27. PROBABILITY

Revised May 1996 by D.E. Groom (LBNL) and F. James (CERN), September 1999 by R. Cousins (UCLA), October 2001 by G. Cowan (RHUL).

27.1. General [1–8]

An abstract definition of probability can be given by considering a set S, called the sample space, and possible subsets $A, B, \ldots$, the interpretation of which is left open. The probability P is a real-valued function defined by the following axioms due to Kolmogorov [9]:

1. For every subset A in S, $P(A) \geq 0$.
2. For disjoint subsets (*i.e.*, $A \cap B = \emptyset$), $P(A \cup B) = P(A) + P(B)$.
3. $P(S) = 1$.

In addition one defines the conditional probability $P(A|B)$ (read P of A given B) as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \, . \tag{27.1}$$

From this definition and using the fact that $A \cap B$ and $B \cap A$ are the same, one obtains *Bayes' theorem*,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \, . \tag{27.2}$$

From the three axioms of probability and the definition of conditional probability, one obtains the *law of total probability*,

$$P(B) = \sum_i P(B|A_i)P(A_i) \, , \tag{27.3}$$

for any subset B and for disjoint A_i with $\cup_i A_i = S$. This can be combined with Bayes' theorem Eq. (27.2) to give

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i (B|A_i)P(A_i)} \, , \tag{27.4}$$

where the subset A could, for example, be one of the A_i.

The most commonly used interpretation of the subsets of the sample space are outcomes of a repeatable experiment. The probability $P(A)$ is assigned a value equal to the limiting frequency of occurrence of A. This interpretation forms the basis of *frequentist statistics*.

The subsets of the sample space can also be interpreted as *hypotheses*, *i.e.*, statements that are either true or false, such as 'The mass of the W boson lies between 80.3 and 80.5 GeV'. In the frequency interpretation, such statements are either always or never true, *i.e.*, the corresponding probabilities would be 0 or 1. Using *subjective probability*, however, $P(A)$ is interpreted as the degree of belief that the hypothesis A is true.

Subjective probability is used in *Bayesian* (as opposed to frequentist) statistics. Bayes' theorem can be written

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}) , \tag{27.5}$$

where 'theory' represents some hypothesis and 'data' is the outcome of the experiment. Here $P(\text{theory})$ is the *prior* probability for the theory, which reflects the experimenter's degree of belief before carrying out the measurement, and $P(\text{data}|\text{theory})$ is the probability to have gotten the data actually obtained, given the theory, which is also called the *likelihood*.

Bayesian statistics provides no fundamental rule for obtaining the prior probability; this is necessarily subjective and may depend on previous measurements, theoretical prejudices, etc. Once this has been specified, however, Eq. (27.5) tells how the probability for the theory must be modified in the light of the new data to give the *posterior* probability, $P(\text{theory}|\text{data})$. As Eq. (27.5) is stated as a proportionality, the probability must be normalized by summing (or integrating) over all possible hypotheses.

27.2. Random variables

A *random variable* is a numerical characteristic assigned to an element of the sample space. In the frequency interpretation of probability, it corresponds to an outcome of a repeatable experiment. Let x be a possible outcome of an observation. If x can take on any value from a continuous range, we write $f(x;\theta)dx$ as the probability that the measurement's outcome lies between x and $x + dx$. The function $f(x;\theta)$ is called the *probability density function* (p.d.f.), which may depend on one or more parameters θ. If x can take on only discrete values (*e.g.*, the non-negative integers), then $f(x;\theta)$ is itself a probability.

The p.d.f. is always normalized to unit area (unit sum, if discrete). Both x and θ may have multiple components and are then often written as vectors. If θ is unknown, we may wish to estimate its value from a given set of measurements of x; this is a central topic of *statistics* (see Sec. 28).

The *cumulative distribution function* $F(a)$ is the probability that $x \le a$:

$$F(a) = \int_{-\infty}^{a} f(x)\, dx \ . \tag{27.6}$$

Here and below, if x is discrete-valued, the integral is replaced by a sum. The endpoint a is expressly included in the integral or sum. Then $0 \le F(x) \le 1$, $F(x)$ is nondecreasing, and $P(a < x \le b) = F(b) - F(a)$. If x is discrete, $F(x)$ is flat except at allowed values of x, where it has discontinuous jumps equal to $f(x)$.

