Appendix A

Glossary

This section summarises many of the terms used in this book. A brief description of the meaning of each term is given, and where appropriate a cross reference to the corresponding section in the text is suggested for further reading.
$\wedge$: Logical and.
$\vee$: Logical or.
$\boldsymbol{\chi^2}$: See chi-square.
$\boldsymbol{\mu}$, $\boldsymbol{\mu(i)}$ **or** $\boldsymbol{\mu_i}$: See mean.
$\boldsymbol{\rho}$, $\boldsymbol{\rho(x,y)}$ **or** $\boldsymbol{\rho_{xy}}$: See correlation.
$\boldsymbol{\sigma}$, $\boldsymbol{\sigma(i)}$ **or** $\boldsymbol{\sigma_i}$: See standard deviation.
$\boldsymbol{\sigma_{xy}}$ **or** $\boldsymbol{\sigma(x,y)}$ **or** $\boldsymbol{\sigma_{ij}}$: See covariance.
Alternate Hypothesis ($\boldsymbol{H_1}$) : The complement of a null hypothesis. If a statistical test is performed for a null hypothesis, and that test fails, then the alternate hypothesis (as the complement of the null hypothesis) is the result supported by the data.
Arithmetic Mean : See mean.
Artificial Neural Network : Also called a neural network. This is a collection of perceptrons fed by some number of input variables, with one or more output nodes. The function of the neural network is to try and optimally distinguish between two or more types of data. See section 9.4.
Bagging : This is the name given to a technique of oversampling data used for training decision trees. See section 9.5.2.
Bayes Theorem : See Eq. (2.2.1).
Bayesian Classifier : A classification algorithm based on Bayes theorem, see Sec. 9.2.
Bayesian Statistics : The branch of statistics derived from the work of Rev. Bayes. This is a subjective approach to the computation of probability that can be used in a variety of situations, including for those that are not repeatable. See section 2.2.
Binomial Distribution : A distribution that describes a bi-modal situation, and is often presented in terms of success with some probability p, and failure with some probability $q = 1 - p$. See section 4.2.
Binomial Error : The error on an observable that is bi-modal, such as the detection efficiency, is given by a binomial error. See section 5.3.
Boosting : This is the name given to a re-weighting of events used for training decision trees. See section 9.5.1.
Blind Analysis : A method of validating and performing an analysis while not being able to determine the true value of the measured observable until the last moment in the measurement process. This methodology can be used to try and minimise experimenter bias, although it is not applicable in all circumstances. Blind analysis is discussed in more detail in section 5.6.
Breit-Wigner : A probability density function often used to describe resonant behaviour in physics. See appendix B.1.2.
Chi-square : The sum of squared deviations of data x_i relative to some model $\theta(x_i)$ normalised by the

standard deviation of the i^{th} point. See section 4.5. Also used in the context of a fit; see chapter 8.
Combination of Errors : The process of computing the error on some function f depending on an number of measurements of observables $x \pm \sigma_x$, $y \pm \sigma_y$, etc. as discussed in section 5.2.
Confidence Level : The confidence level of X% is a statement about the value of some observable obtained as the conclusion of a measurement. Assuming that the original measurement corresponds to the true value, a repeat measurement would yield the same or a compatible conclusion X% of the time. The corollary of this is that 100-X% of the time, one can obtain an equally valid result that would be in disagreement with the original conclusion.
Confidence Interval : The interval in some variable corresponding to a given confidence level.
Correlation : The correlation between two variables x and y is is a measure of how dependent they are on each other. The symbol often used to denote correlation is ρ (with or without subscripts). If the correlation ρ_{xy} is zero, then x and y are independent, and if the correlation has a magnitude of one, then having determined the value one of the variables, the other can be determined as well. See section 3.6.2.
Correlation Matrix : For two or more variables, it is possible to construct a matrix of correlation coefficients, referred to as the correlation matrix.
Covariance : The covariance of two variables x and y is the term given to the average of the sum of the products of residuals of the variables x and y relative to the mean value of those variables. The covariance between x and y is denoted by σ_{xy}, and it is related to the correlation ρ_{xy} by $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$, where the sigmas refer to the standard deviations of x and y. See section 3.6.1.
Covariance Matrix : For two or more variables, it is possible to construct a matrix of variance and covariance coefficients, referred to as the covariance matrix.
Coverage : Corresponds to the confidence level with which an interval or limit is defined by. Under coverage refers to the situation where a limit or interval with a quoted CL actually has less coverage than stated. Over coverage refers to the opposite scenario where the quoted limit or interval has more coverage than stated. Both under coverage and over coverage are undesirable.
Cut-Based Selection : A multivariate technique to distinguish between types or classes of data. See section 9.1.
Decision Tree : A multivariate technique used to distinguish between types or classes of data. See section 9.5.
Error : There are several types of error that one might consider:

- A mistake.
- See Type I error.
- See Type II error.
- See standard error on the mean (or standard deviation).
- See combination of errors.
- Event mis-classification in a multivariate algorithm.

The context that the term error is used in should make the meaning of the word clear.
Error Matrix : See covariance matrix.
Expectation Value : The mean result expected from a given distribution or measurement.
False negative : See type I error.
False positive : See type II error.
Fisher Discriminant : A multivariate technique used to distinguish between types or classes of data. See section 9.3.
Fit/Fitting : Given a set of data $\underline{D}$, the process of optimising a set of parameters $\underline{p}$ of some model θ in order that the model describes the data in the best possible way is called fitting. One fits the model to the data.
Frequentist Statistics : The branch of statistics derived from the axiom that one can repeat measurements many times. A frequentist probability is determined from taking the limit that the number of repeated measurements tends to infinity. Often it is possible to exactly compute a frequentist probability without taking

this limit. This method can not be used if a measurement is not repeatable. See section 2.4.
Full Width at Half Maximum (FWHM) : The full width at half maximum (FWHM) of a function $f(x)$ is the distance in x between the two points on $f(x)$ that are at half of the global maximum of the function. It is usually only sensible to use the FWHM for functions that have a single global maximum such as a Gaussian or Breit-Wigner function.
Gaussian Distribution : A probability density function often used to describe measurements made using large samples of data. See section 4.4.
Goodness of fit : A quantitative measure of how well a data and model agree with each other. See section 4.5.
Graph : A visual display representing a set of points or a function in two or three dimensions. See section 3.1.2.
Gradient Decent : A numerical method used for optimisation of some test statistic. See section 8.1.1.
Histogram : A binned graphical representation of data. See section 3.1.1.
Hypothesis : A postulate compared against a sample of data.
Least-Squares Fit : A fitting technique derived from the χ^2 distribution for a set of data. See section 8.3, also see chi-square.
Likelihood : The likelihood function is a probability density function used to model a sample of data. See section 8.4.
Maximum Likelihood Fit : The use of a likelihood function to fit to a sample of data. See section 8.4.
Extended Maximum Likelihood Fit : The use of an extended likelihood function to fit to a sample of data. See section 8.4.1.
Mean : A measure of the central value of a distribution. See section 3.2. The mean is also called the arithmetic mean or average.
Median : A measure of the central value of a distribution. See section 3.2. The median of some distribution in a variable x is the value of x that corresponds to the mid-point in the distribution where there are equal numbers of data both above and below that point.
Mode : A measure of the central value of a distribution. See section 3.2. The mode of some distribution is the most frequently occurring value. If the data in a distribution are binned as a histogram, then the mode will correspond to the bin that contains the largest number of entries.
Multi-Layer Perceptron : A type of artificial neural network based on more than one layer of perceptrons, with several input variables, and one or more output variable. See section 9.4.2.
Monte Carlo : Monte Carlo is the term given to a technique of using sequences of random numbers in order to simulate or model a given situation.
Null Hypothesis (H_0) : The default hypothesis used to test against a sample of data. Such a test will attempt to ascertain if the null hypothesis is compatible with the data.
Optimisation : The process of varying the parameters of some model or algorithm in order to minimise or maximise some test statistic. See section 8.1.
***p*-value** : This is the probability for obtaining a result corresponding to that observed in data, or a more extreme fluctuation.
Poisson Distribution : The Poisson distribution is a probability density function used to describe processes where there is no way of knowing how many events have occurred. This distribution has the property that the mean and variance are the same value. See section 4.3.
Perceptron : A perceptron is a logical unit that is used to process a set of input values, and compare against some threshold function in order to determine a given output. See section 9.4.1.
Probability : A measure of the chance of something happening. This concept is discussed in chapter 2.
Probability Density Function (PDF) : A function used to describe a distribution of some data. The total integral over the interesting problem space of a PDF is one. See section 2.5. A number of useful PDFs are described in chapter 4 and appendix B.
Relative Error : The error on a quantity expressed as a fraction or percentage of the value of the quantity.
Rectangle Rule : A numerical integration algorithm (see appendix C.1).
Set : A collection of unique elements. See chapter 1.
Skew : The skew of a distribution is a measure of how asymmetric a distribution is about the mean value. A symmetric distribution has a skew of zero. See section 3.5.
Standard Deviation : A measure of the spread of data in some distribution, denoted by the symbol σ.

