39. Probability

Revised August 2025 by G. Cowan (RHUL).

39.1 General

Further discussion of probability can be found, *e.g.*, in Refs. [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 as P of A given B) as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \, . \tag{39.1}$$

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \, . \tag{39.2}$$

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

$$P(B) = \sum_i P(B|A_i)P(A_i) \, , \tag{39.3}$$

for any subset B and for disjoint A_i with $\cup_i A_i = S$. This can be combined with Bayes' theorem (Eq. (39.2)) to give

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)} \, , \tag{39.4}$$

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

The most commonly used interpretation of the elements 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 elements of the sample space might 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.' Upon repetition of a measurement, however, such statements are either always true or always false, *i.e.*, the corresponding probabilities in the frequentist interpretation are either 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

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}) \, , \tag{39.5}$$

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, which may depend on previous measurements, theoretical prejudices, *etc.* Once this has been specified,

however, Eq. (39.5) 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. (39.5) is stated as a proportionality, the probability must be normalized by summing (or integrating) over all possible hypotheses.

39.2 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 we use $f(x;\theta)$ to denote the probability to find the value x. In the following the term p.d.f. is often taken to cover both the continuous and discrete cases, although technically the term density should only be used in the continuous case.

The p.d.f. is always normalized to unity. 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. 40).

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

$$F(a) = \int_{-\infty}^{a} f(x)\, dx \ . \tag{39.6}$$

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 $0 \leq F(x) \leq 1$, $F(x)$ is nondecreasing, and $P(a < x \leq b) = F(b) - F(a)$. 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

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\, f(x)\, dx \ , \tag{39.7}$$

assuming the integral is finite. The expectation value is linear, *i.e.*, for any two functions u and v of x and constants c_1 and c_2, $E[c_1 u + c_2 v] = c_1 E[u] + c_2 E[v]$.

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

$$\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx \ , \tag{39.8a}$$

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

$$m_n \equiv E[(x - \alpha_1)^n] = \int_{-\infty}^{\infty} (x - \alpha_1)^n f(x)\, dx \ . \tag{39.8b}$$

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

$$\mu \equiv \alpha_1 \ , \tag{39.9a}$$

$$\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2 \ . \tag{39.9b}$$

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. 39.4), 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.

The *quantile* x_α is the value of the random variable x at which the cumulative distribution is equal to α. That is, the quantile is the inverse of the cumulative distribution function, *i.e.*, $x_\alpha = F^{-1}(\alpha)$. An important special case 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 $P(x \geq x_{\rm med}) \geq 1/2$ and $P(x \leq x_{\rm med}) \geq 1/2$. If only one value exists, it is called '*the* median.')

Under a monotonic change of variable $x \to y(x)$, the quantiles of a distribution (and hence also the median) obey $y_\alpha = y(x_\alpha)$. In general the expectation value and *mode* (most probable value) of a distribution do not, however, transform in this way.

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

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \ , \tag{39.10}$$

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. (39.2) and (39.4)),

$$f_4(x|y) = \frac{f_3(y|x) f_1(x)}{f_2(y)} = \frac{f_3(y|x) f_1(x)}{\int f_3(y|x') f_1(x')\, dx'} \ . \tag{39.11}$$

The mean of x is

$$\mu_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f(x, y)\, dx\, dy = \int_{-\infty}^{\infty} x\, f_1(x)\, dx \ , \tag{39.12}$$

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

$$\mathrm{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y \ . \tag{39.13}$$

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

$$\rho_{xy} = \mathrm{cov}[x, y]/\sigma_x \sigma_y \ , \tag{39.14}$$

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

$$f(x, y) = f_1(x) f_2(y) \ . \tag{39.15}$$

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

$$g(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y}))|J| \ \ , \tag{39.16}$$

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 into 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.

Transformation of p.d.f.s of continuous random variables of any dimension can also be formulated using the Dirac delta function. Starting from an n-dimensional vector random variable $\boldsymbol{x} = (x_1, \ldots, x_n)$ that follows $f(\boldsymbol{x})$ and given m functions $\boldsymbol{a}(\boldsymbol{x}) = (a_1(\boldsymbol{x}), \ldots, a_m(\boldsymbol{x}))$, the p.d.f. of $\boldsymbol{y} \equiv \boldsymbol{a}(\boldsymbol{x})$ can be expressed as

$$g(\boldsymbol{y}) = \int d^n\boldsymbol{x}\, f(\boldsymbol{x})\, \delta^{(m)}(\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{x}))\ , \tag{39.17}$$

where $\delta^{(m)}(\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{x})) = \prod_{i=1}^{m} \delta(y_i - a_i(\boldsymbol{x}))$. A proof and examples can be found in Ref. [10].

