27. PROBABILITY

Revised May 1996 by D.E. Groom (LBNL) and F. James (CERN). Updated September 1999 by R. Cousins (UCLA).

27.1. General [1–6]

Let x be a possible outcome of an observation. The probability of x is the relative frequency with which that outcome occurs out of a (possibly hypothetical) large set of similar observations. If x can take any value from a *continuous* range, we write $f(x;\theta)\,dx$ as the probability of observing x between x and $x+dx$. The function $f(x;\theta)$ is the *probability density function* (p.d.f.) for the *random variable* x, which may depend upon 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, but we shall still call it a p.d.f. 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 column vectors. If θ is unknown and we wish to estimate its value from a given set of data measuring x, we may use statistics (see Sec. 28).

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

$$F(a) = \int_{-\infty}^{a} f(x)\,dx\ . \qquad (27.1)$$

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 $\mathrm{Prob}(a < x \le b) = F(b) - F(a)$. 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

$$E\left[u(x)\right] = \int_{-\infty}^{\infty} u(x)\ f(x)\ dx\ , \qquad (27.2)$$

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 nth moment of a distribution is

$$\alpha_n \equiv E(x^n) = \int_{-\infty}^{\infty} x^n f(x) dx\ , \qquad (27.3a)$$

and the nth moment about the mean of x, α_1, is

$$m_n \equiv E[(x-\alpha_1)^n] = \int_{-\infty}^{\infty} (x-\alpha_1)^n f(x) dx\ . \qquad (27.3b)$$

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

$$\mu \equiv \alpha_1 \qquad (27.4a)$$

$$\sigma^2 \equiv \mathrm{Var}(x) \equiv m_2 = \alpha_2 - \mu^2\ . \qquad (27.4b)$$

The mean is the location of the "center of mass" of the probability density function, and the variance is a measure of the square of its width. Note that $\mathrm{Var}(cx+k) = c^2\mathrm{Var}(x)$.

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 $\gamma_1 \equiv m_3/\sigma^3$.

Besides the mean, another useful indicator of the "middle" of the probability distribution is the *median* x_{med}, defined by $F(x_{\mathrm{med}}) = 1/2$; *i.e.*, half the probability lies above and half lies below x_{med}. For a given *sample* of events, x_{med} is the value such that half the events have larger x and half have smaller x (not counting any that have the same x as the median). If the sample median lies between two observed x values, it is set by convention halfway between them. If the p.d.f. for x has the form $f(x-\mu)$ and μ is both mean and median, then for a large number of events N, the variance of the median approaches $1/[4Nf^2(0)]$, provided $f(0) > 0$.

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

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy\ , \qquad (27.5)$$

and similarly for the marginal p.d.f. $f_2(y)$. We define $f_3(y|x)$, the *conditional* p.d.f. of y given fixed x, by

$$f_3(y|x)\,f_1(x) = f(x,y)\ . \qquad (27.6a)$$

Similarly, $f_4(x|y)$, the conditional p.d.f. of x given fixed y, is

$$f_4(x|y)\,f_2(y) = f(x,y)\ . \qquad (27.6b)$$

From these definitions we immediately obtain Bayes' theorem [2]:

$$f_4(x|y) = \frac{f_3(y|x)\,f_1(x)}{f_2(y)} = \frac{f_3(y|x)\,f_1(x)}{\int f_3(y|x)\,f_1(x)\,dx}\ . \qquad (27.7)$$

The mean of x is

$$\mu_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\ f(x,y)\ dx\ dy = \int_{-\infty}^{\infty} x\ f_1(x)\ dx\ , \qquad (27.8)$$

and similarly for y. The *correlation* between x and y is

$$\rho_{xy} = E\left[(x-\mu_x)(y-\mu_y)\right]/\sigma_x\,\sigma_y = \mathrm{Cov}(x,y)/\sigma_x\,\sigma_y\ , \qquad (27.9)$$

where σ_x and σ_y are defined in analogy with Eq. (27.4*b*). It can be shown that $-1 \le \rho_{xy} \le 1$. Here "Cov" is the covariance of x and y, a 2-dimensional generalization of the variance.

Two random variables are *independent* if and only if

$$f(x,y) = f_1(x)\ f_2(y)\ . \qquad (27.10)$$

If x and y are independent then $\rho_{xy} = 0$; the converse is not necessarily true except for Gaussian-distributed x and y. If x and y are independent, $E[u(x)\ v(y)] = E[u(x)]\ E[v(y)]$, and $\mathrm{Var}(x+y) = \mathrm{Var}(x)+\mathrm{Var}(y)$; otherwise, $\mathrm{Var}(x+y) = \mathrm{Var}(x)+\mathrm{Var}(y)+ 2\mathrm{Cov}(x,y)$, and $E(u\ v)$ does not factor.