See section 3.3.2.
Standard Error on the Mean : See standard deviation.
Statistic : A term ascribed to mean some calculable quantity. For example arithmetic mean, variance, χ^2 are statistics. Note that often physicists refer to the number of data using the same term, however the meaning inferred should be clear from the context.
Statistical Error : The error determined from a measurement of some observable, whose sources are purely the random result performing the measurement. Also known as the statistical uncertainty.
Systematic Error : An error that is the result of some assumption explicitly or implicitly made in a measurement that may arise in the central value measured being different from the that of a perfect measurement using the same data.
Training : The process of determining a set of coefficients used in a multivariate technique in order to distinguish between classes or types of event.
Training Validation : The process of validating the training procedure, and *optimal* coefficients obtained for a multivariate technique.
Trapezium Rule : A numerical integration algorithm (see appendix C.2).
Type I Error : Rejecting the null hypothesis in a test, where in fact the null hypothesis provides the correct conclusion. Thus a type I error is a false negative response.
Type II Error : Accepting the null hypothesis in a test, where in fact the null hypothesis does not provide the correct conclusion. Thus a type II error is a false positive response.
Uncertainty : This is the spread with which it is possible to determine the value of a parameter via some measurement.
Upper limit : If on performing a search for some effect, one fails to obtain a definite indication that an effect may or may not exist, then one can ascribe a numerical bound on the non-existence of the effect. Above this bound, the effect will have been ruled out, and below this bound, the effect may still exist. This bound is called an upper limit, and as with a confidence interval, there is an associated confidence level ascribed to an upper limit. See section 6.2.
Variance : The square of the standard deviation.
Venn diagram : A graphical way of describing sets. See chapter 1.

Appendix B

Probability Density Functions

This appendix summarises the functional forms of some commonly used probability density functions (PDFs) and follows on from chapter 4. In the following the functional form of the distribution is introduced, and is accompanied with a graphical representation of the PDFs. Several of the PDFs described are specific to problems relating to physics, such as resonant behaviour, and specialist particle physics applications. Where appropriate, there is some elaboration of the physically relevant PDFs for completeness. If there are numerical precision issues relating to the computation of a particular shape, or determination of one or more parameters when used in an optimisation procedure this is noted. Similarly some of the functional forms quoted are not normalised to unity over all space.

While the functional forms discussed here can be used to construct complicated models to represent or test data, sometimes it is necessary to combine PDFs in order to use a composite PDF that is the sum of several parts. This can be achieved by adding together different functional forms with an appropriate weighting factor for each PDF in order to retain a total probability of unity, i.e.:

$$\mathcal{P} = f_1\mathcal{P}_1 + f_2\mathcal{P}_2 + \ldots \left(1 - \sum_{i=1}^{n-1} f_i\right)\mathcal{P}_n, \tag{B.0.1}$$

where the sum of the coefficients is unity. If each of the $\mathcal{P}_i$ are normalised, then the total PDF $\mathcal{P}$ will be properly normalised. If this is not the case, then one has to analytically or numerically integrate $\mathcal{P}$ over the domain of interest to obtain an appropriate normalisation constant.

B.1 Parametric PDFs

Most PDFs used in modelling physical situations are parametric PDFs. Such PDFs depend on one or more discriminating variables represented by either a scalar x or vector $\underline{x}$, and the specific shape of the PDF can be modified by changing one or more parameters $\underline{p}$. In general terms we can write a PDF that depends on several discriminating variables, and parameters as $\mathcal{P}(\underline{x};\underline{p})$. In order for the function to represent a PDF, it must satisfy the normalisation condition

$$\int_{\underline{A}}^{\underline{B}} \mathcal{P}(\underline{x};\underline{p})\mathrm{d}\underline{x} = 1, \tag{B.1.1}$$

where the limits $\underline{A}$ and $\underline{B}$ represent the physically interesting space of the discriminating variables that one wants to use. It follows that for a one-dimensional PDF, we have

$$\int_{A}^{B} \mathcal{P}(x;\underline{p})dx = 1, \tag{B.1.2}$$

where in general we assume that the PDF may depend on several parameters. If a PDF is not naturally normalised to conserve probability, then one has to determine the corresponding normalisation constant required in order to maintain the above relationship. If it is possible to analytically determine a normalisation constant, then it is advisable to do so as this will generally be more efficient (in terms of computing resources) than performing an accurate numerical integration of a function (see appendix C).

The systematic uncertainty associated with the use of a parametric PDF when extracting some quantity from data comes from both the lack of knowledge of parameters, as well as the dependence on the choice of the PDF itself. If the chosen PDF agrees well with the data, then it is likely that using another very similar PDF will produce essentially the same result, and that this second systematic uncertainty is small, or negligible. Uncertainties arising from the lack of knowledge can be addressed by computing the deviation of a result obtained when varying some parameter p_i by its uncertainty, $\pm 1\sigma_{p_i}$. For uncorrelated parameters, the sum in quadrature of the p_i's provides the systematic uncertainty from this source. This approach overestimates the systematic uncertainty for correlated parameters.

B.1.1 Binomial

The binomial distribution (section 4.2) for n trials is given by

$$P(r;p,n) = p^r(1-p)^{n-r}\frac{n!}{r!(n-r)!}, \tag{B.1.3}$$

where r trials are successful and the probability of success is given by p.

B.1.2 Breit-Wigner

The ***Breit-Wigner*** distribution (also called a non-relativistic Breit-Wigner or Cauchy function) is given by

$$\mathcal{P}(x\,;m_0,\Gamma) = \frac{1}{\pi}\frac{\Gamma}{(x-m_0)^2+(\Gamma)^2} \tag{B.1.4}$$

and is shown in Figure B.1. Here m_0 is the position of the peak and Γ is the width of the peak. By construction the distribution is symmetric about the mean value m_0, with a characteristic width given by Γ. This is a properly normalised PDF where

$$\int_{-\infty}^{+\infty} \mathcal{P}(x\,;m_0,\Gamma)dx = 1. \tag{B.1.5}$$

The Breit-Wigner distribution is often used to describe resonant behaviour in nature, for example that exhibited in electronic circuits such as the LCR circuit, and in particle or nuclear interactions where scattering of particles will be enhanced at a resonance[1]. A detailed introduction of the physical significance of this PDF can be found elsewhere, for example in the book (Feynman Leighton and Sands, 1989).

[1] Although it should be noted that modified forms of the Breit-Wigner exist to describe resonances where relativistic factors

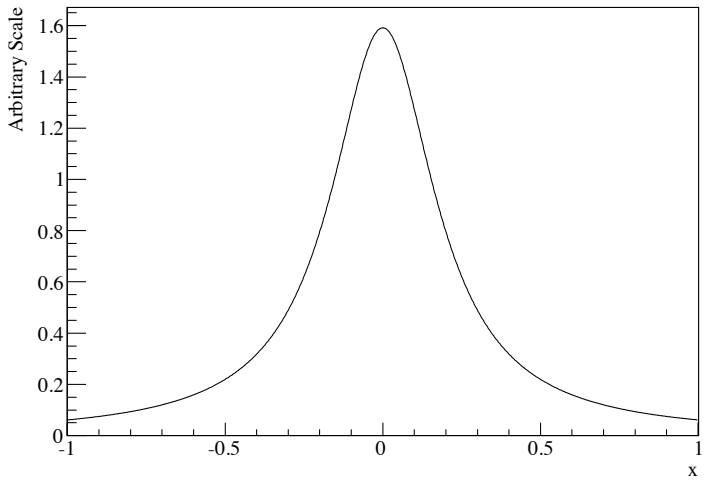

Figure B.1: An example of the Breit-Wigner (Cauchy) PDF.

B.1.3 χ^2

This distribution is defined by χ^2 and the number of degrees of freedom ν, which is given by the number of data N less the number of constraints c imposed. The minimum value for c is one, which is the constraint arising from the total number of data being used in the comparison. The χ^2 distribution has the following form

$$P(\chi^2,\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}(\chi^2)^{(\nu/2-1)}e^{-\chi^2/2}, \tag{B.1.6}$$

where

$$\Gamma(\nu/2) = (\nu/2-1)!, \tag{B.1.7}$$

for positive integer values of ν, and $\Gamma(1/2) = \sqrt{\pi}$. This distribution is described in section 4.5.

B.1.4 Exponential

The ***exponential distribution*** defined as

$$\mathcal{P}(x;\gamma) = \frac{1}{N}e^{\gamma x}, \tag{B.1.8}$$

where γ is the slope (or constant) of the exponential (see Fig. B.2). Often when dealing with the behaviour of physical processes as a function of time (or some other variable) it can be useful to replace the constant γ with $1/\tau$, where the parameter τ is the characteristic decay constant (for negative values of γ) with the same

become important, or where non-zero spin quantum numbers describe the underlying dynamics, for example see the Dalitz plot analysis formalism review compiled by the Particle Data Group (Beringer *et al.*, 2012). In general relativistic Breit-Wigner distributions are asymmetric about the modal value.

units as the variable x. The normalisation constant N is simply given by the integral of the exponential over the desired range in x from A to B, i.e.

$$N = \frac{e^{\gamma B} - e^{\gamma A}}{\gamma}. \tag{B.1.9}$$

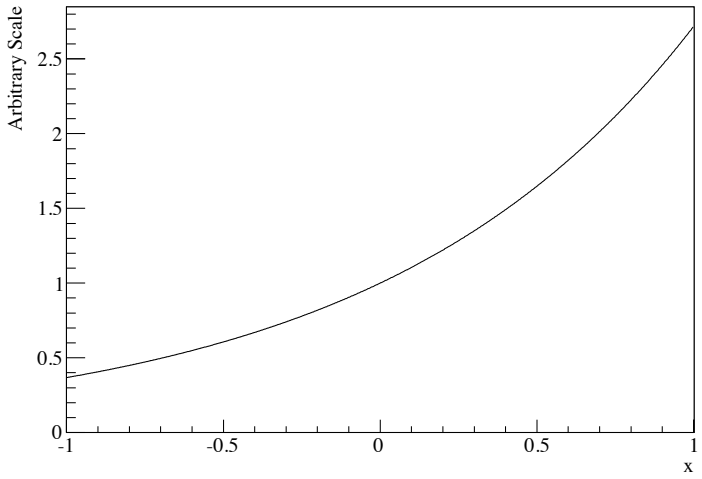

Figure B.2: An example of the exponential PDF with a positive value of γ.

