationship in order to characterize the signal from the source.

4.1.1 General Mean and Variance Formulas

The expectation or mean of the linear combination is $E[Y] = \sum_{i=1}^{N} a_i E[X_i]$ or

$$\mu_y = \sum_{i=1}^{N} a_i \mu_i, \tag{4.2}$$

M. Bonamente, *Statistics and Analysis of Scientific Data*, Graduate Texts in Physics, DOI 10.1007/978-1-4939-6572-4_4

where μ_i is the mean of X_i. This property follows from the linearity of the expectation operator, and it is equivalent to a weighted mean in which the weights are given by the coefficients a_i.

In the case of the variance, the situation is more complex:

$$
\begin{aligned}
Var[Y] = E\left[\left(\sum_{i=1}^{N} a_i X_i - \sum_{i=1}^{N} a_i \mu_i\right)^2\right] &= \sum_{i=1}^{N} a_i^2 E\left[(X_i-\mu_i)^2\right] \\
&+ 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} a_i a_j E[(X_i-\mu_i)(X_j-\mu_j)] \\
&= \sum_{i=1}^{N} a_i^2 Var(X_i) + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} a_i a_j Cov(X_i, X_j).
\end{aligned}
$$

The result can be summarized in a more compact relationship,

$$\sigma_y^2 = \sum_{i=1}^{N} a_i^2 \sigma_i^2 + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} a_i a_j \sigma_{ij}^2. \tag{4.3}$$

Equation (4.3) shows that variances add only for variables that are mutually uncorrelated, or $\sigma_{ij}^2 = 0$, but not in general. The following example illustrates the importance of a non-zero covariance between two variables, and its effect on the variance of the sum.

Example 4.1 (Variance of Anti-correlated Variables) Consider the case of the measurement of two random variables X and Y that are completely anti-correlated, $Corr(X, Y) = -1$, with mean and variance $\mu_x = 1$, $\mu_y = 1$, $\sigma_x^2 = 0.5$ and $\sigma_y^2 = 0.5$.

The mean of $Z = X + Y$ is $\mu = 1 + 1 = 2$ and the variance is $\sigma^2 = \sigma_x^2 + \sigma_y^2 - 2Cov(X, Y) = (\sigma_x - \sigma_y)^2 = 0$; this means that in this extreme case of complete anticorrelation the sum of the two random variables is actually not a random variable any more. If the covariance term had been neglected in (4.3), we would have made the error of inferring a variance of 1 for the sum. ◇

4.1.2 Uncorrelated Variables and the $1/\sqrt{N}$ Factor

For two or more uncorrelated variables the variances add linearly, according to (4.3). Uncorrelated variables are common in statistics. For example, consider repeating the same experiment a number N of times independently, and each time measurements of a random variable X_i is made. After N experiments, one obtains N measurements from identically distributed random variables (since they resulted from the same type of experiment). The variables are independent, and therefore uncorrelated, if the experiments were performed in such a way that the outcome of one specific experiment did not affect the outcome of another.

With N uncorrelated variables X_i all of equal mean μ and variance σ^2, one is often interested in calculating the *relative uncertainty* in the variable

$$Y = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{4.4}$$

which describes the sample mean of N measurements. The relative uncertainty is described by the ratio of the standard deviation and the mean,

$$\frac{\sigma_y}{\mu_y} = \frac{1}{N}\frac{\sqrt{\sigma^2 + \cdots + \sigma^2}}{\mu} = \frac{1}{\sqrt{N}} \times \frac{\sigma}{\mu} \tag{4.5}$$

where we used the property that $Var[aX] = a^2 Var[X]$ and the fact that both means and variances add linearly. The result shows that the N measurements reduced the relative error in the random variable by a factor of $1/\sqrt{N}$, as compared with a single measurement. This observation is a key factor in statistics, and it is the reason why one needs to repeat the same experiment many times in order to reduce the relative statistical error. Equation (4.5) can be recast to show that the variance in the sample mean is given by

$$\sigma_Y^2 = \frac{\sigma^2}{N} \tag{4.6}$$

where σ is the sample variance, or variance associated with one measurement. The interpretation is simple: one expects much less variance between two measurements of the sample mean, than between two individual measurements of the variable, since the statistical fluctuations of individual measurements average down with increasing sample size.

Another important observation is that, in the case of completely correlated variables, then additional measurements introduces no advantages, i.e., the relative error does not decrease with the number of measurements. This can be shown with the aid of (4.3), and is illustrated in the following example.

Example 4.2 (Variance of Correlated Variables) Consider the two measurements in Example 4.1, but now with a correlation of 1. In this case, the covariance of the sum is $\sigma^2 = \sigma_x^2 + \sigma_y^2 + 2Cov(X, Y) = (\sigma_x + \sigma_y)^2$, and therefore the relative error in the sum is

$$\frac{\sigma}{\mu} = \frac{(\sigma_x + \sigma_y)}{\mu_x + \mu_y}$$

which is the same as the relative error of each measurement. Notice that the same conclusion applies to the average of the two measurements, since the sum and the average differ only by a constant factor of 1/2. ◇

4.2 The Moment Generating Function

The mean and the variance provide only partial information on the random variable, and a full description would require the knowledge of all moments. The moment generating function is a convenient mathematical tool to determine the distribution function of random variables and its moments. It is also useful to prove the central limit theorem, one of the key results of statistics, since it establishes the Gaussian distribution as the normal distribution when a random variable is the sum of a large number of measurements.

The *moment generating function* of a random variable X is defined as

$$M(t) = E[e^{tX}], \tag{4.7}$$

and it has the property that all moments can be derived from it, provided they exist and are finite. Assuming a continuous random variable of probability distribution function $f(x)$, the moment generating function can be written as

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx =$$

$$\int_{-\infty}^{+\infty} \left(1 + \frac{tx}{1} + \frac{(tx)^2}{2!} + \ldots\right) f(x)dx = 1 + t\mu_1 + \frac{t^2}{2!}\mu_2 + \ldots$$

and therefore all moments can be obtained as partial derivatives,

$$\mu_r = \left. \frac{\partial^r M(t)}{\partial t^r} \right|_{t=0}. \tag{4.8}$$

The most important property of the moment generating function is that there is a one-to-one correspondence between the moment generating function and the probability distribution function, i.e., the moment generating function is a sufficient description of the random variable. Some distributions do not have a moment generating function, since some of their moments may be infinite, so in principle this method cannot be used for all distributions.

