31. PROBABILITY

Revised May 1996 by D.E. Groom (LBNL) and F. James (CERN), September 1999 by R. Cousins (UCLA), October 2001 and October 2003 by G. Cowan (RHUL).

31.1. 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

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; . \tag{31.1}$$

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \; . \tag{31.2}$$

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

$$P(B) = \sum_i P(B|A_i)P(A_i) \; , \tag{31.3}$$

for any subset B and for disjoint A_i with $\cup_i A_i = S$. This can be combined with Bayes' theorem Eq. (31.2) to give

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i (B|A_i)P(A_i)} \; , \tag{31.4}$$

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

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}) \; , \tag{31.5}$$

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. (31.5) 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. (31.5) is stated as a proportionality, the probability must be normalized by summing (or integrating) over all possible hypotheses.

31.2. 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. 32).

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

$$F(a) = \int_{-\infty}^{a} f(x)\, dx \; . \tag{31.6}$$

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

$$E[u(x)] = \int_{-\infty}^{\infty} u(x)\, f(x)\, dx \; , \tag{31.7}$$

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

$$\alpha_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx \; , \tag{31.8a}$$

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

$$m_n \equiv E[(x - \alpha_1)^n] = \int_{-\infty}^{\infty} (x - \alpha_1)^n f(x)\, dx \; . \tag{31.8b}$$

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

$$\mu \equiv \alpha_1 \; , \tag{31.9a}$$

$$\sigma^2 \equiv V[x] \equiv m_2 = \alpha_2 - \mu^2 \; . \tag{31.9b}$$

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. 31.4) 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_{med}, defined by $F(x_{\text{med}}) = 1/2$, *i.e.*, half the probability lies above and half lies below x_{med}. (More rigorously, x_{med} is a median if $P(x \geq x_{\text{med}}) \geq 1/2$ and $P(x \leq x_{\text{med}}) \geq 1/2$. 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

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \ , \tag{31.10}$$

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. (31.2) and (31.4)),

$$f_4(x|y) = \frac{f_3(y|x)f_1(x)}{f_2(y)} = \frac{f_3(y|x)f_1(x)}{\int f_3(y|x')f_1(x')\,dx'} \ . \tag{31.11}$$

The mean of x is

$$\mu_x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x\, f_1(x)\, dx \ , \tag{31.12}$$

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

$$\text{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)] = E[xy] - \mu_x\mu_y \ . \tag{31.13}$$

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

$$\rho_{xy} = \text{cov}[x,y]/\sigma_x\sigma_y \ , \tag{31.14}$$

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

$$f(x,y) = f_1(x)f_2(y) \ . \tag{31.15}$$

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, $V[x+y] = V[x] + V[y] + 2\text{cov}[x,y]$ 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

$$g(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y}))|J| \ , \tag{31.16}$$

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.

31.3. 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}:

$$\phi(u) = E\left[e^{iux}\right] = \int_{-\infty}^{\infty} e^{iux} f(x)\, dx \ . \tag{31.17}$$

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. (31.8a) and (31.17) that the n^{th} moment of a random variable x that follows $f(x)$ is given by

$$i^{-n}\left.\frac{d^n\phi}{du^n}\right|_{u=0} = \int_{-\infty}^{\infty} x^n f(x)\, dx = \alpha_n \ . \tag{31.18}$$

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

$$\phi(u) = \int \phi_2(u|z)f_1(z)\, dz \ . \tag{31.19}$$

Suppose we can write ϕ_2 in the form

$$\phi_2(u|z) = A(u)e^{ig(u)z} \ . \tag{31.20}$$

Then

$$\phi(u) = A(u)\phi_1(g(u)) \ . \tag{31.21}$$

The semi-invariants κ_n are defined by

$$\phi(u) = \exp\left[\sum_{n=1}^{\infty} \frac{\kappa_n}{n!}(iu)^n\right] = \exp\left(i\kappa_1 u - \tfrac{1}{2}\kappa_2 u^2 + \ldots\right) \ . \tag{31.22}$$

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

$$\begin{aligned}
&\kappa_1 = \alpha_1 \ (= \mu, \text{ the mean}) \\
&\kappa_2 = m_2 \ = \alpha_2 - \alpha_1^2 \ (= \sigma^2, \ \text{the variance}) \\
&\kappa_3 = m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^2 \ .
\end{aligned} \tag{31.23}$$

Table 31.1. 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

31.4. Some probability distributions

Table 31.1 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. 33.4. We comment below on all except the trivial uniform distribution.

31.4.1. *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 31.1. 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).

31.4.2. *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.