Any function of random variables is itself a random variable, with (in general) a different p.d.f. The *expectation value* of any function $u(x)$ is

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\, f(x)\, dx \ , \tag{27.7}$$

assuming the integral is finite. For $u(x)$ and $v(x)$ any two functions of x, $E[u+v] = E[u] + E[v]$. For c and k constants, $E[cu+k] = cE[u] + k$.

The n^{th} moment of a random variable is

$$\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx \ , \tag{27.8a}$$

and the n^{th} central moment of x (or moment about the mean, α_1) is

$$m_n \equiv E[(x-\alpha_1)^n] = \int_{-\infty}^{\infty} (x-\alpha_1)^n f(x)\, dx \ . \tag{27.8b}$$

The most commonly used moments are the mean μ and variance σ^2:

$$\mu \equiv \alpha_1 \ , \tag{27.9a}$$
$$\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2 \ . \tag{27.9b}$$

The mean is the location of the "center of mass" of the p.d.f., and the variance is a measure of the square of its width. Note that $V[cx+k] = c^2 V[x]$. It is often convenient to use the *standard deviation* of x, σ, defined as the square root of the variance.

Any odd moment about the mean is a measure of the skewness of the p.d.f. The simplest of these is the dimensionless coefficient of skewness $\gamma_1 = m_3/\sigma^3$.

The fourth central moment m_4 provides a convenient measure of the tails of a distribution. For the Gaussian distribution (see Sec. 27.4) one has $m_4 = 3\sigma^4$. The *kurtosis* is defined as $\gamma_2 = m_4/\sigma^4 - 3$, *i.e.*, it is zero for a Gaussian, positive for a *leptokurtic* distribution with longer tails, and negative for a *platykurtic* distribution with tails that die off more quickly than those of a Gaussian.

Besides the mean, another useful indicator of the "middle" of the probability distribution is the *median*, $x_{\rm med}$, defined by $F(x_{\rm med}) = 1/2$, *i.e.*, half the probability lies above and half lies below $x_{\rm med}$. (More rigorously, $x_{\rm med}$ is a median if $P(x < x_{\rm med}) \le 1/2$ and $P(x \le x_{\rm med}) \ge 1/2$. If only one value exists it is called 'the median'.)

Let x and y be two random variables with a *joint* p.d.f. $f(x,y)$. The *marginal* p.d.f. of x (the distribution of x with y unobserved) is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \ , \tag{27.10}$$

and similarly for the marginal p.d.f. $f_2(y)$. The *conditional* p.d.f. of y given fixed x (with $f_1(x) \neq 0$) is defined by $f_3(y|x) = f(x,y)/f_1(x)$ and similarly $f_4(x|y) = f(x,y)/f_2(y)$. From these we immediately obtain Bayes' theorem (see Eqs. (27.2) and (27.4)),

$$f_4(x|y) = \frac{f_3(y|x) f_1(x)}{f_2(y)} = \frac{f_3(y|x) f_1(x)}{\int f_3(y|x') f_1(x')\, dx'} \ . \tag{27.11}$$

The mean of x is

$$\mu_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x\, f_1(x)\, dx \ , \tag{27.12}$$

and similarly for y. The *covariance* of x and y is

$$\mathrm{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)] = E[xy] - \mu_x \mu_y \ . \tag{27.13}$$

A dimensionless measure of the covariance of x and y is given by the *correlation coefficient*,

$$\rho_{xy} = \mathrm{cov}[x,y]/\sigma_x \sigma_y \ , \tag{27.14}$$

where σ_x and σ_y are the standard deviations of x and y. It can be shown that $-1 \leq \rho_{xy} \leq 1$.

Two random variables x and y are *independent* if and only if

$$f(x,y) = f_1(x) f_2(y) \ . \tag{27.15}$$

If x and y are independent then $\rho_{xy} = 0$; the converse is not necessarily true. If x and y are independent, $E[u(x)v(y)] = E[u(x)]E[v(y)]$, and $V[x+y] = V[x] + V[y]$; otherwise, $V[x+y] = V[x] + V[y] + 2\mathrm{cov}[x,y]$ and $E[uv]$ does not necessarily factorize.