4.2.1 Properties of the Moment Generating Function

A full treatment of mathematical properties of the moment generating function can be found in textbooks on theory of probability, such as [38]. Two properties of the moment generating function will be useful in the determination of the distribution function of random variables:

- If $Y = a + bX$, where a, b are constants, the moment generating function of Y is

$$M_y(t) = e^{at} M_x(bt). \tag{4.9}$$

Proof This relationship can be proved by the use of the expectation operator, according to the definition of the moment generating function:

$$E[e^{tY}] = E[e^{t(a+bX)}] = E[e^{at}e^{btX}] = e^{at}M_x(bt).$$

□

- If X and Y are independent random variables, with $M_x(t)$ and $M_y(t)$ as moment generating functions, then the moment generating function of $Z = X + Y$ is

$$M_z(t) = M_x(t)M_y(t). \tag{4.10}$$

Proof The relationship is derived immediately by

$$E[e^{tZ}] = E[e^{t(X+Y)}] = M_x(t)M_y(t).$$

□

4.2.2 The Moment Generating Function of the Gaussian and Poisson Distribution

Important cases to study are the Gaussian distribution of mean μ and variance σ^2 and the Poisson distribution of mean μ.

- The moment generating function of the Gaussian is given by

$$M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}. \tag{4.11}$$

Proof Start with

$$M(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

The exponent can be written as

$$\begin{aligned} tx - \frac{1}{2}\frac{x^2+\mu^2-2x\mu}{\sigma^2} &= \frac{2\sigma^2 tx - x^2 - \mu^2 + 2x\mu}{2\sigma^2} \\ &= -\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + \frac{2\mu\sigma^2 t}{2\sigma^2} + \frac{\sigma^2 t}{2\sigma^2}\sigma^2 t. \end{aligned}$$

It follows that

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{\mu t} e^{\frac{\sigma^2 t^2}{2}} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\mu t} e^{\frac{\sigma^2 t^2}{2}} \sqrt{2\pi\sigma^2} = e^{\mu t + \dfrac{\sigma^2 t^2}{2}}. \end{aligned}$$

□

- The moment generating function of the Poisson distribution is given by

$$M(t) = e^{-\mu} e^{\mu e^t}. \tag{4.12}$$

Proof The moment generating function is obtained by

$$M(t) = E[e^{tN}] = \sum_{n=0}^{\infty} e^{nt} \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_{n=0}^{\infty} \frac{(\mu e^t)^n}{n!} = e^{-\mu} e^{\mu e^t}.$$

□

Example 4.3 (Sum of Poisson Variables) The moment generating function can be used to show that the sum of two independent Poisson random variables of mean λ and μ is a Poisson random variable with mean $\lambda+\mu$. In fact that mean of the Poisson appears at the exponent of the moment generating function, and property (4.10), can be used to prove this result. The fact that the mean of two independent Poisson distributions will add is not surprising, given that the Poisson distribution relates to the counting of discrete events. ◇

4.3 The Central Limit Theorem

The Central Limit Theorem is one of statistic's most important results, establishing that a variable obtained as the sum of a large number of independent variables has a Gaussian distribution. This result can be stated as:

Theorem 4.1 (Central Limit Theorem) *The sum of a large number of independent random variables is approximately distributed as a Gaussian. The mean of the distribution is the sum of the means of the variables and the variance of the distribution is the sum of the variances of the variables. This result holds regardless of the distribution of each individual variable.*

Proof Consider the variable Y as the sum of N variables X_i of mean μ_i and variance σ_i^2,

$$Y = \sum_{i=1}^{N} X_i, \tag{4.13}$$

with $M_i(t)$ the moment generating function of the random variable $(X_i - \mu_i)$. Since the random variables are independent, and independence is a stronger statement than uncorrelation, it follows that the mean of Y is $\mu = \sum \mu_i$, and that variances likewise add linearly, $\sigma^2 = \sum \sigma_i^2$. We want to calculate the moment generating function of the variable Z defined by

$$Z = \frac{Y - \mu}{\sigma} = \frac{1}{\sigma} \sum_{i=1}^{N} (X_i - \mu_i).$$

The variable Z has a mean of zero and unit variance. We want to show that Z can be approximated by a standard Gaussian. Using the properties of the moment generating function, the moment generating function of Z is

$$M(t) = \prod_{i=1}^{N} M_i(t/\sigma).$$

The moment generating function of each variable $(X_i - \mu_i)/\sigma$ is

$$M_i(t/\sigma) = 1 + \mu_{(x_i - \mu_i)} \frac{t}{\sigma} + \frac{\sigma_i^2}{2} \left(\frac{t}{\sigma}\right)^2 + \frac{\mu_{i,3}}{3!} \left(\frac{t}{\sigma}\right)^3 + \dots$$

where $\mu_{x_i - \mu_i} = 0$ is the mean of $X_i - \mu_i$. The quantities σ_i^2 and $\mu_{i,3}$ are, respectively, the central moments of the second and third order of X_i.

If a large number of random variables are used, $N \gg 1$, then σ^2 is large, as it is the sum of variances of the random variables, and we can ignore terms of order σ^{-3}. We therefore make the approximation

$$\ln M(t) = \sum \ln M_i\left(\frac{t}{\sigma}\right) =$$

$$\sum \ln\left(1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2\right) \simeq \sum \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 = \frac{1}{2}t^2.$$

This results in the approximation of the moment generating function of $(y - \mu)/\sigma$ as

$$\Rightarrow M(t) \simeq e^{\frac{t^2}{2}},$$

which shows that Z is approximately distributed as a standard Gaussian distribution, according to (4.11). Given that the random variable of interest Y is obtained by a change of variable $Z = (Y - \mu)/\sigma$, we also know that $\mu_y = \mu$ and $Var(Y) = Var(\sigma Z) = \sigma^2 Var(Z) = \sigma^2$, therefore Y is distributed as a Gaussian with mean μ and variance σ^2. □

The central limit theorem establishes that the Gaussian distribution is the limiting distribution approached by the sum of random variables, no matter their original shapes, when the number of variables is large. A particularly illustrative example is the one presented in the following, in which we perform the sum of a number of uniform distributions. Although the uniform distribution does not display the Gaussian-like feature of a centrally peaked distribution, with the increasing number of variables being summed, the sum rapidly approaches a Gaussian distribution.

Example 4.4 (Sum of Uniform Random Variables) We show that the sum of N independent uniform random variables between 0 and 1 tend to a Gaussian with mean $N/2$, given that each variable has a mean of $1/2$. The calculation that the sum of N uniform distribution tends to the Gaussian can be done by first calculating the moment generating function of the uniform distribution, then using the properties of the moment generating function.