B.1.5 Gaussian

This is defined by a mean and width (μ and σ) and is given by

$$\mathcal{P}(x\,;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-[x-\mu]^2/2\sigma^2\right) \tag{B.1.10}$$

where the integral from $-\infty$ to $+\infty$ is one. The Gaussian PDF is described in detail section 4.4, and the multivariate normal distribution is discussed in section 6.8.1. An asymmetric Gaussian distribution is a Gaussian distribution where the values of σ are allowed to differ above and below $x = \mu$.

B.1.6 Gaussian with an exponential tail

A Gaussian with an exponential tail can be used to parameterise data that has a core part of the distribution resembling a Gaussian, but one that also has a significant skew. In limited circles this distribution is also known as the "Crystal Ball" distribution (Gaiser, 1982; Oreglia, 1980; Skwarnicki, 1986). The functional form of this distribution is (see Fig. B.3)

$$\mathcal{P}(x; m_0, \sigma, \alpha, n) = \begin{cases} \frac{1}{N} \times e^{-(x-x_0)^2/(2\sigma^2)} & x > x_0 - \alpha\sigma, \\ \frac{1}{N} \times \frac{(n/\alpha)^n \exp(-\alpha^2/2)}{((x_0-x)/\sigma + n/\alpha - \alpha)^n} & x \leq x_0 - \alpha\sigma. \end{cases} \tag{B.1.11}$$

By allowing the parameters α and n to vary when determining the PDF, one will find that these parameters are all highly correlated. If there is some way of determining a suitable value for one or both of these

parameters using prior information, that will result in a faster convergence of any fit that is performed using this PDF. The normalisation constant N is given by the integral of the PDF over the problem domain in x.

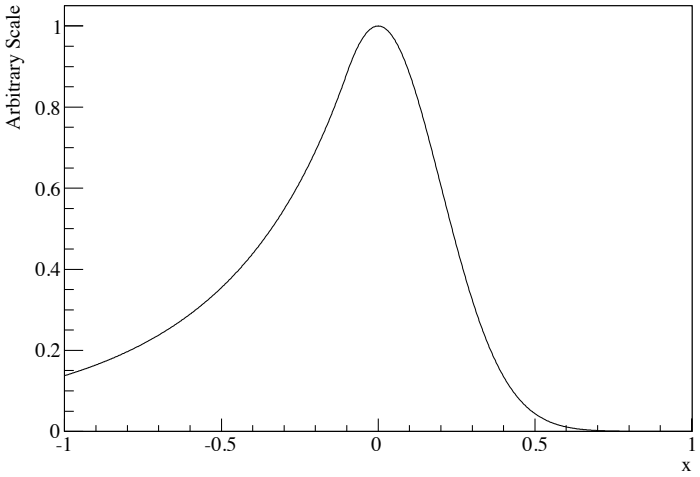

Figure B.3: An example of a Gaussian with an exponential tail (Crystal Ball) PDF.

B.1.7 Landau

This is an important PDF, in particular, for nuclear and particle physicists. The functional form of the Landau distribution is given in Ref. (Landau, 1944) and this is used to describe the fluctuations in energy loss of a charged particle passing through a thin layer of material. Numerically this PDF is time consuming to compute, and there are a number of algorithms that have been used over the years in order to evaluate an estimate of the Landau PDF.

Figure B.4 shows an example of the Landau PDF, which is given by

$$\phi(\lambda) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} e^{u\ln u+\lambda u}du, \tag{B.1.12}$$

where there are two parameters describing the distribution, the most probable value given by λ and a width parameter $\sigma \geq 0$. The underlying numerical algorithm used in order to compute the distribution shown in the figure is described by Kölbig and Schorr (1984). A modern discussion of the computation of this function, with references to earlier works can be found in Fanchiotti Garcia-Canal and Marucho (2006).

B.1.8 Poisson

The Poisson distribution (section 4.3) for r observed events is given by

$$P(r,\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}, \tag{B.1.13}$$

where λ is the mean and variance of the distribution.

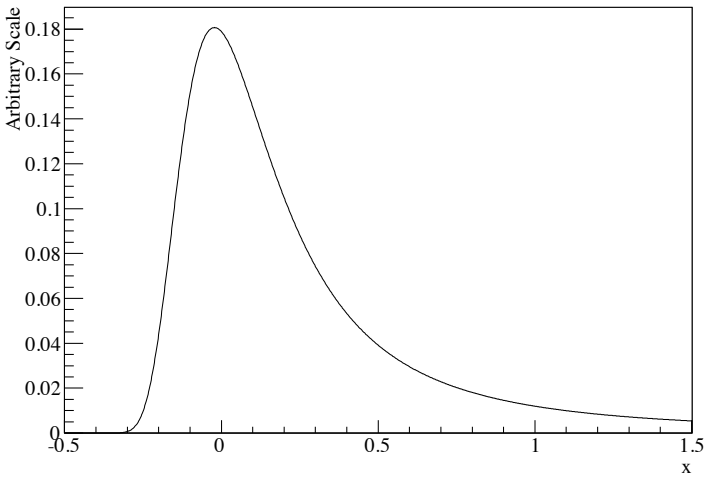

Figure B.4: An example of the Landau PDF with a most probable value of $\lambda = 0.0$, and width parameter $\sigma = 0.1$.

B.1.9 Polynomial

One way to define a ***polynomial distribution*** is given by

$$\mathcal{P}(x; p_k) = \frac{1}{N} \sum_{i=0}^{M} p_k x^k, \tag{B.1.14}$$

where the p_k are coefficients. The normalisation constant N is given by the integral over the interesting domain $A \leq x \leq B$ for the sum

$$\int_A^B \mathcal{P}(x; p_k) dx \quad = \quad \sum_{k=1}^{M} k \times p_k x^{k-1} = N. \tag{B.1.15}$$

In general the coefficients p_k will be correlated with each other. In particular the odd terms (odd k) will be strongly correlated with each other, and the even coefficients (even k) will be strongly correlated with each other. For this reason, it may be desirable to limit the order of a polynomial, or for example only use odd (even) terms for odd (even) distributions of events in order to limit the number of parameters allowed to vary when fitting data. Figure B.5 shows an example of a cubic function. It is possible to define a polynomial distribution in other ways, so if you are using a particular representation, ensure that you use this consistently. Furthermore, if you find that the polynomial coefficients p_k are highly correlated with each other for a particular problem, you may find using an alternative definition of a polynomial function, for example one formed from a set of orthogonal terms, to be a better choice of PDF. Examples of orthogonal polynomials include Legendre, Chebychev, and Hermite polynomials.

One final thing to remember, if a polynomial function is considered a suitable PDF that is related to the probability of some event to occur, then this PDF needs to be non-negative over the relevant domain. This follows as probability is a positive definite quantity bound by zero and one.

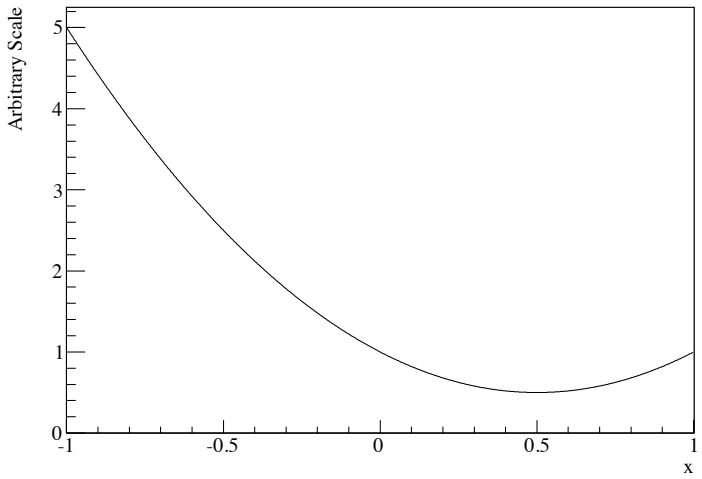

Figure B.5: An example of the polynomial PDF.

B.1.10 Sigmoid

The sigmoid (or logistic) function is often used in machine learning algorithms or to model behaviour at some threshold. The sigmoid function is given by

$$\mathcal{P}(x;a,b) = \frac{1}{N}\frac{1}{1+e^{ax+b}}, \tag{B.1.16}$$

where the coefficients a and b are an exponent scale factor and x-offset, respectively. The factor N is a normalisation constant that has to be determined for the problem domain in x from x_1 to x_2. This can be obtained by substituting a variable for the denominator and integrating Eq. (B.1.16) using partial fractions. The result obtained for the normalisation is

$$N = \left[x - \frac{\ln|1+e^{ax+b}|}{a}\right]_{x_1}^{x_2}. \tag{B.1.17}$$

Figure B.6 shows an example of the sigmoid PDF.