Consider a set of n continuous random variables $\boldsymbol{x} = (x_1, \ldots, x_n)$ with joint p.d.f. $f(\boldsymbol{x})$ and a set of n new variables $\boldsymbol{y} = (y_1, \ldots, y_n)$, related to $\boldsymbol{x}$ by means of a function $\boldsymbol{y}(\boldsymbol{x})$ that is one-to-one, *i.e.*, the inverse $\boldsymbol{x}(\boldsymbol{y})$ exists. The joint p.d.f. for $\boldsymbol{y}$ is given by

$$g(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y}))|J| \ \ , \tag{27.16}$$

where $|J|$ is the absolute value of the determinant of the square matrix $J_{ij} = \partial x_i / \partial y_j$ (the Jacobian determinant). If the transformation from $\boldsymbol{x}$ to $\boldsymbol{y}$ is not one-to-one, the $\boldsymbol{x}$-space must be broken in to regions where the function $\boldsymbol{y}(\boldsymbol{x})$ can be inverted and the contributions to $g(\boldsymbol{y})$ from each region summed.

Given a set of functions $\boldsymbol{y} = (y_1, \ldots, y_m)$ with $m < n$, one can construct $n - m$ additional independent functions, apply the procedure above, then integrate the resulting $g(\boldsymbol{y})$ over the unwanted y_i to find the marginal distribution of those of interest.

To change variables for discrete random variables simply substitute; no Jacobian is necessary because now f is a probability rather than a probability density. If f depends on a set of parameters $\boldsymbol{\theta}$, a change to a different parameter set $\boldsymbol{\eta}(\boldsymbol{\theta})$ is made by simple substitution; no Jacobian is used.

27.3. Characteristic functions

The characteristic function $\phi(u)$ associated with the p.d.f. $f(x)$ is essentially its Fourier transform, or the expectation value of e^{iux}:

$$\phi(u) = E\left[e^{iux}\right] = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx \ . \tag{27.17}$$

Once $\phi(u)$ is specified, the p.d.f. $f(x)$ is uniquely determined and vice versa; knowing one is equivalent to the other. Characteristic functions are useful in deriving a number of important results about moments and sums of random variables.

It follows from Eqs. (27.8a) and (27.17) that the n^{th} moment of a random variable x that follows $f(x)$ is given by

$$i^{-n} \left. \frac{d^n\phi}{du^n}\right|_{u=0} = \int_{-\infty}^{\infty} x^n f(x)\, dx = \alpha_n \ . \tag{27.18}$$

Thus it is often easy to calculate all the moments of a distribution defined by $\phi(u)$, even when $f(x)$ cannot be written down explicitly.

If the p.d.f.s $f_1(x)$ and $f_2(y)$ for independent random variables x and y have characteristic functions $\phi_1(u)$ and $\phi_2(u)$, then the characteristic function of the weighted sum $ax + by$ is $\phi_1(au)\phi_2(bu)$. The addition rules for several important distributions (*e.g.*, that the sum of two Gaussian distributed variables also follows a Gaussian distribution) easily follow from this observation.

Let the (partial) characteristic function corresponding to the conditional p.d.f. $f_2(x|z)$ be $\phi_2(u|z)$, and the p.d.f. of z be $f_1(z)$. The characteristic function after integration over the conditional value is

$$\phi(u) = \int \phi_2(u|z) f_1(z)\, dz \ . \tag{27.19}$$

Suppose we can write ϕ_2 in the form

$$\phi_2(u|z) = A(u) e^{ig(u)z} \ . \tag{27.20}$$

Then

$$\phi(u) = A(u)\phi_1(g(u)) \ . \tag{27.21}$$

The semi-invariants κ_n are defined by

$$\phi(u) = \exp\left[\sum_{n=1}^{\infty} \frac{\kappa_n}{n!}(iu)^n\right] = \exp\left(i\kappa_1 u - \tfrac{1}{2}\kappa_2 u^2 + \ldots\right) \ . \tag{27.22}$$

The values κ_n are related to the moments α_n and m_n. The first few relations are

$$\begin{aligned}
\kappa_1 &= \alpha_1 \ (= \mu, \text{ the mean}) \\
\kappa_2 &= m_2 \ = \alpha_2 - \alpha_1^2 \ (= \sigma^2, \ \text{the variance}) \\
\kappa_3 &= m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^2 \ .
\end{aligned} \tag{27.23}$$

27.4. Some probability distributions

Table 27.1 gives a number of common probability density functions and corresponding characteristic functions, means, and variances. Further information may be found in Refs. [1– 8] and [10]; Ref. [10] has particularly detailed tables. Monte Carlo techniques for generating each of them may be found in our Sec. 29.4. We comment below on all except the trivial uniform distribution.