We can show that the uniform distribution in the range [0, 1] has $\mu_i = 1/2$, $\sigma_i^2 = 1/12$, and a moment generating function

$$M_i(t) = \frac{(e^t - 1)}{t};$$

the sum of N independent such variables therefore has $\mu = N/2$ and $\sigma^2 = N/12$. To prove that the sum is asymptotically distributed like a Gaussian with this mean and variance, we must show that

$$\lim_{N\to\infty} M(t) = e^{\frac{N}{2}t + \frac{t^2}{2}\frac{N}{12}}$$

Proof Using the property of the moment generating function of independent variables, we write

$$M(t) = M_i(t)^N = \left(\frac{1}{t}(e^t - 1)\right)^N$$
$$= \left(\frac{1 + t + t^2/2! + t^3/3! \ldots . - 1}{t}\right)^N \simeq \left(1 + \frac{t}{2} + \frac{t^2}{6} + \ldots\right)^N .$$

Neglect terms of order $O(t^3)$ and higher, and work with logarithms:

$$\ln(M(t)^N) \simeq N \ln\left(1 + \frac{t}{2} + \frac{t^2}{6}\right)$$

Use the Taylor series expansion $\ln(1 + x) \simeq (x - x^2/2 + \ldots)$, to obtain

$$\ln(M_i(t)) \simeq N\left(\frac{t}{2} + \frac{t^2}{6} - \frac{1}{2}\left(\frac{t}{2} + \frac{t^2}{6}\right)^2\right) =$$
$$N(t/2 + t^2/6 - t^2/8 + O(t^3)) \simeq N(t/2 + t^2/24)$$

in which we continued neglecting terms of order $O(t^3)$. The equation above shows that the moment generating function can be approximated as

$$M(t) \simeq e^{N\left(\frac{t}{2}+\frac{t^2}{24}\right)} \tag{4.14}$$

which is in fact the moment generating function of a Gaussian with mean $N/2$ and variance $N/12$. □

In Figure 4.1 we show the simulations of, respectively, 1000 and 100,000 samples drawn from $N = 100$ uniform and independent variables between 0 and 1. The sample distributions approximate well the limiting Gaussian with $\mu = N/2$, $\sigma = \sqrt{N/12}$. The approximation is improved when a larger number of samples are drawn, also illustrating the fact that the sample distribution approximates the parent distribution in the limit of a large number of samples collected. ◇

Example 4.5 (Sum of Two Uniform Distributions) An analytic way to develop a practical sense of how the sum of non-Gaussian distributions progressively develops the peaked Gaussian shape can be illustrated with the sum of just two uniform distributions. We start with a uniform distribution in the range of -1 to 1, which can be shown to have

$$M(t) = 1/(2t)(e^t - e^{-t}).$$

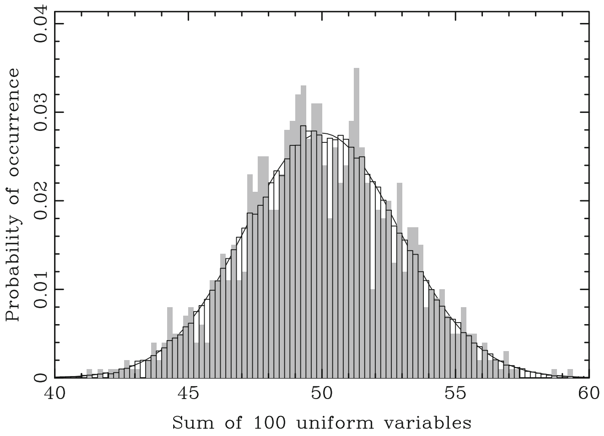

Fig. 4.1 Sample distribution functions of the sum of $N = 100$ independent uniform variables between 0 and 1, constructed from 1000 simulated measurements (*grey* histograms) and 100,000 measurements (histogram plot with *black outline*). The *solid curve* is the $N(\mu, \sigma)$ Gaussian, with $\mu = N/2$, $\sigma = \sqrt{N/12}$, the limiting distribution according to the Central Limit Theorem

The sum of two such variables will have a *triangular distribution*, given by the analytical form

$$f(x) = \begin{cases} \dfrac{1}{2} + \dfrac{x}{4} & \text{if } -2 \leq x \leq 0 \\ \dfrac{1}{2} - \dfrac{x}{4} & \text{if } 0 \leq x \leq 2. \end{cases}$$

This is an intuitive result that can be proven by showing that the moment generating function of the triangular distribution is equal to $M(t)^2$ (see Problem 4.3). The calculation follows from the definition of the moment generating function for a variable of known distribution function. The triangular distribution is the first step in the development of a peaked, Gaussian-like distribution. ◇

4.4 The Distribution of Functions of Random Variables

The general case of a variable that is a more complex function of other variables can be studied analytically when certain conditions are met. In this book we present the method of change of variables which can be conveniently applied to one-dimensional transformations and a method based on the cumulative distribution function which can be used for multi-dimensional transformations. Additional information on this subject can be found, e.g., in the textbook by Ross [38].

4.4.1 The Method of Change of Variables

A simple method for obtaining the probability distribution function of the dependent variable $Y = Y(X)$ is by using the method of *change of variables*, which applies only if the function $Y(x)$ is strictly increasing. In this case the probability distribution of $g(y)$ of the dependent variable is related to the distribution $f(x)$ of the independent variable via

$$g(y) = f(x)\frac{dx}{dy} \tag{4.15}$$

In the case of a decreasing function, the same method can be applied but the term dx/dy must be replaced with the absolute value, $|dx/dy|$.

Example 4.6 Consider a variable X distributed as a uniform distribution between 0 and 1, and the variable $Y = X^2$. The method automatically provides the information that the variable Y is distributed as

$$g(y) = \frac{1}{2}\frac{1}{\sqrt{y}}$$

with $0 \leq y \leq 1$. You can prove that the distribution is properly normalized in this domain. ◇

The method can be naturally extended to the joint distribution of several random variables. The multi-variable version of (4.15) is

$$g(u, v) = h(x, y)\,|J| \tag{4.16}$$

in which

$$J = \left(\frac{d(x, y)}{d(u, v)}\right) = \begin{pmatrix} \dfrac{dx}{du} & \dfrac{dx}{dv} \\ \dfrac{dy}{du} & \dfrac{dy}{dv} \end{pmatrix}$$

is the *Jacobian* of the transformation, in this case a 2 by 2 matrix, $h(x, y)$ is the joint probability distribution of the independent variables X, Y, and U, V are the new random variables related to the original ones by a transformation $U = u(X, Y)$ and $V = v(X, Y)$.

Example 4.7 (Transformation of Cartesian to Polar Coordinates) Consider two random variables X,Y distributed as standard Gaussians, and independent of one another. The joint probability distribution function is

$$h(x, y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.$$

Consider a transformation of variables from Cartesian coordinates x, y to polar coordinates r, θ, described by

$$\begin{cases} x = r \cdot \cos(\theta) \\ y = r \cdot \sin(\theta) \end{cases}$$

The Jacobian of the transformation is

$$J = \begin{pmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{pmatrix}$$

and its determinant is $|J| = r$. Notice that to apply the method described by (4.16) one only needs to know the inverse transformation of (x, y) as function of (r, θ). It follows that the distribution of (r, θ) is given by

$$g(r, \theta) = \frac{1}{2\pi} r e^{-\frac{r^2}{2}}$$

for $r \geq 0$, $0 \leq \theta \leq 2\pi$. The distribution $re^{-\frac{r^2}{2}}$ is called the *Rayleigh distribution*, and $1/2\pi$ can be interpreted as a uniform distribution for the angle θ between 0 and π. One important conclusion is that, since $g(r, \theta)$ can be factored out into two functions that contain separately the two variables r and θ, the two new variables are also independent. ◇