In a *change of continuous random variables* from $\boldsymbol{x} \equiv (x_1,\ldots,x_n)$, with p.d.f. $f(\boldsymbol{x}) = f(x_1,\ldots,x_n)$, to $\boldsymbol{y} \equiv (y_1,\ldots,y_n)$, a one-to-one function of the x_i's, the p.d.f. $g(\boldsymbol{y}) = g(y_1,\ldots,y_n)$ is found by substitution for $(x_1,\ldots,x_n)$ in f followed by multiplication by the absolute value of the Jacobian of the transformation; that is,

$$g(\boldsymbol{y}) = f\left[w_1(\boldsymbol{y}),\ldots,w_n(\boldsymbol{y})\right]|J|\ . \qquad (27.11)$$

The functions w_i express the *inverse* transformation, $x_i = w_i(\boldsymbol{y})$ for $i = 1,\ldots,n$, and $|J|$ is the absolute value of the determinant of the square matrix $J_{ij} = \partial x_i/\partial y_j$. If the transformation from $\boldsymbol{x}$ to $\boldsymbol{y}$ is not one-to-one, the situation is more complex and a unique solution may not exist. For example, if the change is to $m < n$ variables, then a given y may correspond to more than one x, leading to multiple integrals over the contributions [1].

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 upon a parameter set $\boldsymbol{\alpha}$, a change to a different parameter set $\phi_i = \phi_i(\boldsymbol{\alpha})$ is made by simple substitution; no Jacobian is used.

27.2. Characteristic functions

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

$$\phi(u) = E(e^{iux}) = \int_{-\infty}^{\infty} e^{iux} f(x) dx\ . \qquad (27.12)$$

It is often useful, and several of its properties follow [1].

It follows from Eqs. (27.3*a*) and (27.12) that the nth moment of the distribution $f(x)$ is given by

$$i^{-n}\frac{d^n\phi}{du^n}\bigg|_{u=0} = \int_{-\infty}^{\infty} x^n f(x) dx = \alpha_n\ . \qquad (27.13)$$

Thus it is often easy to calculate all the moments of a distribution defined by $\phi(u)$, even when $f(x)$ is difficult to obtain.

If $f_1(x)$ and $f_2(y)$ 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 common distributions (*e.g.*, that the sum of two numbers from Gaussian distributions also has 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

$$\phi(u) = \int \phi_2(u|z)\, f_1(z) dz \ . \tag{27.14}$$

Suppose we can write ϕ_2 in the form

$$\phi_2(u|z) = A(u) e^{ig(u)z} \ . \tag{27.15}$$

Then

$$\phi(u) = A(u)\phi_1(g(u)) \ . \tag{27.16}$$

The semi-invariants κ_n are defined by

$$\phi(u) = \exp\left(\sum_1^\infty \frac{\kappa_n}{n!}(iu)^n\right) = \exp\left(i\kappa_1 u - \tfrac{1}{2}\kappa_2 u^2 + \ldots\right) \ . \tag{27.17}$$

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

$$\begin{aligned} &\kappa_1 = \alpha_1 \ (= \mu, \text{ the mean}) \\ &\kappa_2 = m_2 \ = \alpha_2 - \alpha_1^2 \ (= \sigma^2, \ \text{ the variance}) \\ &\kappa_3 = m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^2 \ . \end{aligned} \tag{27.18}$$

27.3. 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. 1–7; Ref. 7 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.

27.3.1. ***Binomial distribution***: A random process with exactly two possible outcomes is called a *Bernoulli* process. If the probability of obtaining a certain outcome (a "success") in each trial is p, then the probability of obtaining exactly r successes ($r = 0, 1, 2, \ldots, n$) in n 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 successes are observed in n_r trials with probability p of a success, and if s successes are observed in n_s similar trials, then $t = r + s$ is also binomial with $n_t = n_r + n_s$.

27.3.2. ***Poisson distribution***: The Poisson distribution $f(r; \mu)$ gives the probability of finding exactly r 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 μ per the given interval. The variance σ^2 equals μ. It is the limiting case $p \to 0$, $n \to \infty$, $np = \mu$ of the binomial distribution. The Poisson distribution approaches the Gaussian distribution for large μ.