B.1.11 Step and veto functions

It can be useful to impose a sharp cut-off to a PDF. In such cases it is useful use a ***step function*** PDF where

$$\mathcal{P}(x) = \begin{cases} \frac{1}{N}, & x_a < x < x_b \\ 0, & x < x_a \text{ or } x > x_b \end{cases} \tag{B.1.18}$$

The normalisation N for a step function depends on the values of x_a and x_b, as can be seen from the following:

$$\int_{x_a}^{x_b} \mathcal{P}(x)dx = \int_{x_a}^{x_b} \frac{1}{N}dx = \frac{(x_b - x_a)}{N} = 1. \tag{B.1.19}$$

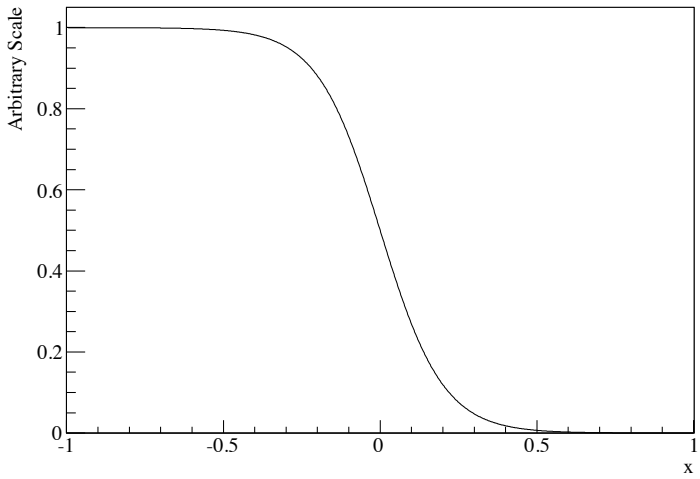

Figure B.6: An example of the sigmoid PDF.

Thus in order to satisfy the normalisation condition that the total probability is one, $N = x_b - x_a$, thus

$$\mathcal{P}(x) = \frac{1}{(x_b - x_a)}, \tag{B.1.20}$$

Figure B.7 shows an example of the step function where $x_a = -0.3$ and $x_b = 0.3$.

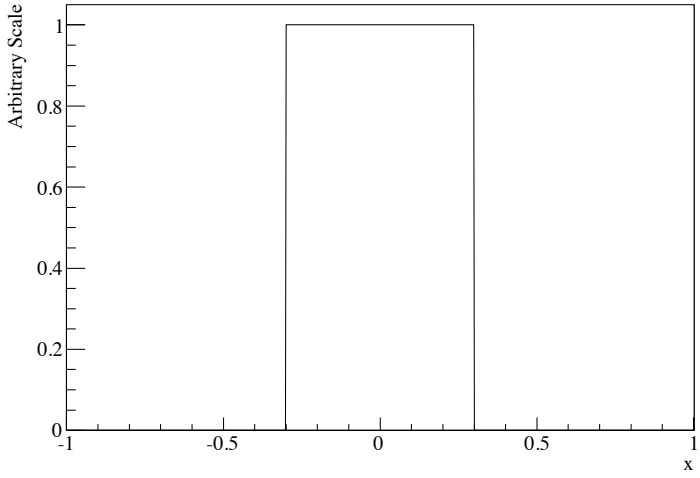

Figure B.7: An example of a step function.

The complement of a step function is a ***veto function***, where the PDF has a value of zero for finite range $x_a < x < x_b$, and is otherwise constant. Thus the functional form of the veto function is

$$\mathcal{P}(x) = \begin{cases} \frac{1}{N}, & x < x_a \text{ or } x > x_b \\ 0, & x_a < x < x_b \end{cases}. \tag{B.1.21}$$

The normalisation for the veto function depends not only on x_a and x_b, but also on the domain boundaries of interest A and B. Thus

$$\int_A^B \mathcal{P}(x)dx = \int_A^{x_a} \frac{1}{N}dx + \int_{x_b}^B \frac{1}{N}dx \quad \text{(B.1.22)}$$

$$= \frac{x_a - A + B - x_b}{N}, \quad \text{(B.1.23)}$$

hence in order for this PDF to be properly normalised we require $N = x_a - A + B - x_b$.

B.2 Non-parametric PDFs

Those PDFs where there are no parameters $\underline{p}$ used to determine the shape are called ***non-parametric*** PDFs. For such functions, the PDF is completely pre-determined and can not vary in an optimisation process. Such PDFs can be useful when control sample data show complicated behaviour that would be difficult to parameterise in a sensible way. Several examples of non-parametric PDFs are discussed in the following. The PDF normalisation constants for non-parametric PDFs are generally determined numerically (see appendix C for a simple introduction to this topic). If the non-parametric PDF is a binned PDF, then the normalisation can be determined by simply summing the weights of each bin in that distribution.

It should be noted that unlike parametric PDFs, where one can compute systematic uncertainties on an extracted quantity from the lack of knowledge of parameters, as well as dependence on the choice of PDF, the systematic uncertainty associated with the use of non-parametric PDFs arises from dependence on the choice of PDF and binning. Care should be taken to address this issue properly when analysing data.

B.2.1 Histogram

It is possible to construct a non-parametric PDF based on an input histogram. Such a PDF will have a pre-determined level as a function of the discriminating variable. The function will take the form

$$\mathcal{P}(x) = B(x), \quad \text{(B.2.1)}$$

where B is a discrete function with a constant value across a bin interval. Figure B.8 shows an example of a histogram PDF.

While the content of a histogram is pre-determined but discrete, it may be considered undesirable to have a discrete distribution as a PDF. One can interpolate between the bin content of adjacent bins to determine a more smoothly varying form for $\mathcal{P}$ at the bin boundaries. Such a process is referred to as smoothing a histogram. In practice this finesse usually results in an improved visual representation of the PDF, while the fitted result obtained is very similar to the original one.

While the histogram shape can be well defined given sufficient prior knowledge such as a control sample of data, or a sample of Monte Carlo simulated data, there are some instances when one does not have the luxury of an independent control sample to fix the PDF. One possible variant on the histogram PDF that can be used for such situation is the so-called parametric step function (PSF) (Aubert *et al.*, 2003). This is a PDF where data are binned in one or more dimensions, and the parameters of the PDF are the fractions of the total number of events in each bin. As binning can be non-uniform by definition, the fitted fractions are, in general, not the heights of each bin when plotted, but also depend on the bin width.

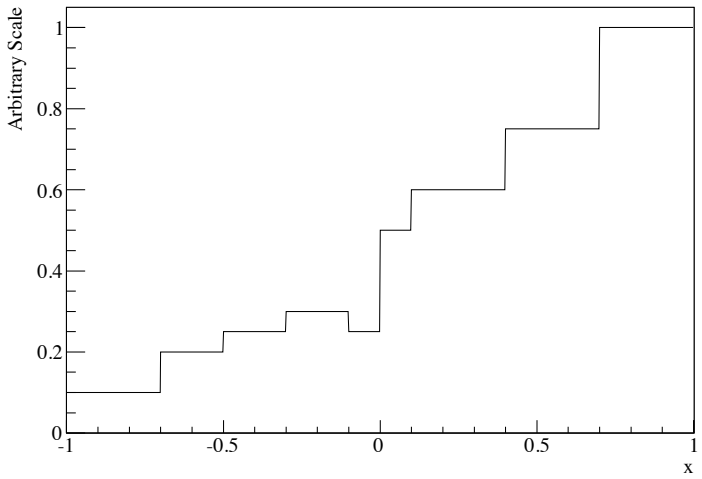

Figure B.8: An example of a histogram PDF.

B.2.2 Kernal estimation PDF

A ***Kernal Estimation*** (KE) algorithm, for example see (Cranmer, 2001; Hastie Tibshirani and Friedman, 2009) can be used to obtain a smoothly varying non-parametric PDF representation of a sample of data or Monte Carlo simulated data. The PDF is constructed from individual events used to represent a particular type of data, for example Monte Carlo simulated reference samples, or control samples of data. The individual events are referred to as kernals. Given a reference sample of data, one can define the properties of a kernal to be used in order to construct the PDF. A narrow step function would correspond to a shape similar to a histogram, however a smoothly varying shape, such as a Gaussian could be used to provide a smooth non-parametric shape. For such a PDF the mean of the Gaussian is given by the point in space of i^{th} event $e_i(\underline{x})$, and the width of the Gaussian kernal is taken from some representation of the data, for example the RMS of the data, some multiple of the RMS, or the RMS of a local set of the data. Given that each event in the reference sample contributes to the total PDF, one has to numerically integrate a KE PDF in order to normalise the total probability to unity.

A KE PDF for some data set Ω with N elements, with Gaussian kernals is given by

$$\mathcal{P}(x) = A\sum_{i=1}^{N} \frac{1}{\lambda\sqrt{2\pi}} e^{-(x-x_i)^2/2\lambda^2}, \tag{B.2.2}$$

where A is a normalisation coefficient, and λ is the kernal width, which is related to the σ of the data. In general $\lambda = a\sigma$, where a is a scale factor. This is set to one in order to use a kernal width corresponding to the RMS of Ω. The kernals can be made narrower ($a < 1$) or wider ($a > 1$) as desired in order to compute a PDF that is a better representation of the data. The use of different kernal elements is discussed in Hastie Tibshirani and Friedman (2009).

The use case of the KE PDF is similar to that of the histogram PDF described above. The main difference between using a histogram PDF and a KE PDF is that the latter requires a lot more computation than the former. The reason for this is that a Gaussian kernal is computed for each event in the data set used to construct a KE PDF, and the PDF is evaluated at each point by computing a sum over all events. In practice it is usually beneficial to compute a KE PDF once and then store the output shape as a histogram, or in a look-up-table, for use in all later fitting and validations. In essence the KE PDF is a smoothed representation of reference data. In contrast to a histogram PDF, where data are binned, the KE PDF

constructs its representation of the data on an event-by-event basis.