4.4.2 A Method for Multi-dimensional Functions

We will consider the case in which the variable Z is a function of two random variables X and Y, since this is a case of common use in statistics, e.g., $X + Y$, or X/Y. We illustrate the methodology with the case of the function $Z = X + Y$, when the two variables are independent. The calculation starts with the cumulative distribution function of the random variable of interest,

$$F_Z(a) = P(Z \leq a) = \int \int_{x+y \leq a} f(x)g(y)dxdy$$

in which $f(x)$ and $g(y)$ are, respectively, the probability distribution functions of X and Y, and the limits of integration must be chosen so that the sum of the two variables is less or equal than a. The portion of parameter space such that $x + y \leq a$ includes all values $x \leq a - y$, for any given value of y, or

$$F_Z(a) = \int_{-\infty}^{+\infty} dy \int_{\infty}^{a-y} f(x)g(y)dx = \int_{-\infty}^{+\infty} g(y)dyF_x(a - y)$$

where F_x is the cumulative distribution for the variable X. It is often more convenient to express the relationship in terms of the probability distribution function, which is related to the cumulative distribution function via a derivative,

$$f_Z(a) = \frac{d}{da}F_Z(a) = \int_{-\infty}^{\infty} f(a-y)g(y)dy. \tag{4.17}$$

This relationship is called the *convolution* of the distributions $f(x)$ and $g(y)$.

Example 4.8 (Sum of Two Independent Uniform Variables) Calculate the probability distribution function of the sum of two independent uniform random variables between −1 and +1.

The probability distribution function of a uniform variable between −1 and +1 is $f(x) = 1/2$, defined for $-1 \leq x \leq 1$. The convolution gives the following integral

$$f_Z(a) = \int_{-1}^{+1} \frac{1}{2} f(a-y)dy.$$

The distribution function of the sum Z can have values $-2 \leq a \leq 2$, and the convolution must be divided into two integrals, since $f(a-y)$ is only defined between −1 and +1. We obtain

$$f_Z(a) = \frac{1}{4} \times \begin{cases} \int_{-1}^{a+1} dy & \text{if } -2 \leq a \leq 0 \\ \int_{a-1}^{1} dy & \text{if } 0 \leq a \leq 2. \end{cases}$$

This results in

$$f_Z(a) = \frac{1}{4} \times \begin{cases} (a+2) & \text{if } -2 \leq a \leq 0 \\ (2-a) & \text{if } 0 \leq a \leq 2 \end{cases}$$

which is the expected triangular distribution between −2 and +2. ◇

Another useful application is for the case of $Z = X/Y$, where X and Y are again independent variables. We begin with the cumulative distribution,

$$F_Z(z) = P(Z < z) = P(X/Y < z) = P(X < zY).$$

For a given value y of the random variable Y, this probability equals $F_X(zy)$; since Y has a probability $f_Y(y)dy$ to be in the range between y and $y + dy$, we obtain

$$F_Z(z) = \int F_X(zy) f_Y(y)dy.$$

Following the same method as for the derivation of the distribution of $X + Y$, we must take the derivative of $F_Z(z)$ with respect to z to obtain:

$$f_Z(z) = \int f_X(zy) y f_Y(y) dy. \tag{4.18}$$

This is the integral than must be solved to obtain the distribution of X/Y.

4.5 The Law of Large Numbers

Consider N random variables X_i that are identically distributed, and μ is their common mean. The *Strong Law of Large Numbers* states that, under suitable conditions on the variance of the random variables, the sum of the N variables tends to the mean μ, which is a deterministic number and not a random variable. This result can be stated as

$$\lim_{n\to\infty} \frac{X_1 + \ldots + X_N}{N} = \mu, \tag{4.19}$$

and it is, together with the Central Limit Theorem, one of the most important results of the theory of probability, and of great importance for statistics. Equation (4.19) is a very strong statement because it shows that, asymptotically, the sum of random variables becomes a constant equal to the sample mean of the N variables, or N measurements. Although no indication is given towards establishing how large N should be in order to achieve this goal, it is nonetheless an important result that will be used in determining the asymptotic behavior of random variables. Additional mathematical properties of this law can be found in books of theory of probability, such as [38] or [26].

Instead of providing a formal proof of this law, we want to focus on an important consequence. Given a function $y(x)$, we would like to estimate its expected value $E[y(X)]$ from the N measurements of the variables X_i. According to the law of large numbers, we can say that

$$\lim_{n\to\infty} \frac{y(X_1) + \ldots + y(X_N)}{N} = E[y(X)]. \tag{4.20}$$

Equation (4.20) states that a large number of measurements of the variables X_i can be used to measure the expectation of $E[y(X)]$, entirely bypassing the probability distribution function of the function $y(X)$. This property is used in the following section.

4.6 The Mean of Functions of Random Variables

For a function of random variables it is often necessary or convenient to develop methods to estimate the mean and the variance without having full knowledge of its probability distribution function.

For functions of a single variable $Y = y(X)$

$$E[y(X)] = \int y(x)f(x)dx \tag{4.21}$$

where $f(x)$ is the distribution function of X. This is in fact a very intuitive result, stating that the distribution function of X is weighted by the function of interest, and it makes it straightforward to compute expectation values of variables without first having to calculate their full distribution. According to the law of large numbers, this expectation can be estimated from N measurements x_i as per (4.20),

$$\overline{y(x)} = \frac{y(x_1) + \ldots + y(x_n)}{N}. \tag{4.22}$$

An important point is that the mean of the function is *not* equal to the function of the mean, $\overline{y(x)} \neq y(\overline{x})$, as will be illustrated in the following example. Equation (4.22) says that we must have access to the *individual measurements* of the variable X, if we want to make inferences on the mean of a function of X. If, for example, we only had the mean $\overline{x}$, we cannot measure $\overline{u(x)}$. This point is relevant when one has limited access to the data, e.g., when the experimenter does not report all information on the measurements performed.

Example 4.9 (Mean of Square of a Uniform Variable) Consider the case of a uniform variable U in the range 0–1, with mean 1/2. If we want to evaluate the parent mean of $X = U^2$, we calculate

$$\mu = \int_0^1 u^2 du = 1/3.$$

It is important to see that the mean of U^2 is not just the square of the mean of U, and therefore the means do not transform following the same analytic expression as the random variables. You can convince yourself of this fact by assuming to draw five "fair" samples from a uniform distribution, 0.1, 0.3, 0.5, 0.7 and 0.9—they can be considered as a dataset of measurements. Clearly their mean is 1/2, but the mean of their squares is 1/3 and not 1/4, in agreement with the theoretical calculation of the parent mean. ◇

Another example where the mean of the function does not equal to the function of the mean is reported in Problem 4.5, in which you can show that using the means of I and W/Q do not give the mean of m/e for the Thomson experiment to measure the mass to charge ratio of the electron. The problem provides a multi-dimensional extension to (4.22), since the variable m/e is a function of two variables that have been measured in pairs.