27.4.1. *Binomial distribution*:

A random process with exactly two possible outcomes which occur with fixed probabilities is called a *Bernoulli* process. If the probability of obtaining a certain outcome (a "success") in each trail is p, then the probability of obtaining exactly r successes ($r = 0, 1, 2, \ldots, N$) in N independent trials, without regard to the order of the successes and failures, is given by the binomial distribution $f(r; N, p)$ in Table 27.1. If r and s are binomially distributed with parameters (N_r, p) and (N_s, p), then $t = r + s$ follows a binomial distribution with parameters $(N_r + N_s, p)$.

27.4.2. *Poisson distribution*:

The Poisson distribution $f(n; \nu)$ gives the probability of finding exactly n events in a given interval of x (*e.g.*, space and time) when the events occur independently of one another and of x at an average rate of ν per the given interval. The variance σ^2 equals ν. It is the limiting case $p \to 0$, $N \to \infty$, $Np = \nu$ of the binomial distribution. The Poisson distribution approaches the Gaussian distribution for large ν.

27.4.3. *Normal or Gaussian distribution*:

The normal (or Gaussian) probability density function $f(x; \mu, \sigma^2)$ given in Table 27.1 has mean $E[x] = \mu$ and variance $V[x] = \sigma^2$. Comparison of the characteristic function $\phi(u)$ given in Table 27.1 with Eq. (27.22) shows that all semi-invariants κ_n beyond κ_2 vanish; this is a unique property of the Gaussian distribution. Some other properties are:

$P(x \text{ in range } \mu \pm \sigma) = 0.6827$,

$P(x \text{ in range } \mu \pm 0.6745\sigma) = 0.5$,

$E[|x - \mu|] = \sqrt{2/\pi}\sigma = 0.7979\sigma$,

half-width at half maximum $= \sqrt{2 \ln 2}\sigma = 1.177\sigma$.

For a Gaussian with $\mu = 0$ and $\sigma^2 = 1$ (the *standard* Gaussian), the cumulative distribution, Eq. (27.6), is related to the error function $\mathrm{erf}(y)$ by

$$F(x; 0, 1) = \tfrac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right] \ . \tag{27.24}$$

The error function and standard Gaussian are tabulated in many references (*e.g.*, Ref. [10]) and are available in libraries of computer routines such as CERNLIB. For a mean μ and variance σ^2, replace x by $(x-\mu)/\sigma$. The probability of x in a given range can be calculated with Eq. (28.43).

For x and y independent and normally distributed, $z = ax + by$ follows $f(z; a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$; that is, the weighted means and variances add.

The Gaussian derives its importance in large part from the *central limit theorem*: If independent random variables $x_1, \ldots, x_n$ are distributed according to *any* p.d.f.s with finite means and variances, then the sum $y = \sum_{i=1}^{n} x_i$ will have a p.d.f. that approaches a Gaussian for large n. The mean and variance are given by the sums of corresponding terms from the individual x_i. Therefore the sum of a large number of fluctuations x_i will be distributed as a Gaussian, even if the x_i themselves are not.

(Note that the *product* of a large number of random variables is not Gaussian, but its logarithm is. The p.d.f. of the product is *log-normal*. See Ref. [8] for details.)

For a set of n Gaussian random variables $\boldsymbol{x}$ with means $\boldsymbol{\mu}$ and corresponding Fourier variables $\boldsymbol{u}$, the characteristic function for a one-dimensional Gaussian is generalized to

$$\phi(\boldsymbol{u}; \boldsymbol{\mu}, V) = \exp\left[i\boldsymbol{\mu} \cdot \boldsymbol{u} - \tfrac{1}{2}\boldsymbol{u}^T V \boldsymbol{u} \right] \ . \tag{27.25}$$

From Eq. (27.18), the covariance of x_i and x_j is

$$E\left[(x_i - \mu_i)(x_j - \mu_j) \right] = V_{ij} \ . \tag{27.26}$$

If the components of $\boldsymbol{x}$ are independent, then $V_{ij} = \delta_{ij}\sigma_i^2$, and Eq. (27.25) is the product of the c.f.s of n Gaussians.