KE PDFs are ideal for describing smoothly varying distributions that fall to zero at the extremities of the distribution. If this is not the case for a given situation, then one will encounter boundary effects that are associated with the fact that the value of the PDF at a given point in space x has a contribution from all events. Near boundaries, a PDF will only obtain contributions from events that are within the physical domain of the reference kernals being used to construct it, and hence the PDF will underestimate the distribution of events used to generate it. There are two ways to mitigate boundary effects in KE PDFs. Firstly the contributions from kernals near the boundary can be reflected at the boundary in order to compensate for missing contributions. This reflection or mirroring approach will tend to over-estimate the PDF at boundaries. A second possible way to mitigate boundary effects with a KE PDF is to provide events that go beyond the discriminating variable range required. This way the boundary effect is moved to the edges of the range of data being used, which is no longer in the valid range of the discriminating variable that one is trying to describe. Unfortunately the second approach is not always usable.

In general KE PDFs are numerically resource hungry, and a minimum amount of kernal data is required in order to obtain a reasonable representation of the distribution. How much kernal data is required depends on the distribution being approximated, and the dimensionality of the problem. Thus far it has implicitly been assumed that the KE PDF is a one-dimensional PDF, however one can extend the algorithm to an arbitrary number of dimensions. The penalties for doing this are that the computing power and the number of kernal data N required to compute the PDF increases considerably for each dimension added. This PDF suffers from the curse of dimensionality described by Bellman (1961).

B.2.3 Uniform distribution

This is a special case of the step function PDF described above and shown in Figure B.7. Consider the situation where the step function limits x_a and x_b coincide with the domain boundaries A and B in the problem. As a result of allowing the step function boundaries to match the limits of the interesting problem space, there are no intrinsic parameters required to determine the form of this PDF. Thus we can write

$$\mathcal{P}(x) = \frac{1}{B-A}, \tag{B.2.3}$$

where it can be shown that the normalisation condition for the PDF is satisfied as

$$\int_A^B \mathcal{P}(x)dx = \int_A^B \frac{1}{B-A}dx = 1. \tag{B.2.4}$$

If either A or B tends to infinity the normalisation of this PDF tends to zero. Hence, a uniform distribution is only well defined for some finite interval $x \in [A, B]$.

Appendix C

Numerical integration methods

This section provides a brief introduction to numerical integration techniques. A general approach that can be used for complicated distributions is to adopt a Monte Carlo method to compute a numerical integral. This follows from the discussion of how to compute an upper limit from an arbitrary PDF in section 6.7. One draw back of this approach is that in order to obtain a precise estimate of an integral, one needs to generate a large number of simulated data. Thus if a PDF is easily integrable, it is highly inefficient to try and use Monte Carlo techniques to compute the integral. A number of robust numerical recipes are available to facilitate computation of integrals of single valued functions. Two such algorithms are discussed in this appendix. With modern computers there can be less pressure on an individual to use efficient algorithms to compute numerical integrations, however if an algorithm will be used on a regular basis, it would probably be beneficial to investigate the potential of more sophisticated algorithms than those presented here. Such techniques can be found in texts such as the Numerical Recipes series (Press *et al.*, 2002).

C.1 Rectangle rule

For some function $y = f(x)$, one can numerically determine an approximation for the integral

$$I = \int_{x_a}^{x_b} f(x)dx, \tag{C.1.1}$$

between the limits x_a and x_b. In the limit where dx becomes large, this integral approaches the approximation

$$I \simeq \sum_{i=1}^{n} f(x_i + \Delta x/2)\Delta x, \tag{C.1.2}$$

which is known as the ***rectangle rule***. Here Δx is the finite interval corresponding to the infinitesimal element dx and the sum is over n bins in x. Thus if we consider this sum graphically we represent the curve $f(x)$ by a set of discrete rectangles, each of some width Δx along the x direction, centred about the mid-point of a bin, $x + \Delta x/2$. This rectangle rule is also known as the mid-point rule, and it is the simplest approximation. If one requires a high level of accuracy in a numerical integral using this rule, then Δx must be small. As such the bin width Δx can be considered as a tuneable parameter.

By tuning or changing the value of Δx, one can estimate the numerical uncertainty on the use of this algorithm. With this information in hand for a given problem, the user can decide if a sufficient level of numerical precision has been achieved or not. One way of determining if a sufficient level of precision has

been reached is to use the result of the numerical integral I_0 with the bin width Δx_0 to whatever end is desired (e.g. computation of a confidence level), and compare this with the corresponding result I_1 when the bin width is $\Delta x_1 = \Delta x_0/2$. If the result obtained in both cases is the same within the desired precision, then the numerical integration performed is sufficient. If not then the value of Δx is too large, and should be reduced further until the desired stability is encountered.

C.2 Trapezium rule

A more sophisticated numerical integration technique can be developed as a result of the lessons learned with the rectangle rule. Each contribution to the rectangle rule is an approximation of the curve that has no sensitivity to the change in gradient of the curve across a given bin i. One modification to this procedure would be to compute the numerical integral from a sum of trapeziums instead of rectangles. The error in this approximation should, in general (but not always), be smaller for each bin than for the rectangle rule, hence one would expect an algorithm based on a sum of trapeziums to compute a more accurate estimate of an integral than the rectangle rule for a given bin width.

The area of a trapezium A_i is given by

$$A_i = \frac{f(x_i) + f(x_{i+1})}{2}\Delta x. \tag{C.2.1}$$

If we consider n bins, then the corresponding area computed by the ***trapezium rule*** for integration is

$$A = \sum_{i=0}^{n} \left[\frac{f(x_i) + f(x_{i+1})}{2} \right] \Delta x. \tag{C.2.2}$$

The error associated with a particular integral can again be estimated by tuning the bin width Δx and observing how the computed integral impacts upon the end result.

C.3 Reflection on the use of numerical integration

Table C.1 summarises the results obtained on numerically integrating a Gaussian distribution between $\mu - 1\sigma$ and $\mu + 1\sigma$ for different numbers of bins, hence a number of different values of Δx. If we are content with obtaining a precision of three significant figures for our result, then the trapezium rule reaches this with 30 bins. In contrast the rectangle rule achieves the same precision with only 10 bins and the integral is stable up to six significant figures when 900,000 bins are used using the rectangle rule. In this instance the more sophisticated algorithm provides a worse performance than the rectangle rule. If one considers the error computed on the integral (normalising to the result obtained with 10^6 bins, then it can be seen that the rectangle rule overestimates the integral and the trapezium rule underestimates the integral.

A second example is illustrated in Table C.2, where the results of performing a numerical integration of the cubic equation

$$y = x^3 + 2x^2 + 1, \tag{C.3.1}$$

are shown for $0 \leq x \leq 1$. Analytically this integral is 23/12. For this particular function the rectangle rule is not as effective as the trapezium rule in computing an accurate estimate of the integral. The integral is stable up to six significant figures when one million steps are used. These two examples illustrate that the user should investigate the suitability and uncertainty obtained with different numerical techniques for each particular problem faced, as the effectiveness of a given numerical integration algorithm can change from

Table C.1: Comparison of numerical integrals of a Gaussian obtained using the rectangle and trapezium rules for different numbers of bins. These integrals are over the range $\mu - 1\sigma$ to $\mu + 1\sigma$. The 3^{rd} and 5^{th} columns report the estimated error (ϵ) on the numerical integral for the two methods.

Number of bins	Rectangle rule	ϵ	Trapezium rule	ϵ
5	0.685946	3.3×10^{-3}	0.676202	-6.5×10^{-3}
10	0.683498	8.1×10^{-4}	0.681074	-1.6×10^{-3}
50	0.682722	3.3×10^{-5}	0.682625	-6.4×10^{-5}
100	0.682698	8.6×10^{-6}	0.682673	-1.6×10^{-5}
10^6	0.682689	–	0.682689	–

Table C.2: Comparison of numerical integrals of the cubic equation, Eq. (C.3.1), obtained using the rectangle and trapezium rules for different numbers of bins. These integrals are over the range $0 \leq x \leq 1$. The 3^{rd} and 5^{th} columns report the estimated error (ϵ) on the numerical integral for the two methods.

Number of bins	Rectangle rule	ϵ	Trapezium rule	ϵ
5	1.64000	-0.277	1.94000	0.023
10	1.77250	-0.144	1.92250	0.006
50	1.88690	-0.030	1.91690	2.3×10^{-4}
100	1.90172	-0.015	1.91672	5.8×10^{-5}
10^6	1.91667	-1.5×10^{-7}	1.91667	-1.5×10^{-13}

case to case. It should be noted that if it is possible to perform the integral analytically then one should endeavour to do so.

One last thing to consider about numerical integration techniques is that the error on a computation is related to how quickly the gradient is changing over the width of a single bin. If you have to numerically integrate a function that has well defined structure where parts of the function are smoothly varying and others change rapidly, then there may be advantages in trying to use an adaptive bin width. For example, the error penalty for using a large bin width when the function is varying smoothly may be relatively small, and by using a large bin width for such areas of the function you may be able to spend more CPU time computing the integral with a small bin width where that is more appropriate.

Appendix D

Reference tables

This appendix provides a number of quick-reference tables that can be used to compute probabilities based on the distributions introduced in chapter 4. A more extensive set of information can be found in Lindley and Scott (1995).