4.7 The Variance of Functions of Random Variables and Error Propagation Formulas

A random variable Z that is a function of other variables can have its variance estimated directly if the measurements of the independent variables are available, similar to the case of the estimation of the mean. Considering, for example, the case of a function $Z = z(U)$ that depends on just one variable, for which we have N measurements $u_1, \ldots, u_N$ available. With the mean estimated from (4.22), the variance can accordingly be estimated as

$$s_u^2 = \frac{(z(u_1) - \overline{z})^2 + \ldots + (z(u_N) - \overline{z})^2}{N - 1}, \tag{4.23}$$

as one would normally do, treating the numbers $z(u_1), \ldots, z(u_N)$ as samples from the dependent variable. This method can naturally be extended to more than one variable, as illustrated in the following example. When the measurements of the independent variables are available, this method is the straightforward way to estimate the variance of the function of random variables.

Example 4.10 Using the Thomson experiment described on page 23, consider the data collected for Tube 1, consisting of 11 measurements of W/Q and I, from which the variable of interest v is calculated as

$$v = 2\frac{W/Q}{I}$$

From the reported data, one obtains 11 measurements of v, from which the mean and standard deviation can be immediately calculated as $\overline{v} = 7.9 \times 10^9$, and $s_v = 2.8 \times 10^9$. ◇

There are a number of instances in which one does not have access to the original measurements of the independent variable or variables, required for an accurate estimate of the variance according to (4.23). In this care, an approximate method to estimate the variance must be used instead. This method takes the name of *error propagation*. Consider a random variable Z that is a function of a number of variables, $Z = z(U, V, \ldots)$. A method to approximate the variance of Z in terms of the variance of the independent variables U, V, etc. starts by expanding Z in a

Taylor series about the means of the independent variables, to obtain

$$z(u, v, \dots) = z(\mu_u, \mu_v, \dots) + (u - \mu_u) \left.\frac{\partial z}{\partial u}\right|_{\mu_u} + (v - \mu_v) \left.\frac{\partial z}{\partial v}\right|_{\mu_v} + \dots + O(u - \mu_u)^2 + O(v - \mu_v)^2 + \dots$$

Neglecting terms of the second order, the expectation of Z would be given by $E[Z] = z(\mu_u, \mu_v, \dots)$, i.e., the mean of X would be approximated as $\mu_X = z(\mu_u, \mu_v, \dots)$. This is true only if the function is linear, and we have shown in Sect. 4.6 that this approximation may not be sufficiently accurate in the case of nonlinear functions such as U^2. This approximation for the mean is used to estimate the variance of Z, for which we retain only terms of first order in the Taylor expansion:

$$\begin{aligned} E[(Z - E[Z])^2] &\simeq E\left[\left((u - \mu_u) \left.\frac{\partial z}{\partial u}\right|_{\mu_u} + (v - \mu_v) \left.\frac{\partial z}{\partial v}\right|_{\mu_v} + \dots\right)^2\right] \\ &\simeq E\left[\left((u - \mu_u) \left.\frac{\partial z}{\partial u}\right|_{\mu_u}\right)^2 \left((v - \mu_v) \left.\frac{\partial z}{\partial v}\right|_{\mu_v}\right)^2 \right. \\ &\left. + \, 2(u - \mu_u) \left.\frac{\partial z}{\partial u}\right|_{\mu_u} \cdot (v - \mu_v) \left.\frac{\partial z}{\partial v}\right|_{\mu_v} + \dots \right]. \end{aligned}$$

This formula can be rewritten as

$$\sigma_X^2 \simeq \sigma_u^2 \left.\frac{\partial f}{\partial u}\right|_{\mu_u}^2 + \sigma_v^2 \left.\frac{\partial f}{\partial v}\right|_{\mu_v}^2 + 2 \cdot \sigma_{uv}^2 \left.\frac{\partial f}{\partial u}\right|_{\mu_u} \left.\frac{\partial f}{\partial v}\right|_{\mu_v} + \dots \tag{4.24}$$

which is usually referred to as the *error propagation formula*, and can be used for any number of independent variables. This result makes it possible to estimate the variance of a function of variable, knowing simply the variance of each of the independent variables and their covariances. The formula is especially useful for all cases in which the measured variables are independent, and all that is known is their mean and standard deviation (but not the individual measurements used to determine the mean and variance). This method must be considered as an approximation when there is only incomplete information about the measurements. Neglecting terms of the second order in the Taylor expansion can in fact lead to large errors, especially when the function has strong nonlinearities. In the following we provide a few specific formulas for functions that are of common use.

4.7.1 Sum of a Constant

Consider the case in which a constant a is added to the variable U,

$$Z = U + a$$

where a is a deterministic constant which can have either sign. It is clear that $\partial z/\partial a = 0$, $\partial z/\partial u = 1$, and therefore the addition of a constant has no effect on the uncertainty of X,

$$\sigma_z^2 = \sigma_u^2. \tag{4.25}$$

The addition or subtraction of a constant only changes the mean of the variable by the same amount, but leaves its standard deviation unchanged.

4.7.2 Weighted Sum of Two Variables

The variance of the weighted sum of two variables,

$$Z = aU + bV$$

where a, b are constants of either sign, can be calculated using $\partial z/\partial u = a$, $\partial z/\partial v = b$. We obtain

$$\sigma_z^2 = a^2\sigma_u^2 + b^2\sigma_v^2 + 2ab\sigma_{uv}^2. \tag{4.26}$$

The special case in which the two variables U, V are uncorrelated leads to the weighted sum of the variances.

Example 4.11 Consider a decaying radioactive source which is found to emit $N_1 = 50$ counts and $N_2 = 35$ counts in two time intervals of same duration, during which $B = 20$ background counts are recorded. This is an idealized situation in which we have directly available the measurement of the background counts. In the majority of real-life experiments one simply measures the sum of signal plus background, and in those cases additional considerations must be used. We want to calculate the background subtracted source counts in the two time intervals and estimate their *signal-to-noise ratio*, defined as $S/N = \mu/\sigma$. The inverse of the signal-to-noise ratio is the relative error of the variable.