Two or more Poisson processes (*e.g.*, *signal* + *background*, with parameters μ_s and μ_b) that independently contribute amounts n_s and n_b to a given measurement will produce an observed number $n = n_s + n_b$, which is distributed according to a new Poisson distribution with parameter $\mu = \mu_s + \mu_b$.

27.3.3. ***Normal or Gaussian distribution***: The normal (or Gaussian) probability density function $f(x; \mu, \sigma^2)$ given in Table 27.1 has mean $\overline{x} = \mu$ and variance σ^2. Comparison of the characteristic function $\phi(u)$ given in Table 27.1 with Eq. (27.17) shows that all semi-invariants κ_n beyond κ_2 vanish; this is a unique property of the Gaussian distribution. Some properties of the distribution are:

rms deviation $= \sigma$

probability x in the range $\mu \pm \sigma = 0.6827$

probability x in the range $\mu \pm 0.6745\sigma = 0.5$

expection value of $|x - \mu|$, $E(|x - \mu|) = (2/\pi)^{1/2}\sigma = 0.7979\sigma$

half-width at half maximum $= (2 \ln 2)^{1/2}\sigma = 1.177\sigma$

The cumulative distribution, Eq. (27.1), for a Gaussian with $\mu = 0$ and $\sigma^2 = 1$ is related to the error function erf(y) by

$$F(x; 0, 1) = \tfrac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right] \ . \tag{27.19}$$

The error function is tabulated in Ref. 7 and is available in computer math libraries and personel computer spreadsheets. 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. (28.36).

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

The Gaussian gets its importance in large part from the *central limit theorem*: If a continuous random variable x is distributed according to *any* p.d.f. with finite mean and variance, then the sample mean, $\overline{x}_n$, of n observations of x will have a p.d.f. that approaches a Gaussian as n increases. Therefore the end result $\sum^n x_i \equiv n\overline{x}_n$ of a large number of small 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 *lognormal*. See Ref. 6 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

$$\phi(\boldsymbol{x}; \boldsymbol{\mu}, S) = \exp\left[i\boldsymbol{\mu} \cdot \boldsymbol{u} - \tfrac{1}{2}\boldsymbol{u}^T S \boldsymbol{u}\right] \ . \tag{27.20}$$

From Eq. (27.13), the covariance about the mean is

$$E\left[(x_j - \mu_j)(x_k - \mu_k)\right] = S_{jk} \ . \tag{27.21}$$

If the $\boldsymbol{x}$ are independent, then $S_{jk} = \delta_{jk}\sigma_j^2$, and Eq. (27.20) is the product of the c.f.'s of n Gaussians.

The covariance matrix S can be related to the correlation matrix defined by Eq. (27.9) (a sort of normalized covariance matrix). With the definition $\sigma_k^2 \equiv S_{kk}$, we have $\rho_{jk} = S_{jk}/\sigma_j\sigma_k$.

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

$$f(\boldsymbol{x}; \boldsymbol{\mu}, S) = \frac{1}{(2\pi)^{n/2}\sqrt{|S|}} \ \exp\left[-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T S^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right] \tag{27.22}$$

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

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	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu}-e^{iau}}{(b-a)iu}$	$\overline{x} = \dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;n,p) = \dfrac{n!}{r!(n-r)!}\, p^r q^{n-r}$ $r = 0,1,2,\ldots,n\ ;\quad 0 \le p \le 1\ ;\quad q = 1-p$	$(q+pe^{iu})^n$	$\overline{r} = np$	npq
Poisson	$f(r;\mu) = \dfrac{\mu^r e^{-\mu}}{r!}\ ;\quad r = 0,1,2,\ldots\ ;\quad \mu > 0$	$\exp[\mu(e^{iu}-1)]$	$\overline{r} = \mu$	μ
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\,\exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\ ;\quad -\infty < \mu < \infty\ ;\quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	$\overline{x} = \mu$	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},S) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert S\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T S^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty;\quad -\infty < \mu_j < \infty;\quad \det S > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T S\boldsymbol{u}\right]$	$\boldsymbol{\mu}$	S_{jk}
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\ ;\quad z \ge 0$	$(1-2iu)^{-n/2}$	$\overline{z} = n$	$2n$
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	—	$\overline{t} = 0$ 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	$(1-iu/\lambda)^{-k}$	$\overline{x} = k/\lambda$	k/λ^2

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

$$f(x_1,x_2;\ \mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times\ \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \tag{27.23}$$