D.1 Binomial probability

Section 4.2 introduced the binomial probability distribution, given by

$$P(r; p, n) = p^r (1-p)^{n-r} \frac{n!}{r!(n-r)!}, \tag{D.1.1}$$

for a given number of trials n with r successful outcomes, where each successful outcome has a probability given by p. Tables D.1 through D.5 summarise the cumulative probabilities for binomial distributions with $n = 2, 3, 4, 5$, and 10, and p values between 0.1 and 0.9.

Table D.1: Cumulative probability table for a Binomial distribution with values of p specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column. This table is for n =2.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.81000	0.64000	0.49000	0.36000	0.25000	0.16000	0.09000	0.04000	0.01000
1	0.99000	0.96000	0.91000	0.84000	0.75000	0.64000	0.51000	0.36000	0.19000
2	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table D.2: Cumulative probability table for a Binomial distribution with values of p specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column. This table is for n =3.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.72900	0.51200	0.34300	0.21600	0.12500	0.06400	0.02700	0.00800	0.00100
1	0.97200	0.89600	0.78400	0.64800	0.50000	0.35200	0.21600	0.10400	0.02800
2	0.99900	0.99200	0.97300	0.93600	0.87500	0.78400	0.65700	0.48800	0.27100
3	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table D.3: Cumulative probability table for a Binomial distribution with values of p specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column. This table is for n =4.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.65610	0.40960	0.24010	0.12960	0.06250	0.02560	0.00810	0.00160	0.00010
1	0.94770	0.81920	0.65170	0.47520	0.31250	0.17920	0.08370	0.02720	0.00370
2	0.99630	0.97280	0.91630	0.82080	0.68750	0.52480	0.34830	0.18080	0.05230
3	0.99990	0.99840	0.99190	0.97440	0.93750	0.87040	0.75990	0.59040	0.34390
4	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table D.4: Cumulative probability table for a Binomial distribution with values of p specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column. This table is for n =5.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.59049	0.32768	0.16807	0.07776	0.03125	0.01024	0.00243	0.00032	0.00001
1	0.91854	0.73728	0.52822	0.33696	0.18750	0.08704	0.03078	0.00672	0.00046
2	0.99144	0.94208	0.83692	0.68256	0.50000	0.31744	0.16308	0.05792	0.00856
3	0.99954	0.99328	0.96922	0.91296	0.81250	0.66304	0.47178	0.26272	0.08146
4	0.99999	0.99968	0.99757	0.98976	0.96875	0.92224	0.83193	0.67232	0.40951
5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table D.5: Cumulative probability table for a Binomial distribution with values of p specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column. This table is for n =10.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.34868	0.10737	0.02825	0.00605	0.00098	0.00010	0.00001	0.00000	0.00000
1	0.73610	0.37581	0.14931	0.04636	0.01074	0.00168	0.00014	0.00000	0.00000
2	0.92981	0.67780	0.38278	0.16729	0.05469	0.01229	0.00159	0.00008	0.00000
3	0.98720	0.87913	0.64961	0.38228	0.17188	0.05476	0.01059	0.00086	0.00001
4	0.99837	0.96721	0.84973	0.63310	0.37695	0.16624	0.04735	0.00637	0.00015
5	0.99985	0.99363	0.95265	0.83376	0.62305	0.36690	0.15027	0.03279	0.00163
6	0.99999	0.99914	0.98941	0.94524	0.82812	0.61772	0.35039	0.12087	0.01280
7	1.00000	0.99992	0.99841	0.98771	0.94531	0.83271	0.61722	0.32220	0.07019
8	1.00000	1.00000	0.99986	0.99832	0.98926	0.95364	0.85069	0.62419	0.26390
9	1.00000	1.00000	0.99999	0.99990	0.99902	0.99395	0.97175	0.89263	0.65132
10	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

D.2 Poisson probability

Section 4.3 introduced the Poisson probability distribution, given by

$$P(r,\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}. \tag{D.2.1}$$

Poisson probabilities $P(r,\lambda)$ for integer values of λ from one to ten and values of r up to twenty are summarised in Table D.6. The first column indicates the value of r, and the top row indicates the value of λ considered.

Table D.6: Probability table for a Poisson distribution $P(r,\lambda)$ with values of λ specified at the top of each column, for a given value of r specified in the first column.

	1	2	3	4	5	6	7	8	9	10
0	0.36788	0.13534	0.04979	0.01832	0.00674	0.00248	0.00091	0.00034	0.00012	0.00005
1	0.36788	0.27067	0.14936	0.07326	0.03369	0.01487	0.00638	0.00268	0.00111	0.00045
2	0.18394	0.27067	0.22404	0.14653	0.08422	0.04462	0.02234	0.01073	0.00500	0.00227
3	0.06131	0.18045	0.22404	0.19537	0.14037	0.08924	0.05213	0.02863	0.01499	0.00757
4	0.01533	0.09022	0.16803	0.19537	0.17547	0.13385	0.09123	0.05725	0.03374	0.01892
5	0.00307	0.03609	0.10082	0.15629	0.17547	0.16062	0.12772	0.09160	0.06073	0.03783
6	0.00051	0.01203	0.05041	0.10420	0.14622	0.16062	0.14900	0.12214	0.09109	0.06306
7	0.00007	0.00344	0.02160	0.05954	0.10444	0.13768	0.14900	0.13959	0.11712	0.09008
8	0.00001	0.00086	0.00810	0.02977	0.06528	0.10326	0.13038	0.13959	0.13176	0.11260
9	0.00000	0.00019	0.00270	0.01323	0.03627	0.06884	0.10140	0.12408	0.13176	0.12511
10	0.00000	0.00004	0.00081	0.00529	0.01813	0.04130	0.07098	0.09926	0.11858	0.12511

The corresponding table for cumulative probabilities (summing up from $r = 0$ to the specified value) given by

$$P = \sum_{r=0}^{r=n} P(r,\lambda), \tag{D.2.2}$$

can be found in Table D.7.

Table D.7: Cumulative probability table for a Poisson distribution with values of λ specified at the top of each column, summing up from $r = 0$ to the r value specified in the first column.

	1	2	3	4	5	6	7	8	9	10
0	0.36788	0.13534	0.04979	0.01832	0.00674	0.00248	0.00091	0.00034	0.00012	0.00005
1	0.73576	0.40601	0.19915	0.09158	0.04043	0.01735	0.00730	0.00302	0.00123	0.00050
2	0.91970	0.67668	0.42319	0.23810	0.12465	0.06197	0.02964	0.01375	0.00623	0.00277
3	0.98101	0.85712	0.64723	0.43347	0.26503	0.15120	0.08177	0.04238	0.02123	0.01034
4	0.99634	0.94735	0.81526	0.62884	0.44049	0.28506	0.17299	0.09963	0.05496	0.02925
5	0.99941	0.98344	0.91608	0.78513	0.61596	0.44568	0.30071	0.19124	0.11569	0.06709
6	0.99992	0.99547	0.96649	0.88933	0.76218	0.60630	0.44971	0.31337	0.20678	0.13014
7	0.99999	0.99890	0.98810	0.94887	0.86663	0.74398	0.59871	0.45296	0.32390	0.22022
8	1.00000	0.99976	0.99620	0.97864	0.93191	0.84724	0.72909	0.59255	0.45565	0.33282
9	1.00000	0.99995	0.99890	0.99187	0.96817	0.91608	0.83050	0.71662	0.58741	0.45793
10	1.00000	0.99999	0.99971	0.99716	0.98630	0.95738	0.90148	0.81589	0.70599	0.58304

D.3 Gaussian probability

Section 4.4 introduced the Gaussian probability distribution, which is given by

$$P(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \tag{D.3.1}$$

One and two sided confidence intervals for the Gaussian probability distribution can be found in Tables D.8 and D.9. The one sided confidence interval is given by

$$P = \int_{-\infty}^{n\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx, \tag{D.3.2}$$

$$= \int_{-\infty}^{n} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \tag{D.3.3}$$

where the lower limit is $-\infty$ and the upper limit is the number of σ above the mean value. The table is given in terms of $z = (x-\mu)/\sigma$, where the first column indicates the number of σ up to one decimal place, and the first row indicates the number of σ at the level of the second decimal place.

For example, if one is interested in the probability corresponding to a one sided Gaussian interval from $-\infty$ to $\mu + 1.55\sigma$, then this would be given by 0.9394, which corresponds to the element in the 0.05 column and row marked with 1.5 in Table D.8.

The two sided confidence interval is given by

$$P = \int_{-n\sigma}^{n\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx, \tag{D.3.4}$$

$$= \int_{-n}^{n} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \tag{D.3.5}$$

where n defines the number of σ over which the integral is performed. For example, if one is interested in the probability corresponding to the two sided Gaussian interval between $\mu - 1.64\sigma$ and $\mu + 1.64\sigma$, then this would correspond to 0.8990, which is the table element in the 0.04 column and the row marked with 1.6 in Table D.9.