Each random variable N_1, N_2, and B obeys the Poisson distribution, since it comes from a counting process. Therefore, we can estimate the following parent means and variances from the sample measurements,

$$\begin{cases} \mu_1 = 50 & \sigma_1 = \sqrt{50} = 7.1 \\ \mu_2 = 35 & \sigma_2 = \sqrt{35} = 5.9 \\ \mu_B = 20 & \sigma_B = \sqrt{20} = 4.5 \end{cases}$$

Since the source counts are given by $S_1 = N_1 - B$ and $S_2 = N_2 - B$, we can now use the approximate variance formulas *assuming* that the variables are uncorrelated, $\sigma_{S_1} = \sqrt{50 + 20} = 8.4$ and $\sigma_{S_2} = \sqrt{35 + 20} = 7.4$. The two measurements of the source counts would be reported as $S_1 = 30 \pm 8.4$ and $S_2 = 15 \pm 7.4$, from which the signal-to-noise ratios are given, respectively, as $\mu_{S1}/\sigma_{S_1} = 3.6$ and $\mu_{S2}/\sigma_{S_2} = 2.0$. ◇

4.7.3 Product and Division of Two Random Variables

Consider the product of two random variables U, V, optionally also with a constant factor a of either sign,

$$Z = aUV. \tag{4.27}$$

The partial derivatives are $\partial z/\partial u = av$, $\partial z/\partial v = au$, leading to the approximate variance of

$$\sigma_z^2 = a^2 v^2 \sigma_u^2 + a^2 u^2 \sigma_v^2 + 2auv\sigma_{uv}^2.$$

This can be rewritten as

$$\frac{\sigma_z^2}{z^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + 2\frac{\sigma_{uv}^2}{uv}. \tag{4.28}$$

Similarly, the division between two random variables,

$$Z = a\frac{U}{V}, \tag{4.29}$$

leads to

$$\frac{\sigma_z^2}{z^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - 2\frac{\sigma_{uv}^2}{uv}. \tag{4.30}$$

Notice the equations for product and division differ by just one sign, meaning that a positive covariance between the variables leads to a reduction in the standard deviation for the division, and an increase in the standard deviation for the product.

Example 4.12 Using the Thomson experiment of page 23, consider the data for Tube 1, and assume that the only number available are the mean and standard deviation of W/Q and I. From these two numbers we want to estimate the mean and variance of v. The measurement of the two variables are $W/Q = 13.3 \pm 8.5 \times 10^{11}$ and $I = 312.9 \pm 93.4$, from which the mean of v would have to be estimated as $\overline{v} = 8.5 \times 10^9$—compare with the value of 7.9×10^9 obtained from the individual measurements.

The estimate of the variance requires also a knowledge of the covariance between the two variables W/Q and I. In the absence of any information, we will assume that the two variables are uncorrelated, and use the error propagation formula to obtain

$$\sigma_v \simeq 2 \times \frac{13.3 \times 10^{11}}{312.9} \times \left(\left(\frac{8.5}{13.3} \right)^2 + \left(\frac{93.4}{312.9} \right)^2 \right)^{1/2} = 6 \times 10^9,$$

which is a factor of 2 larger than estimated directly from the data (see Example 4.10). Part of the discrepancy is to be attributed to the neglect of the covariance between the measurement, which can be found to be positive, and therefore would reduce the variance of v according to (4.30). Using this approximate method, we would estimate the measurement as $v = 8.5 \pm 6 \times 10^9$, instead of $7.9 \pm 2.8 \times 10^9$. ◇

4.7.4 Power of a Random Variable

A random variable may be raised to a constant power, and optionally multiplied by a constant,

$$Z = aU^b \tag{4.31}$$

where a and b are constants of either sign. In this case, $\partial z/\partial u = abu^{b-1}$ and the error propagation results in

$$\frac{\sigma_z}{z} = |b| \frac{\sigma_u}{u}. \tag{4.32}$$

This results states that the relative error in Z is b times the relative error in U.

4.7.5 Exponential of a Random Variable

Consider the function

$$Z = ae^{bU}, \tag{4.33}$$

where a and b are constants of either sign. The partial derivative is $\partial z/\partial u = abe^{bu}$, and we obtain

$$\frac{\sigma_z}{z} = |b|\sigma_u. \tag{4.34}$$

4.7.6 Logarithm of a Random Variable

For the function

$$Z = a\ln(bU), \tag{4.35}$$

where a is a constant of either sign, and $b > 0$. The partial derivative is $\partial z/\partial u = a/U$, leading to

$$\sigma_z = |a|\frac{\sigma_u}{u}. \tag{4.36}$$

A similar result applies for a base-10 logarithm,

$$Z = a\log(bU), \tag{4.37}$$

where a is a constant of either sign, and $b > 0$. The partial derivative is $\partial z/\partial u = a/(U\ln(10))$, leading to

$$\sigma_z = |a|\frac{\sigma_u}{u\ln(10)}. \tag{4.38}$$

Similar error propagation formulas can be obtained for virtually any analytic function for which derivatives can be calculated. Some common formulas are reported for convenience in Table 4.1, where the terms z, u, and v refer to the random variables evaluated at their estimated mean value.

Example 4.13 With reference to Example 4.11, we want to give a quantitative answer to the following question: what is the *probability* that during the second time interval the radioactive source was actually detected? In principle a fluctuation of the number of background counts could give rise to all detected counts.

Table 4.1 Common error propagation formulas

Function	Error propagation formula	Notes
$Z = U + a$	$\sigma_z^2 = \sigma_u^2$	a is a constant
$Z = aU + bV$	$\sigma_z^2 = a^2\sigma_u^2 + b^2\sigma_b^2 + 2ab\sigma_{uv}^2$	a, b are constants
$Z = aUV$	$\dfrac{\sigma_z^2}{z^2} = \dfrac{\sigma_u^2}{u^2} + \dfrac{\sigma_v^2}{v^2} + 2\dfrac{\sigma_{uv}^2}{uv}$	a is a constant
$Z = a\dfrac{U}{V}$	$\dfrac{\sigma_z^2}{z^2} = \dfrac{\sigma_u^2}{u^2} + \dfrac{\sigma_v^2}{v^2} - 2\dfrac{\sigma_{uv}^2}{uv}$	a is a constant
$Z = aU^b$	$\dfrac{\sigma_z}{z} = b\dfrac{\sigma_u}{u}$	a, b are constants
$Z = ae^{bU}$	$\dfrac{\sigma_z}{z} = \lvert b\rvert\sigma_u$	a, b are constants
$Z = a\ln(bU)$	$\sigma_z = \lvert a\rvert\dfrac{\sigma_u}{u}$	a, b are constants, $b > 0$
$Z = a\log(bU)$	$\sigma_z = \lvert a\rvert\dfrac{\sigma_u}{u\ln(10)}$	a, b are constants, $b > 0$

A solution to this question can be provided by stating the problem in a Bayesian way:

$$\mathrm{P(detection)} = \mathrm{P}(S_2 > 0/\mathrm{data})$$

where the phrase “data” refers also to the available measurement of the background where $S_2 = N_2 - B$ is the number of source counts. This could be elaborated by stating that the data were used to estimate a mean of 15 and a standard deviation of 7.4 for S_2, and therefore we want to calculate the probability to exceed zero for such random variable. We can use the Central Limit Theorem to say that the sum of two random variables—each approximately distributed as a Gaussian since the number of counts is sufficiently large—is Gaussian, and the probability of a positive detection of the radioactive source therefore becomes equivalent to the probability of a Gaussian-distributed variable to have values larger than approximately $\mu - 2\sigma$. According to Table A.3, this probability is approximately 97.7 %. We can therefore conclude that source were detected in the second time period with such confidence. ◇

4.8 The Quantile Function and Simulation of Random Variables

In data analysis one often needs to simulate a random variable, that is, drawing random samples from a parent distribution. The simplest such case is the generation of a random number between two limits, which is equivalent to drawing samples from a uniform distribution. In particular, several Monte Carlo methods including the Markov chain Monte Carlo method discussed in Chap. 16 will require random variables with different distributions. Most computer languages and programs do have available a uniform distribution, and thus it is useful to learn how to simulate any distribution based on the availability of a simulator for a uniform variable.