For a one-to-one transformation of discrete random variables, the probability is obtained by simple substitution; no Jacobian is necessary because in this case f is a probability rather than a probability density. If the transformation is not one-to-one, then one must sum the probabilities for all values of the original variable that contribute to a given value of the transformed variable. 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.

39.2.1 *Propagation of errors*

Consider n random variables $\boldsymbol{x} = (x_1, \ldots, x_n)$ and m functions $\boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \ldots, y_m(\boldsymbol{x}))$. Suppose here that the mean values $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) = E[\boldsymbol{x}]$ are known, although in practice they will only be estimated, and suppose we also know or have estimated the covariance matrix $V_{ij} = \mathrm{cov}[x_i, x_j]$. The goal of *error propagation* is to determine the covariance matrix for the functions, $U_{ij} = \mathrm{cov}[y_i, y_j]$. In particular, the diagonal elements $U_{ii} = V[y_i]$ give the variances. The new covariance matrix can be found by expanding the functions $\boldsymbol{y}(\boldsymbol{x})$ about the means $\boldsymbol{\mu}$ to first order in a Taylor series. Using this one finds

$$U_{ij} \approx \sum_{k,l} \frac{\partial y_i}{\partial x_k} \frac{\partial y_j}{\partial x_l}\bigg|_{\boldsymbol{\mu}} V_{kl}\ . \tag{39.18}$$

This can be written in matrix notation as $U \approx AVA^T$ where the matrix of derivatives A is

$$A_{ij} = \left.\frac{\partial y_i}{\partial x_j}\right|_{\boldsymbol{\mu}}\ , \tag{39.19}$$

and A^T is its transpose. The approximation is exact if $\boldsymbol{y}(\boldsymbol{x})$ is linear. If this is not the case, the approximation can break down if, for example, $\boldsymbol{y}(\boldsymbol{x})$ is significantly nonlinear close to $\boldsymbol{\mu}$ in a region of a size comparable to the standard deviations of $\boldsymbol{x}$.

39.3 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}:

$$\phi(u) = E\left[e^{iux}\right] = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx\ . \tag{39.20}$$

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. (39.8) and (39.20) that the n^{th} moment of a random variable x that follows $f(x)$ is given by

$$i^{-n} \left.\frac{d^n\phi}{du^n}\right|_{u=0} = \int_{-\infty}^{\infty} x^n f(x)\,dx = \alpha_n \ . \tag{39.21}$$

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 rules of addition 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

$$\phi(u) = \int \phi_2(u|z) f_1(z)\,dz \ . \tag{39.22}$$

Suppose we can write ϕ_2 in the form

$$\phi_2(u|z) = A(u)e^{ig(u)z} \ . \tag{39.23}$$

Then

$$\phi(u) = A(u)\phi_1(g(u)) \ . \tag{39.24}$$

The cumulants (semi-invariants) κ_n of a distribution with characteristic function $\phi(u)$ are defined by the relation

$$\phi(u) = \exp\left[\sum_{n=1}^{\infty} \frac{\kappa_n}{n!}(iu)^n\right] = \exp\left(i\kappa_1 u - \frac{1}{2}\kappa_2 u^2 + \ldots\right) \ . \tag{39.25}$$

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

$$\begin{aligned} \kappa_1 &= \alpha_1 \ (= \mu, \text{ the mean}) \\ \kappa_2 &= m_2 \ = \alpha_2 - \alpha_1^2 \ (= \sigma^2, \text{ the variance}) \\ \kappa_3 &= m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3. \end{aligned} \tag{39.26}$$

39.4 Commonly used probability distributions

Table 39.1 gives a number of common probability density functions and corresponding characteristic functions, means, and variances. Further information may be found in Refs. [1–8], [11] and [12], which has particularly detailed tables. Monte Carlo techniques for generating each of them may be found in our Sec. 42.4 and in Ref. [11]. We comment below on all except the trivial uniform distribution.