The covariance matrix V can be related to the correlation matrix defined by Eq. (27.14) (a sort of normalized covariance matrix) as $\rho_{ij} = V_{ij}/\sigma_i\sigma_j$. Note that by construction $\rho_{ii} = 1$, since $V_{ii} = \sigma_i^2$.

The characteristic function may be inverted to find the corresponding p.d.f.,

$$f(\boldsymbol{x}; \boldsymbol{\mu}, V) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} \exp\left[-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T V^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right] \tag{27.27}$$

where the determinant $|V|$ must be greater than 0. For diagonal V (independent variables), $f(\boldsymbol{x}; \boldsymbol{\mu}, V)$ is the product of the p.d.f.s of n Gaussian distributions.

For $n = 2$, $f(\boldsymbol{x}; \boldsymbol{\mu}, V)$ is

$$f(x_1, x_2;\ \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times\ \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} . \qquad (27.28)$$

The marginal distribution of any x_i is a Gaussian with mean μ_i and variance V_{ii}. V is $n \times n$, symmetric, and positive definite. Therefore for any vector $\boldsymbol{X}$, the quadratic form $\boldsymbol{X}^T V^{-1} \boldsymbol{X} = C$, where C is any positive number, traces an n-dimensional ellipsoid as $\boldsymbol{X}$ varies. If $X_i = x_i - \mu_i$, then C is a random variable obeying the χ^2 distribution with n degrees of freedom, discussed in the following section. The probability that $\boldsymbol{X}$ corresponding to a set of Gaussian random variables x_i lies outside the ellipsoid characterized by a given value of C $(= \chi^2)$ is given by $1 - F_{\chi^2}(C; n)$, where F_{χ^2} is the cumulative χ^2 distribution. This may be read from Fig. 28.1. For example, the "s-standard-deviation ellipsoid" occurs at $C = s^2$. For the two-variable case ($n = 2$), the point $\boldsymbol{X}$ lies outside the one-standard-deviation ellipsoid with 61% probability. The use of these ellipsoids as indicators of probable error is described in Sec. 28.4.2.3; the validity of those indicators assumes that $\boldsymbol{\mu}$ and V are correct.

27.4.4. χ^2 *distribution*:

If $x_1, \ldots, x_n$ are independent Gaussian random variables, the sum $z = \sum_{i=1}^{n}(x_i - \mu_i)^2/\sigma_i^2$ follows the χ^2 p.d.f. with n *degrees of freedom*, which we denote by $\chi^2(n)$. Under a linear transformation to n correlated Gaussian variables x_i', the value of z is invariant; then $z = \boldsymbol{X}'^T V^{-1} \boldsymbol{X}'$ as in the previous section. For a set of z_i, each of which follows $\chi^2(n_i)$, $\sum z_i$ follows $\chi^2(\sum n_i)$. For large n, the χ^2 p.d.f. approaches a Gaussian with mean $\mu = n$ and variance $\sigma^2 = 2n$.

The χ^2 p.d.f. is often used in evaluating the level of compatibility between observed data and a hypothesis for the p.d.f. that the data might follow. This is discussed further in Sec. 28.3.2 on tests of goodness-of-fit.

27.4.5. *Student's t distribution*:

Suppose that x and $x_1, \ldots, x_n$ are independent and Gaussian distributed with mean 0 and variance 1. We then define

$$z = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad t = \frac{x}{\sqrt{z/n}} . \qquad (27.29)$$

The variable z thus follows a $\chi^2(n)$ distribution. Then t is distributed according to Student's t distribution with n degrees of freedom, $f(t; n)$, given in Table 27.1.

The Student's t distribution resembles a Gaussian with wide tails. As $n \to \infty$, the distribution approaches a Gaussian. If $n = 1$, it is a *Cauchy* or *Breit–Wigner* distribution. The mean is finite only for $n > 1$ and the variance is finite only for $n > 2$, so the central limit theorem is not applicable to sums of random variables following the t distribution for $n = 1$ or 2.