Table D.8: One sided probability table for a Gaussian distribution in terms of the number of standard deviations from the mean value for $z > 0$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Table D.9: Two sided probability table for a Gaussian distribution in terms of the number of standard deviations from the mean value.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00798	0.01596	0.02393	0.03191	0.03988	0.04784	0.05581	0.06376	0.07171
0.1	0.07966	0.08759	0.09552	0.10343	0.11134	0.11924	0.12712	0.13499	0.14285	0.15069
0.2	0.15852	0.16633	0.17413	0.18191	0.18967	0.19741	0.20514	0.21284	0.22052	0.22818
0.3	0.23582	0.24344	0.25103	0.25860	0.26614	0.27366	0.28115	0.28862	0.29605	0.30346
0.4	0.31084	0.31819	0.32551	0.33280	0.34006	0.34729	0.35448	0.36164	0.36877	0.37587
0.5	0.38292	0.38995	0.39694	0.40389	0.41080	0.41768	0.42452	0.43132	0.43809	0.44481
0.6	0.45149	0.45814	0.46474	0.47131	0.47783	0.48431	0.49075	0.49714	0.50350	0.50981
0.7	0.51607	0.52230	0.52848	0.53461	0.54070	0.54675	0.55275	0.55870	0.56461	0.57047
0.8	0.57629	0.58206	0.58778	0.59346	0.59909	0.60467	0.61021	0.61570	0.62114	0.62653
0.9	0.63188	0.63718	0.64243	0.64763	0.65278	0.65789	0.66294	0.66795	0.67291	0.67783
1.0	0.68269	0.68750	0.69227	0.69699	0.70166	0.70628	0.71086	0.71538	0.71986	0.72429
1.1	0.72867	0.73300	0.73729	0.74152	0.74571	0.74986	0.75395	0.75800	0.76200	0.76595
1.2	0.76986	0.77372	0.77754	0.78130	0.78502	0.78870	0.79233	0.79592	0.79945	0.80295
1.3	0.80640	0.80980	0.81316	0.81648	0.81975	0.82298	0.82617	0.82931	0.83241	0.83547
1.4	0.83849	0.84146	0.84439	0.84728	0.85013	0.85294	0.85571	0.85844	0.86113	0.86378
1.5	0.86639	0.86896	0.87149	0.87398	0.87644	0.87886	0.88124	0.88358	0.88589	0.88817
1.6	0.89040	0.89260	0.89477	0.89690	0.89899	0.90106	0.90309	0.90508	0.90704	0.90897
1.7	0.91087	0.91273	0.91457	0.91637	0.91814	0.91988	0.92159	0.92327	0.92492	0.92655
1.8	0.92814	0.92970	0.93124	0.93275	0.93423	0.93569	0.93711	0.93852	0.93989	0.94124
1.9	0.94257	0.94387	0.94514	0.94639	0.94762	0.94882	0.95000	0.95116	0.95230	0.95341
2.0	0.95450	0.95557	0.95662	0.95764	0.95865	0.95964	0.96060	0.96155	0.96247	0.96338
2.1	0.96427	0.96514	0.96599	0.96683	0.96765	0.96844	0.96923	0.96999	0.97074	0.97148
2.2	0.97219	0.97289	0.97358	0.97425	0.97491	0.97555	0.97618	0.97679	0.97739	0.97798
2.3	0.97855	0.97911	0.97966	0.98019	0.98072	0.98123	0.98173	0.98221	0.98269	0.98315
2.4	0.98360	0.98405	0.98448	0.98490	0.98531	0.98571	0.98611	0.98649	0.98686	0.98723
2.5	0.98758	0.98793	0.98826	0.98859	0.98891	0.98923	0.98953	0.98983	0.99012	0.99040
2.6	0.99068	0.99095	0.99121	0.99146	0.99171	0.99195	0.99219	0.99241	0.99264	0.99285
2.7	0.99307	0.99327	0.99347	0.99367	0.99386	0.99404	0.99422	0.99439	0.99456	0.99473
2.8	0.99489	0.99505	0.99520	0.99535	0.99549	0.99563	0.99576	0.99590	0.99602	0.99615
2.9	0.99627	0.99639	0.99650	0.99661	0.99672	0.99682	0.99692	0.99702	0.99712	0.99721
3.0	0.99730	0.99739	0.99747	0.99755	0.99763	0.99771	0.99779	0.99786	0.99793	0.99800
3.1	0.99806	0.99813	0.99819	0.99825	0.99831	0.99837	0.99842	0.99848	0.99853	0.99858
3.2	0.99863	0.99867	0.99872	0.99876	0.99880	0.99885	0.99889	0.99892	0.99896	0.99900
3.3	0.99903	0.99907	0.99910	0.99913	0.99916	0.99919	0.99922	0.99925	0.99928	0.99930
3.4	0.99933	0.99935	0.99937	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	0.99952
3.5	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99963	0.99964	0.99966	0.99967
3.6	0.99968	0.99969	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	0.99977	0.99978
3.7	0.99978	0.99979	0.99980	0.99981	0.99982	0.99982	0.99983	0.99984	0.99984	0.99985
3.8	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989	0.99989	0.99990	0.99990
3.9	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99993	0.99993	0.99993	0.99993

D.4 χ^2 probability

The probability for obtaining a given χ^2 for different numbers of degrees of freedom can be found in Tables D.10 and D.11. The χ^2 probability for a given number of degrees of freedom is

$$P(\chi^2, \nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}(\chi^2)^{(\nu/2-1)}e^{-\chi^2/2}, \tag{D.4.1}$$

and is discussed in section 4.5.

Table D.10: χ^2 probability table for observed values of (rows) χ^2 and (columns) degrees of freedom ν.

	1	2	3	4	5	6	7	8	9	10
0.1000	0.7518	0.9512	0.9918	0.9988	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
0.2000	0.6547	0.9048	0.9776	0.9953	0.9991	0.9998	1.0000	1.0000	1.0000	1.0000
0.3000	0.5839	0.8607	0.9600	0.9898	0.9976	0.9995	0.9999	1.0000	1.0000	1.0000
0.4000	0.5271	0.8187	0.9402	0.9825	0.9953	0.9989	0.9997	0.9999	1.0000	1.0000
0.5000	0.4795	0.7788	0.9189	0.9735	0.9921	0.9978	0.9994	0.9999	1.0000	1.0000
0.6000	0.4386	0.7408	0.8964	0.9631	0.9880	0.9964	0.9990	0.9997	0.9999	1.0000
0.7000	0.4028	0.7047	0.8732	0.9513	0.9830	0.9945	0.9983	0.9995	0.9999	1.0000
0.8000	0.3711	0.6703	0.8495	0.9384	0.9770	0.9921	0.9974	0.9992	0.9998	0.9999
0.9000	0.3428	0.6376	0.8254	0.9246	0.9702	0.9891	0.9963	0.9988	0.9996	0.9999
1.0000	0.3173	0.6065	0.8013	0.9098	0.9626	0.9856	0.9948	0.9982	0.9994	0.9998
1.1000	0.2943	0.5769	0.7771	0.8943	0.9541	0.9815	0.9931	0.9975	0.9992	0.9997
1.2000	0.2733	0.5488	0.7530	0.8781	0.9449	0.9769	0.9909	0.9966	0.9988	0.9996
1.3000	0.2542	0.5220	0.7291	0.8614	0.9349	0.9717	0.9884	0.9956	0.9984	0.9994
1.4000	0.2367	0.4966	0.7055	0.8442	0.9243	0.9659	0.9856	0.9942	0.9978	0.9992
1.5000	0.2207	0.4724	0.6823	0.8266	0.9131	0.9595	0.9823	0.9927	0.9971	0.9989
1.6000	0.2059	0.4493	0.6594	0.8088	0.9012	0.9526	0.9786	0.9909	0.9963	0.9986
1.7000	0.1923	0.4274	0.6369	0.7907	0.8889	0.9451	0.9746	0.9889	0.9954	0.9982
1.8000	0.1797	0.4066	0.6149	0.7725	0.8761	0.9371	0.9701	0.9865	0.9942	0.9977
1.9000	0.1681	0.3867	0.5934	0.7541	0.8628	0.9287	0.9652	0.9839	0.9930	0.9971
2.0000	0.1573	0.3679	0.5724	0.7358	0.8491	0.9197	0.9598	0.9810	0.9915	0.9963
2.1000	0.1473	0.3499	0.5519	0.7174	0.8351	0.9103	0.9541	0.9778	0.9898	0.9955
2.2000	0.1380	0.3329	0.5319	0.6990	0.8208	0.9004	0.9479	0.9743	0.9879	0.9946
2.3000	0.1294	0.3166	0.5125	0.6808	0.8063	0.8901	0.9414	0.9704	0.9858	0.9935
2.4000	0.1213	0.3012	0.4936	0.6626	0.7915	0.8795	0.9344	0.9662	0.9835	0.9923
2.5000	0.1138	0.2865	0.4753	0.6446	0.7765	0.8685	0.9271	0.9617	0.9809	0.9909
2.6000	0.1069	0.2725	0.4575	0.6268	0.7614	0.8571	0.9194	0.9569	0.9781	0.9893
2.7000	0.1003	0.2592	0.4402	0.6092	0.7461	0.8454	0.9113	0.9518	0.9750	0.9876
2.8000	0.0943	0.2466	0.4235	0.5918	0.7308	0.8335	0.9029	0.9463	0.9717	0.9857
2.9000	0.0886	0.2346	0.4073	0.5747	0.7154	0.8213	0.8941	0.9405	0.9681	0.9837

Table D.11: χ^2 probability table for observed values of (rows) χ^2 and (columns) degrees of freedom ν.