The marginal distribution of any x_i is a Gaussian with mean μ_i and variance S_{ii}. S is $n \times n$, symmetric, and positive definite. Therefore for any vector $\boldsymbol{X}$, the quadratic form $\boldsymbol{X}^T S^{-1}\boldsymbol{X} = C$, where C is any positive number, traces an n-dimensional ellipsoid as $\boldsymbol{X}$ varies. If $X_i = (x_i-\mu_i)/\sigma_i$, then C is a random variable obeying the $\chi^2(n)$ distribution, discussed in the following section. The probability that $\boldsymbol{X}$ corresponding to a set of Gaussian random variables $\boldsymbol{x}_i$ lies *outside* the ellipsoid characterized by a given value of $C\,(=\chi^2)$ is given by Eq. (27.24) and may be read from Fig. 27.1. 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. (This assumes that μ_i and σ_i are correct.) For $X_i = x_i/\sigma_i$, the ellipsoids of constant χ^2 have the same size and orientation but are centered at $\boldsymbol{\mu}$. The use of these ellipsoids as indicators of probable error is described in Sec. 28.6.2.

27.3.4. χ^2 *distribution*: If $x_1,\ldots,x_n$ are independent Gaussian distributed random variables, the sum $z = \sum^n (x_i-\mu_i)^2/\sigma_i^2$ is distributed as a χ^2 with n *degrees of freedom*, $\chi^2(n)$. Under a linear transformation to n dependent Gaussian variables x_i', the χ^2 at each transformed point retains its value; then $z = \boldsymbol{X}'^T\,V^{-1}\boldsymbol{X}'$ as in the previous section. For a set of z_i, each of which is $\chi^2(n_i)$, $\sum z_i$ is a new random variable which is $\chi^2\left(\sum n_i\right)$.

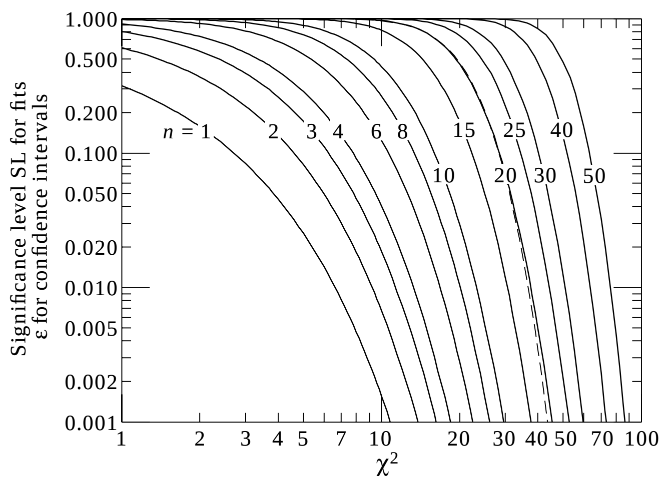

Figure 27.1: The significance level versus χ^2 for n degrees of freedom, as defined in Eq. (27.24). The curve for a given n gives the probability that a value at least as large as χ^2 will be obtained in an experiment; *e.g.*, for $n = 10$, a value $\chi^2 \gtrsim 18$ will occur in 5% of a large number of experiments. For a fit, the SL is a measure of goodness-of-fit, in that a good fit to a correct model is expected to yield a low χ^2 (see Sec. 28.5.0). For a confidence interval, ε measures the probability that the interval *does not* cover the true value of the quantity being estimated (see Sec. 28.6). The dashed curve for $n = 20$ is calculated using the approximation of Eq. (27.25).

Fig. 27.1 shows the signficance level (SL) obtained by integrating the tail of $f(z; n)$:

$$\mathrm{SL}(\chi^2) = \int_{\chi^2}^{\infty} f(z;\, n)\, dz\ . \qquad (27.24)$$

This is shown for a special case in Fig. 27.2, and is equal to 1.0 minus the cumulative distribution function $F(z = \chi^2;\, n)$. It is useful in evaluating the consistency of data with a model (see Sec. 28): The SL is the probability that a random repeat of the given experiment would observe a greater χ^2, assuming the model is correct. It is also useful for confidence intervals for statistical estimators (see Sec. 28.6), in which case one is interested in the unshaded area of Fig. 27.2.