39.4.1 *Binomial and multinomial distributions*

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 an individual trial 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 39.1. 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 $(N_r + N_s, p)$. If there are are m possible outcomes for each trial having probabilities $p_1, p_2, \ldots, p_m$, then the joint probability to find $r_1, r_2, \ldots, r_m$ of each outcome after a total of N independent trials is given by the multinomial distribution as shown in Table 39.1. We can regard outcome i as "success" and all the rest as "failure", so individually, any of the r_i follow a binomial distribution for N trials and a success probability p_i.

39.4.2 *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 or 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.

For example, a large number of radioactive nuclei of a given type will result in a certain number of decays in a fixed time interval. If this interval is small compared to the mean lifetime, then the probability for a given nucleus to decay is small, and thus the number of decays in the time interval is well modeled as a Poisson variable.

39.4.3 *Normal or Gaussian distribution*

The normal (or Gaussian) probability density function $f(x; \mu, \sigma^2)$ given in Table 39.1 has mean $E[x] = \mu$ and variance $V[x] = \sigma^2$. Comparison of the characteristic function $\phi(u)$ given in Table 39.1 with Eq. (39.25) shows that all cumulants 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 normal*) the cumulative distribution, often written $\Phi(x)$, is related to the error function erf by

$$F(x; 0, 1) \equiv \Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right] \ . \tag{39.27}$$

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. (40.78).

For x and y independent and normally distributed, $z = ax + by$ follows a normal p.d.f. $f(z; a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$; 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. with finite mean and variance, then the sum $y = \sum_{i=1}^{n} x_i$ will have a p.d.f. that approaches a Gaussian for large n. If the p.d.f.s of the x_i are not identical, the theorem still holds under somewhat more restrictive conditions. 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.

For a set of n Gaussian random variables $\boldsymbol{x}$ with means $\boldsymbol{\mu}$ and covariances $V_{ij} = \mathrm{cov}[x_i, x_j]$, the p.d.f. for the one-dimensional Gaussian is generalized to

$$f(\boldsymbol{x}; \boldsymbol{\mu}, V) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T V^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right] \ , \tag{39.28}$$

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

$$f(x_1,x_2;\ \mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \tag{39.29}$$

The characteristic function for the multivariate Gaussian is

$$\phi(\boldsymbol{u};\boldsymbol{\mu},V) = \exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V\boldsymbol{u}\right]. \tag{39.30}$$

If the components of $\boldsymbol{x}$ are independent, then Eq. (39.30) is the product of the characteristic functions of n Gaussians.

For an n-dimensional Gaussian distribution for $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix V, the marginal distribution for any single x_i is is a one-dimensional Gaussian with mean μ_i and variance V_{ii}. The equation $(\boldsymbol{x}-\boldsymbol{a})^T V^{-1}(\boldsymbol{x}-\boldsymbol{a}) = C$, where C is any positive number, defines an n-dimensional ellipse centered about $\boldsymbol{a}$. If $\boldsymbol{a}$ is equal to the mean $\boldsymbol{\mu}$, then C is a random variable obeying the χ^2 distribution for n degrees of freedom, which is discussed in the following section. The probability that $\boldsymbol{x}$ lies outside the ellipsoid for a given value of C 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. 40.2. 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. 40.4.2.3; the validity of those indicators assumes that $\boldsymbol{\mu}$ and V are correct.

39.4.4 *Log-normal distribution*

If a random variable y follows a Gaussian distribution with mean μ and variance σ^2, then $x = e^y$ follows a log-normal distribution, as given in Table 39.1. As a consequence of the central limit theorem described in Sec. 39.4.3, the distribution of the product of a large number of positive random variables approaches a log-normal. It is bounded below by zero and is thus well suited for modeling quantities that are intrinsically non-negative such as an efficiency. One can implement a log-normal model for a random variable x by defining $y = \ln x$ so that y follows a Gaussian distribution.

39.4.5 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 $\chi^2(n)$. More generally, for n correlated Gaussian variables as components of a vector $\boldsymbol{X}$ with covariance matrix V, $z = \boldsymbol{X}^T V^{-1}\boldsymbol{X}$ follows $\chi^2(n)$ 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 a mean and variance given by $\mu = n$ and $\sigma^2 = 2n$, respectively (here the formulae for μ and σ^2 are valid for all n).

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. 40.3.2 on significance tests.

Table 39.1: 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; ${}_1F_1$ is the confluent hypergeometric function of the 1st kind [12].