As an example, consider the *sample mean* $\overline{x} = \sum x_i/n$ and the *sample variance* $s^2 = \sum(x_i - \overline{x})^2/(n-1)$ for normally distributed x_i with unknown mean μ and variance σ^2. The sample mean has a Gaussian distribution with a variance σ^2/n, so the variable $(\overline{x} - \mu)/\sqrt{\sigma^2/n}$ is normal with mean 0 and variance 1. Similarly, $(n-1)s^2/\sigma^2$ is independent of this and follows $\chi^2(n-1)$. The ratio

$$t = \frac{(\overline{x} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/\sigma^2(n-1)}} = \frac{\overline{x} - \mu}{\sqrt{s^2/n}} \tag{27.30}$$

is distributed as $f(t; n-1)$. The unknown variance σ^2 cancels, and t can be used to test the probability that the true mean is some particular value μ.

In Table 27.1, n in $f(t; n)$ is not required to be an integer. A Student's t distribution with non-integral $n > 0$ is useful in certain applications.

27.4.6. *Gamma distribution*:

For a process that generates events as a function of x (*e.g.*, space or time) according to a Poisson distribution, the distance in x from an arbitrary starting point (which may be some particular event) to the k^{th} event follows a *gamma* distribution, $f(x; \lambda, k)$. The Poisson parameter μ is λ per unit x. The special case $k = 1$ (*i.e.*, $f(x; \lambda, 1) = \lambda e^{-\lambda x}$) is called the *exponential* distribution. A sum of k' exponential random variables x_i is distributed as $f(\sum x_i; \lambda, k')$.

The parameter k is not required to be an integer. For $\lambda = 1/2$ and $k = n/2$, the gamma distribution reduces to the $\chi^2(n)$ distribution.

References:

1. H. Cramér, *Mathematical Methods of Statistics,* (Princeton Univ. Press, New Jersey, 1958).
2. A. Stuart and A.K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. 1 *Distribution Theory* 5th Ed., (Oxford Univ. Press, New York, 1987), and earlier editions by Kendall and Stuart.
3. W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam, and London, 1971).
4. L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, New York, 1986).
5. B.R. Roe, *Probability and Statistics in Experimental Physics,* 2nd Ed., (Springer, New York, 2001).
6. R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (John Wiley, New York, 1989).

7. S. Brandt, *Data Analysis*, 3rd Ed., (Springer, New York, 1999).
8. G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
9. A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin 1933); *Foundations of the Theory of Probability*, 2nd Ed., (Chelsea, New York 1956).
10. M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1972).

Table 27.1. Some common probability density functions, with corresponding characteristic functions and means and variances. In the Table, $\Gamma(k)$ is the gamma function, equal to $(k-1)!$ when k is an integer.

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu} - e^{iau}}{(b-a)iu}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;N,p) = \dfrac{N!}{r!(N-r)!}\; p^r q^{N-r}$ $r = 0,1,2,\ldots,N\ ;\quad 0 \le p \le 1\ ;\quad q = 1-p$	$(q + pe^{iu})^N$	Np	Npq
Poisson	$f(n;\nu) = \dfrac{\nu^n e^{-\nu}}{n!}\ ;\quad n = 0,1,2,\ldots\ ;\quad \nu > 0$	$\exp[\nu(e^{iu}-1)]$	ν	ν
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\; \exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\ ;\quad -\infty < \mu < \infty\ ;\quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},V) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert V\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty;\quad -\infty < \mu_j < \infty;\quad \det V > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V \boldsymbol{u}\right]$	$\boldsymbol{\mu}$	V_{jk}
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\ ;\quad z \ge 0$	$(1-2iu)^{-n/2}$	n	$2n$
Student's t	$f(t;n) = \dfrac{1}{\sqrt{n\pi}}\; \dfrac{\Gamma[(n+1)/2]}{\Gamma(n/2)} \left(1 + \dfrac{t^2}{n}\right)^{-(n+1)/2}$ $-\infty < t < \infty\ ;\quad n$ not required to be integer	—	0 for $n \ge 2$	$n/(n-2)$ for $n \ge 3$
Gamma	$f(x;\lambda,k) = \dfrac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}\ ;\quad 0 < x < \infty\ ;$ k not required to be integer	$(1 - iu/\lambda)^{-k}$	k/λ	k/λ^2