Given a variable X with a distribution $f(x)$ and a cumulative distribution function $F(x)$, we start by defining the *quantile function* $F^{-1}(p)$ as

$$F^{-1}(p) = \min\{x\varepsilon\mathscr{R}, p \leq F(x)\} \tag{4.39}$$

with the meaning that x is the minimum value of the variable at which the cumulative distribution function reaches the value $0 \leq p \leq 1$. The word "minimum" in the definition of the quantile function is necessary to account for those distributions that have steps—or discontinuities—in their cumulative distribution, but in the more common case of a strictly increasing cumulative distribution, the quantile function is simply defined by the relationship $p = F(x)$. This equation can be solved for x, to obtain the quantile function $x = F^{-1}(p)$.

Example 4.14 (Quantile Function of a Uniform Distribution) For a uniform variable in the range 0–1, the quantile function has a particularly simple form. In fact, $F(x) = x$, and the quantile function defined by the equation $p = F(x)$ yields $x = p$, and therefore

$$x = F^{-1}(p) = p. \tag{4.40}$$

Therefore the analytical form of both the cumulative distribution and the quantile function is identical for the uniform variable in 0–1, meaning that, e.g., the value 0.75 of the random variable is the $p = 0.75$, or 75 % quantile of the distribution. $\diamond$

The basic property of the quantile function can be stated mathematically as

$$p \leq F(x) \Leftrightarrow x \leq F^{-1}(p) \tag{4.41}$$

meaning that the value of $F^{-1}(p)$ is the value x at which the probability of having $X \geq x$ is p.

Example 4.15 (Quantile Function of an Exponential Distribution) Consider a random variable distributed like an exponential,

$$f(x) = \lambda e^{-\lambda x},$$

with $x \geq 0$. Its cumulative distribution function is

$$F(x) = 1 - e^{-\lambda x}.$$

The quantile function is obtained from,

$$p = F(x) = 1 - e^{-\lambda x},$$

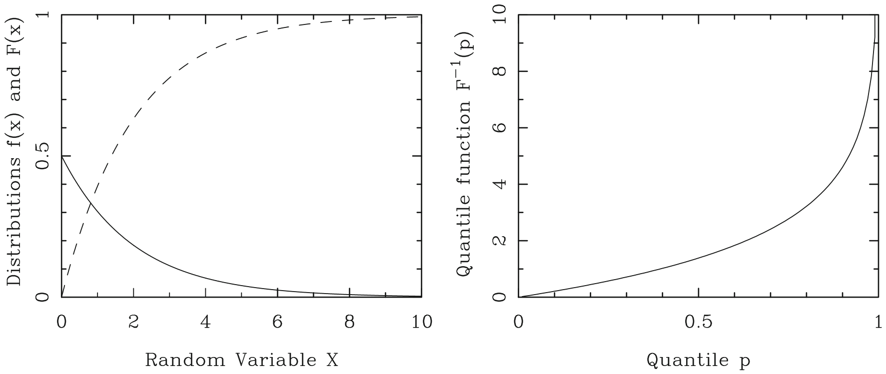

Fig. 4.2 Distribution function $f(x)$, cumulative distribution $F(x)$, and quantile function $F^{-1}(p)$ of an exponential variable with $\lambda = 1/2$

leading to $x = \ln(1-p)/(-\lambda)$, and therefore the quantile function is

$$x = F^{-1}(p) = \frac{ln(1-p)}{-\lambda}.$$

Figure 4.2 shows the cumulative distribution and the quantile function for the exponential distribution. ◇

4.8.1 General Method to Simulate a Variable

The method to simulate a random variable is summarized in the following equation,

$$X = F^{-1}(U), \tag{4.42}$$

which states that any random variable X can be expressed in terms of the uniform variable U between 0 and 1, F is the cumulative distribution of the variable X, and F^{-1} is the quantile function. If a closed analytic form for F is available for that distribution, this equation results in a simple method to simulate the random variable.

Proof We have already seen that for the uniform variable the quantile function is $F^{-1}(U) = U$, i.e., it is the uniform random variable itself. The proof therefore simply consists of showing that, assuming (4.42), then the cumulative distribution of X is indeed $F(X)$, or $P(X \leq x) = F(x)$. This can be shown by writing

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x),$$

in which the second equality follows from the definition of the quantile function, and the last equality follows from the fact that $P(U \leq u) = u$, for u a number between 0 and 1, for a uniform variable. □

Example 4.16 (Simulation of an Exponential Variable) Consider a random variable distributed like an exponential, $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Given the calculations developed in the example above, the exponential variable can be simulated as

$$X = \frac{\ln(1-U)}{-\lambda}.$$

Notice that, although this relationship is between random variables, its practical use is to draw random samples u from U, and a random sample x from X is obtained by simply using the equation

$$x = \frac{\ln(1-u)}{-\lambda}.$$

Therefore, for a large sample of values u, the above equation returns a random sample of values for the exponential variable X. ◇

Example 4.17 (Simulation of the Square of Uniform Variable) It can be proven that the simulation of the square of a uniform random variable $Y = U^2$ is indeed achieved by squaring samples from a uniform distribution, a very intuitive result.

In fact, we start with the distribution of Y as $g(y) = {}^1\!/_2 \; y^{-1/2}$. Since its cumulative distribution is given by $G(y) = \sqrt{y}$, the quantile function is defined by $p = \sqrt{y}$, or $y = p^2$ and therefore the quantile function for U^2 is

$$y = G^{-1}(p) = p^2.$$

This result, according to (4.42), defines U^2, or the square of a uniform distribution, as the function that needs to be simulated to draw fair samples from Y. ◇

4.8.2 Simulation of a Gaussian Variable

This method of simulation of random variables relies on the knowledge of $F(x)$ and the fact that such a function is analytic and invertible. In the case of the Gaussian distribution, the cumulative distribution function is a special function,

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

which cannot be inverted analytically. Therefore, this method cannot be applied. This complication must be overcome, given the importance of Gaussian distribution in probability and statistics. Fortunately, a relatively simple method is available that permits the simulation of two Gaussian distributions from two uniform random variables.