Since the mean of the χ^2 distribution is equal to n, one expects in a "reasonable" experiment to obtain $\chi^2 \approx n$. Hence the "reduced χ^2" $\equiv \chi^2/n$ is sometimes reported. Since the p.d.f. of χ^2/n depends on n, one must report n as well in order to make a meaningful statement. Figure 27.3 shows χ^2/n for useful SL's as a function of n.

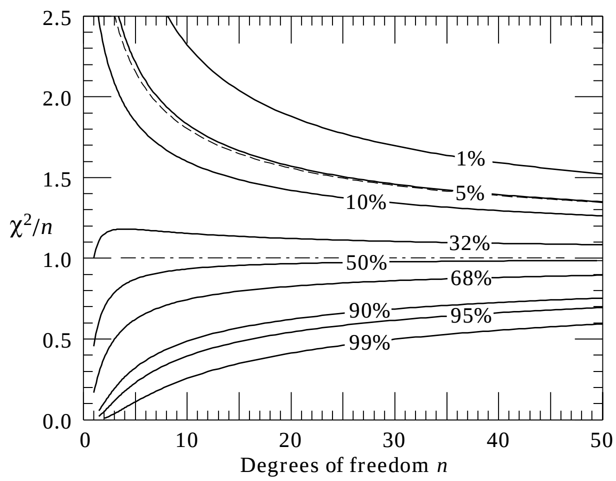

Figure 27.3: Significance levels as a function of the "reduced χ^2" $\equiv \chi^2/n$ and the number of degrees of freedom n. Curves are labeled by the probability that a measurement will give a value of χ^2/n greater than that given on the y axis; *e.g.*, for $n = 10$, a value $\chi^2/n \gtrsim 1.8$ can be expected 5% of the time.

For large n, the SL is approximately given by [1,8]

$$\mathrm{SL}(\chi^2) \approx \frac{1}{\sqrt{2\pi}} \int_y^{\infty} e^{-x^2/2}\, dx\ , \qquad (27.25)$$

where $y = \sqrt{2\chi^2} - \sqrt{2n-1}$. This approximation was used to draw the dashed curves in Fig. 27.1 (for $n = 20$) and Fig. 27.3 (for SL = 5%). Since all the functions and their inverses are now readily available in standard mathematical libraries (such as IMSL, used to generate these figures, and personal computer spreadsheets, such as Microsoft® Excel [9]), the approximation (and even figures and tables) are seldom needed.

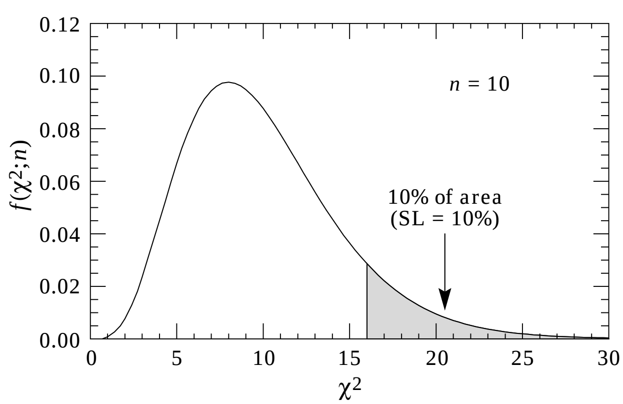

Figure 27.2: Illustration of the significance level integral given in Eq. (27.24). This particlar example is for $n = 10$, where the area above 15.99 is 0.1.

27.3.5. ***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

$$z = \sum_1^n x_i^2\ , \quad \text{and} \quad t = \frac{x}{\sqrt{z/n}}\ . \qquad (27.26)$$

The variable z thus belongs to a $\chi^2(n)$ distribution. Then t is distributed according to a 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 distribution with wide tails. As $n \to \infty$, the distribution approaches a Gaussian. If $n = 1$, the distribution 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 for $n = 1$ or $n = 2$, the central limit theorem is not applicable to t.

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 random variables 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 is χ^2 distributed with $n - 1$ degrees of freedom. The ratio

$$t = \frac{(\overline{x} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)\, s^2/\sigma^2\ (n-1)}} = \frac{\overline{x} - \mu}{\sqrt{s^2/n}} \qquad (27.27)$$

is distributed as $f(t;\, n-1)$. The unknown true 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 nonintegral $n > 0$ is useful in certain applications.

27.3.6. ***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 belongs to 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,* (Springer-Verlag, New York, 1992).
6. G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
7. M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1972).
8. R.A. Fisher, *Statistical Methods for Research Workers,* 8th edition, Edinburgh and London (1941).
9. Microsoft® is a registered trademark of Microsoft corporation.