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{ibu}-e^{iau}}{(b-a)iu}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Binomial	$f(r;N,p) = \frac{N!}{r!(N-r)!}\, p^r q^{N-r}$ $r = 0,1,2,\ldots,N;\quad 0 \le p \le 1;\quad q = 1-p$	$(q + pe^{iu})^N$	Np	Npq
Multinomial	$f(r_1,\ldots,r_m;N,p_1,\ldots,p_m) = \frac{N!}{r_1!\cdots r_m!} p_1^{r_1}\cdots p_m^{r_m}$	$\left(\sum_{k=1}^m p_k e^{iu_k}\right)^N$	$E[r_i] = Np_i$	$\mathrm{cov}[r_i,r_j] = Np_i(\delta_{ij} - p_j)$
Poisson	$f(n;\nu) = \frac{\nu^n e^{-\nu}}{n!};\ n = 0,1,2,\ldots;\ \nu > 0$	$\exp[\nu(e^{iu}-1)]$	ν	ν
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\ \exp(-(x-\mu)^2/2\sigma^2)$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},V) = \frac{1}{(2\pi)^{n/2}\sqrt{\lvert V\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty;\quad -\infty < \mu_j < \infty;\quad \lvert V\rvert > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V \boldsymbol{u}\right]$	$\boldsymbol{u}$	V_{jk}
Log-normal	$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x}\ \exp(-(\ln x-\mu)^2/2\sigma^2)$ $0 < x < \infty;\quad -\infty < \mu < \infty;\quad \sigma > 0$	—	$\exp(\mu+\sigma^2/2)$	$\exp(2\mu+\sigma^2)$ $\times[\exp(\sigma^2)-1]$
χ^2	$f(z;n) = \frac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)};\quad z \ge 0$	$(1-2iu)^{-n/2}$	n	$2n$
Student's t	$f(t;n) = \frac{1}{\sqrt{n\pi}}\ \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}$ $-\infty < t < \infty;\qquad$ n not required to be integer	—	0 for $n > 1$	$n/(n-2)$ for $n > 2$
Gamma	$f(x;\lambda,k) = \frac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)};\quad 0 \le x < \infty\ ;$ k not required to be integer	$(1-iu/\lambda)^{-k}$	k/λ	k/λ^2
Beta	$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $0 \le x \le 1$	${}_1F_1(\alpha;\alpha+\beta;iu)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

39.4.6 *Student's t distribution*

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

$$z = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad t = \frac{y}{\sqrt{z/n}}\ . \tag{39.31}$$

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 39.1.

If defined through gamma functions as in Table 39.1, the parameter n is not required to be an integer. As $n \to \infty$, the distribution approaches a Gaussian, and for $n = 1$ it is a *Cauchy* or *Breit–Wigner* distribution.

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. The quantity $(n-1)s^2/\sigma^2$ is independent of this and follows $\chi^2(n-1)$.

The ratio

$$t = \frac{(\overline{x} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/\sigma^2(n-1)}} = \frac{\overline{x} - \mu}{\sqrt{s^2/n}} \tag{39.32}$$

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

39.4.7 *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.

39.4.8 *Beta distribution*

The beta distribution describes a continuous random variable x in the interval $[0, 1]$. By scaling and translation one can easily generalize it to have arbitrary endpoints. In Bayesian inference about the parameter p of a binomial process, if the prior p.d.f. is a beta distribution $f(p; \alpha, \beta)$ then the observation of r successes out of N trials gives a posterior beta distribution $f(p; r + \alpha, N - r + \beta)$ (Bayesian methods are discussed further in Sec. 40). The uniform distribution is a beta distribution with $\alpha = \beta = 1$.

References

[1] H. Cramér, *Mathematical Methods of Statistics,* (Princeton Univ. Press, New Jersey, 1958).

[2] A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. 1 *Distribution Theory* 6th Ed., (Halsted Press, New York, 1994), and earlier editions by Kendall and Stuart.

[3] F.E. James, *Statistical Methods in Experimental Physics*, 2nd Ed., (World Scientific, Singapore, 2006).

[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] D. T. Gillespie, American Journal of Physics **51**, 6, 520 (1983).

[11] Ch. Walck, *Hand-book on Statistical Distributions for Experimentalists*, University of Stockholm Internal Report SUF-PFY/96-01, available from staff.fysik.su.se/~walck/.

[12] M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions*, (Dover, New York, 1972).