In Sect. 4.4 we showed that the transformation from Cartesian to polar coordinates results in two random variables R, Θ that are distributed, respectively, like a Rayleigh and a uniform distribution:

$$\begin{cases} h(r) = re^{-\frac{r^2}{2}} & r \geq 0 \\ i(\theta) = \dfrac{1}{2\pi} & 0 \leq \theta \leq 2\pi. \end{cases} \tag{4.43}$$

Since these two distributions have an analytic form for their cumulative distributions, R and Θ can be easily simulated. We can then use the transformation given by (4.7) to simulate a pair of independent standard Gaussians. We start with the Rayleigh distribution, for which the cumulative distribution function is

$$H(r) = 1 - e^{-\frac{r^2}{2}}.$$

The quantile function is given by

$$p = 1 - e^{-\frac{r^2}{2}},$$

and from this we obtain

$$r = \sqrt{-2\ln(1-p)} = H^{-1}(p)$$

and therefore $R = \sqrt{-2\ln(1-U)}$ simulates a Rayleigh distribution, given the uniform variable U. For the uniform variable Θ, it is clear that the cumulative distribution is given by

$$I(\theta) = \begin{cases} \theta/(2\pi) & 0 \leq \theta \leq 2\pi \\ 0 & \text{otherwise;} \end{cases}$$

the quantile function is $\theta = 2\pi p = I^{-1}(p)$, and therefore $\Theta = 2\pi V$ simulates a uniform distribution between 0 and 2π, with V the uniform distribution between 0 and 1.

Therefore, with the use of two uniform distributions U, V, we can use R and Θ to simulate a Rayleigh and a uniform angular distribution

$$\begin{cases} R = \sqrt{-2\ln(1-U)} \\ \Theta = 2\pi V. \end{cases} \tag{4.44}$$

Then, using the Cartesian-Polar coordinate transformation, we arrive at the formulas needed to simulate a pair of Gaussians X and Y:

$$\begin{cases} X = R\cos(\Theta) = \sqrt{-2\ln(1-U)} \cdot \cos(2\pi V) \\ Y = R\sin(\Theta) = \sqrt{-2\ln(1-U)} \cdot \sin(2\pi V) \end{cases} \tag{4.45}$$

Equations (4.45) can be easily implemented by having available two simultaneous and independent uniform variables between 0 and 1.

Summary of Key Concepts for this Chapter

- ☐ Linear combination of variables: The formulas for the mean and variance of the linear combination of variables are

$$\begin{cases} \mu = \sum a_i \mu_i \\ \sigma^2 = \sum_{i=1}^N a_i^2 \sigma_i^2 + 2\sum_{i=1}^N \sum_{j=i+1}^N a_i a_j \sigma_{ij}^2 \end{cases}$$

- ☐ *Variance of uncorrelated variables*: When variables are uncorrelated the variances add linearly. The variance of the mean of N independent measurements is $\sigma_Y^2 = \sigma^2/N$.
- ☐ *Moment generating function*: It is a mathematical function that enables the calculation of moments of a distribution, $M(t) = E[e^{tX}]$.
- ☐ *Central Limit theorem*: The sum of a large number of independent variables is distributed like a Gaussian of mean equal to the sum of the means and variance equal to the sum of the variances.
- ☐ *Method of change of variables*: A method to obtain the distribution function of a variable Y that is a function of another variable X, $g(y) = f(x)dx/dy$.
- ☐ *Law of Large Numbers*: The sum of a large number of random variables with mean μ tends to a constant number equal to μ.
- ☐ *Error propagation formula*: It is an approximation for the variance of a function of random variables. For a function $x = f(u, v)$ of two uncorrelated variables U and V, the variance of X is given by

$$\sigma_x^2 = \sigma_u^2 \left.\frac{\partial f}{\partial u}\right|^2 + \sigma_v^2 \left.\frac{\partial f}{\partial v}\right|^2$$

- ☐ *Quantile function*: It is the function $x = F^{-1}(p)$ used to find the value x of a variable that corresponds to a given quantile p.
- ☐ *Simulation of a Gaussian*: Two Gaussians can be obtained from two uniform random variables U, V via

$$\begin{cases} X = \sqrt{-2\ln(1-U)}\cos(2\pi V) \\ Y = \sqrt{-2\ln(1-U)}\sin(2\pi V) \end{cases}$$

Problems

4.1 Consider the data from Thomson's experiment of Tube 1, from page 23.

(a) Calculate the mean and standard deviation of the measurements of v.
(b) Use the results from Problem 2.3, in which the mean and standard deviation of W/Q and I were calculated, to calculate the approximate values of mean and standard deviation of v using the relevant error propagation formula, assuming no correlation between the two measurements.

This problem illustrates that the error propagation formulas may give different results than direct measurement of the mean and variance of a variable, when the individual measurements are available.

4.2 Calculate the mean, variance, and moment generating function $M(t)$ for a uniform random variable in the range 0–1.

4.3 Consider two uniform independent random variables X, Y in the range -1 to 1.

(a) Determine the distribution function, mean and variance, and the moment generating function of the variables.
(b) We speculate that the sum of the two random variables is distributed like a "triangular" distribution between the range -2 to 2, with distribution function

$$f(x) = \begin{cases} \dfrac{1}{2} + \dfrac{x}{4} \text{ if } -2 \leq x \leq 0 \\ \dfrac{1}{2} - \dfrac{x}{4} \text{ if } 0 \leq x \leq 2 \end{cases}$$

Using the moment generation function, prove that the variable $Z = X + Y$ is distributed like the triangular distribution above.

4.4 Using a computer language of your choice, simulate the sum of $N = 100$ uniform variables in the range 0–1, and show that the sampling distribution of the sum of the variables is approximately described by a Gaussian distribution with mean equal to the mean of the N uniform variables and variance equal to the sum of the variances. Use 1,000 and 100,000 samples for each variable.

4.5 Consider the J.J. Thomson experiment of page 23.

(a) Calculate the sample mean and the standard deviation of m/e for Tube 1.
(b) Calculate the approximate mean and standard deviation of m/e from the mean and standard deviation of W/Q and I, according to the equation

$$\frac{m}{e} = \frac{I^2}{2}\frac{Q}{W};$$

Assume that W/Q and I are uncorrelated.

4.6 Use the data provided in Example 4.11. Calculate the probability of a positive detection of source counts S in the first time period (where there are $N_1 = 50$ total counts and $B = 20$ background counts), and the probability that the source emitted $\geq$10 source counts. You will need to assume that the measured variable can be approximated by a Gaussian distribution.

4.7 Consider the data in the Thomson experiment for Tube 1 and the fact that the variables W/Q and I are related to the variable v via the relationship

$$v = \frac{2W}{QI}.$$

Calculate the sample mean and variance of v from the direct measurements of this variable, and then using the measurements of W/Q and I and the error propagation formulas. By comparison of the two estimates of the variance, determine if there is a positive or negative correlation between W/Q and I.

4.8 Provide a general expression for the error propagation formula when three independent random variables are present, to generalize (4.24) that is valid for two variables.