	1	2	3	4	5	6	7	8	9	10
0.5000	0.4795	0.7788	0.9189	0.9735	0.9921	0.9978	0.9994	0.9999	1.0000	1.0000
1.0000	0.3173	0.6065	0.8013	0.9098	0.9626	0.9856	0.9948	0.9982	0.9994	0.9998
1.5000	0.2207	0.4724	0.6823	0.8266	0.9131	0.9595	0.9823	0.9927	0.9971	0.9989
2.0000	0.1573	0.3679	0.5724	0.7358	0.8491	0.9197	0.9598	0.9810	0.9915	0.9963
2.5000	0.1138	0.2865	0.4753	0.6446	0.7765	0.8685	0.9271	0.9617	0.9809	0.9909
3.0000	0.0833	0.2231	0.3916	0.5578	0.7000	0.8088	0.8850	0.9344	0.9643	0.9814
3.5000	0.0614	0.1738	0.3208	0.4779	0.6234	0.7440	0.8352	0.8992	0.9411	0.9671
4.0000	0.0455	0.1353	0.2615	0.4060	0.5494	0.6767	0.7798	0.8571	0.9114	0.9473
4.5000	0.0339	0.1054	0.2123	0.3425	0.4799	0.6093	0.7207	0.8094	0.8755	0.9220
5.0000	0.0253	0.0821	0.1718	0.2873	0.4159	0.5438	0.6600	0.7576	0.8343	0.8912
5.5000	0.0190	0.0639	0.1386	0.2397	0.3579	0.4815	0.5992	0.7030	0.7887	0.8554
6.0000	0.0143	0.0498	0.1116	0.1991	0.3062	0.4232	0.5397	0.6472	0.7399	0.8153
6.5000	0.0108	0.0388	0.0897	0.1648	0.2606	0.3696	0.4827	0.5914	0.6890	0.7717
7.0000	0.0082	0.0302	0.0719	0.1359	0.2206	0.3208	0.4289	0.5366	0.6371	0.7254
7.5000	0.0062	0.0235	0.0576	0.1117	0.1860	0.2771	0.3787	0.4838	0.5852	0.6775
8.0000	0.0047	0.0183	0.0460	0.0916	0.1562	0.2381	0.3326	0.4335	0.5341	0.6288
8.5000	0.0036	0.0143	0.0367	0.0749	0.1307	0.2037	0.2906	0.3862	0.4846	0.5801
9.0000	0.0027	0.0111	0.0293	0.0611	0.1091	0.1736	0.2527	0.3423	0.4373	0.5321
9.5000	0.0021	0.0087	0.0233	0.0497	0.0907	0.1473	0.2187	0.3019	0.3925	0.4854
10.0000	0.0016	0.0067	0.0186	0.0404	0.0752	0.1247	0.1886	0.2650	0.3505	0.4405
10.5000	0.0012	0.0052	0.0148	0.0328	0.0622	0.1051	0.1620	0.2317	0.3115	0.3978
11.0000	0.0009	0.0041	0.0117	0.0266	0.0514	0.0884	0.1386	0.2017	0.2757	0.3575
11.5000	0.0007	0.0032	0.0093	0.0215	0.0423	0.0741	0.1182	0.1749	0.2430	0.3199
12.0000	0.0005	0.0025	0.0074	0.0174	0.0348	0.0620	0.1006	0.1512	0.2133	0.2851
12.5000	0.0004	0.0019	0.0059	0.0140	0.0285	0.0517	0.0853	0.1303	0.1866	0.2530
13.0000	0.0003	0.0015	0.0046	0.0113	0.0234	0.0430	0.0721	0.1118	0.1626	0.2237
13.5000	0.0002	0.0012	0.0037	0.0091	0.0191	0.0357	0.0608	0.0958	0.1413	0.1970
14.0000	0.0002	0.0009	0.0029	0.0073	0.0156	0.0296	0.0512	0.0818	0.1223	0.1730
14.5000	0.0001	0.0007	0.0023	0.0059	0.0127	0.0245	0.0430	0.0696	0.1056	0.1514
15.0000	0.0001	0.0006	0.0018	0.0047	0.0104	0.0203	0.0360	0.0591	0.0909	0.1321
15.5000	0.0001	0.0004	0.0014	0.0038	0.0084	0.0167	0.0301	0.0501	0.0781	0.1149
16.0000	0.0001	0.0003	0.0011	0.0030	0.0068	0.0138	0.0251	0.0424	0.0669	0.0996
16.5000	0.0000	0.0003	0.0009	0.0024	0.0056	0.0113	0.0209	0.0358	0.0571	0.0862
17.0000	0.0000	0.0002	0.0007	0.0019	0.0045	0.0093	0.0174	0.0301	0.0487	0.0744
17.5000	0.0000	0.0002	0.0006	0.0015	0.0036	0.0076	0.0144	0.0253	0.0414	0.0640
18.0000	0.0000	0.0001	0.0004	0.0012	0.0029	0.0062	0.0120	0.0212	0.0352	0.0550
18.5000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0051	0.0099	0.0178	0.0298	0.0471
19.0000	0.0000	0.0001	0.0003	0.0008	0.0019	0.0042	0.0082	0.0149	0.0252	0.0403
19.5000	0.0000	0.0001	0.0002	0.0006	0.0016	0.0034	0.0068	0.0124	0.0213	0.0344
20.0000	0.0000	0.0000	0.0002	0.0005	0.0012	0.0028	0.0056	0.0103	0.0179	0.0293

Bibliography

Abe, K. *et al.* 2010, *Phys. Rev. Lett.*, **107**, 041801.

Aubert, B. *et al.* 2003, *Phys. Rev. Lett.*, **91**, 241801.

Aubert, B. *et al.* 2007, *Phys. Rev. D*, **76**, 052007.

Aubert, B. *et al.* 2009, *Phys. Rev. D*, **79**, 072009.

Barlow, R. 1989, *Statistics*, Wiley.

Barlow, R. 1990, *NIM A*, **297**, 496-506.

Baum, E. and Haussler, D. 1989 *Neural Comp.*, **1**, 151-160.

Bayes, T. 1763, Reprinted in 1958, *Biometrika*, **45**, (3-4), 296-315.

Bellman, R. 1961, *Adaptive Control Processes: A guided tour*, Oxford University Press.

Beringer, J. *et al.* (Particle Data Group) 2012, *Phys. Rev. D*, **86**, 010001.

Cowan, G. 1998, *Statistical Data Analysis*, Oxford University Press.

Cranmer, K. S. 2001, *Comp. Phys. Comm.*, **136**, 198.

Davidson, A. 2003, *Statistical Models*, Cambridge University Press.

Dunnington, F. 1933, *Phys. Rev.*, **43**, 404.

Edwards, A. 1992, *Likelihood*, John Hopkins University Press.

Fanchiotti, H. Garcia Canal, C. A. and Marucho, M. 2006, *Int. J. Mod. Phys.* C**17** 1461-1476.

Feldman, G. and Cousins, R. 1998, *Phys. Rev. D*, **57**, 3873-3889.

Feynman, R. P., Leighton, R. B. and Sands, M. L. 1989, *The Feynman Lectures on Physics Volume I*, Addison-Wesley.

Fisher, R. A. 1936, *Ann. Eug.*, **7**, 179188.

Gaiser, J. E. 1982, *Ph.D Thesis*, SLAC-255, Appendix F.

Gross, E. and Vitells, O. 2010, *Eur. Phys. J. C*, **70**, 525-530.

Hastie, T., Tibshirani, R. and Friedman, J. 2009, *The Elements of Statistical Learning*, Springer.

James, F. 2007, *Statistical Methods in Experimental Physics*, World Scientific.

Kölbig, K. S. and Schorr, B. 1984 *Comp. Phys. Comm.*, **31**, 97, erratum: ibid 2008 **178**, *972*.

Kendall, M. and Stuart, A. 1979, *The Advanced Theory of Statistics (Vols. 1, 2, and 3)*, Charles Griffin & Company Limited.

Kirkup, L. and Frenkel, B. 2006, *An Introduction to uncertainty in measurement*, Cambridge University Press.

Knuth, D. E. 1998, *Art of Computer Programming*, Addison-Wesley.

Landau, L. 1994, *J. Phys. USSR*, **8**, 201; see also W. Allison and J. Cobb 1980, *Ann. Rev. Nucl. Part. Sci.*, **30**, 253.

Lindley, D. V. and Scott, W. F. 1995, *New Cambridge Statistical Tables*, Cambridge University Press.

MacKay, D. 2011, *Information theory, inference, and learning algorithms*, Cambridge University Press.

Mars Climate Orbiter 1998, see `http://mars.jpl.nasa.gov/msp98/orbiter/` for details.

Oreglia, M. J. 1980, *Ph.D Thesis* SLAC-236, Appendix D.

Press, W. H. *et al.* 2002, *Numerical Recipes in C++*, Cambridge University Press.

Quinn, H. and Nir, Y. 2007, *The Mystery of the Missing Antimatter*, Princeton University Press.

Rojas, R. 1996, *Neural Networks: A Systematic Introduction*, Springer.

Sahami, M., Dumais, S., Heckerman, D. and Horvitz, E. 1998, *Learning for Text Categorization: Papers from the 1998 Workshop*, AAAI Technical Report WS-98-05.

Silvia, D. S. and Skilling, J. 2008, *Data Analysis, a Bayesian tutorial*, Oxford University Press.

Skwarnicki, T. 1986, *Ph.D Thesis*, DESY F31-86-02, Appendix E.

Somov, A. *et al.* 2007, *Phys. Rev. D*, **76**, 011104.

Srivastava, A. Sahami, M. 2009, *Text Mining: Classification, Clustering, and Applications*, Chapman and Hall/CRC.

Wilks, S. S. 1937, *Ann. of Math. Stat.*, **9**, 60-62.