31.4.3. *Normal or Gaussian distribution* :

The normal (or Gaussian) probability density function $f(x;\mu,\sigma^2)$ given in Table 31.1 has mean $E[x] = \mu$ and variance $V[x] = \sigma^2$. Comparison of the characteristic function $\phi(u)$ given in Table 31.1 with Eq. (31.22) shows that all semi-invariants κ_n beyond κ_2 vanish; this is a unique property of the Gaussian distribution. Some other properties are:

$$P(x \text{ in range } \mu \pm \sigma) = 0.6827,$$
$$P(x \text{ in range } \mu \pm 0.6745\sigma) = 0.5,$$
$$E[|x-\mu|] = \sqrt{2/\pi}\,\sigma = 0.7979\sigma,$$
$$\text{half-width at half maximum} = \sqrt{2\ln 2}\,\sigma = 1.177\sigma.$$

For a Gaussian with $\mu = 0$ and $\sigma^2 = 1$ (the *standard* Gaussian), the cumulative distribution, Eq. (31.6), is related to the error function $\text{erf}(y)$ by

$$F(x;0,1) = \tfrac{1}{2}\left[1 + \text{erf}(x/\sqrt{2})\right]\ . \tag{31.24}$$

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. (32.43).

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

$$\phi(\boldsymbol{u};\boldsymbol{\mu},V) = \exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \tfrac{1}{2}\boldsymbol{u}^T V\boldsymbol{u}\right] . \tag{31.25}$$

From Eq. (31.18), the covariance of x_i and x_j is

$$E\left[(x_i-\mu_i)(x_j-\mu_j)\right] = V_{ij} . \tag{31.26}$$

If the components of $\boldsymbol{x}$ are independent, then $V_{ij} = \delta_{ij}\sigma_i^2$, and Eq. (31.25) 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. (31.14) (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.,

$$f(\boldsymbol{x};\boldsymbol{\mu},V) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}}\exp\left[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{31.27}$$

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

$$f(x_1,x_2;\ \mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} . \tag{31.28}$$

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. 32.1. 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. 32.3.2.3; the validity of those indicators assumes that $\boldsymbol{\mu}$ and V are correct.

31.4.4. 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. 32.2.2 on tests of goodness-of-fit.

31.4.5. *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

$$z = \sum_{i=1}^n x_i^2 \quad\text{and}\quad t = \frac{x}{\sqrt{z/n}} . \tag{31.29}$$

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 31.1.

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

$$t = \frac{(\overline{x}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/\sigma^2(n-1)}} = \frac{\overline{x}-\mu}{\sqrt{s^2/n}} \tag{31.30}$$

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 31.1, 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.

31.4.6. *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).

32. STATISTICS

Revised April 1998 by F. James (CERN); February 2000 by R. Cousins (UCLA); October 2001, October 2003, and August 2005 by G. Cowan (RHUL).

This chapter gives an overview of statistical methods used in High Energy Physics. In statistics we are interested in using a given sample of data to make inferences about a probabilistic model, *e.g.*, to assess the model's validity or to determine the values of its parameters. There are two main approaches to statistical inference, which we may call frequentist and Bayesian. In frequentist statistics, probability is interpreted as the frequency of the outcome of a repeatable experiment. The most important tools in this framework are parameter estimation, covered in Section 32.1, and statistical tests, discussed in Section 32.2. Frequentist confidence intervals, which are constructed so as to cover the true value of a parameter with a specified probability, are treated in Section 32.3.2. Note that in frequentist statistics one does not define a probability for a hypothesis or for a parameter.

Frequentist statistics provides the usual tools for reporting objectively the outcome of an experiment without needing to incorporate prior beliefs concerning the parameter being measured or the theory being tested. As such they are used for reporting essentially all measurements and their statistical uncertainties in High Energy Physics.

In Bayesian statistics, the interpretation of probability is more general and includes *degree of belief*. One can then speak of a probability density function (p.d.f.) for a parameter, which expresses one's state of knowledge about where its true value lies. Bayesian methods allow for a natural way to input additional information such as physical boundaries and subjective information; in fact they *require* as input the *prior* p.d.f. for the parameters, *i.e.*, the degree of belief about the parameters' values before carrying out the measurement. Using Bayes' theorem Eq. (31.4), the prior degree of belief is updated by the data from the experiment. Bayesian methods for interval estimation are discussed in Sections 32.3.1 and 32.3.2.5

Bayesian techniques are often used to treat systematic uncertainties, where the author's subjective beliefs about, say, the accuracy of the measuring device may enter. Bayesian statistics also provides a useful framework for discussing the validity of different theoretical interpretations of the data. This aspect of a measurement, however, will usually be treated separately from the reporting of the result.

For many inference problems, the frequentist and Bayesian approaches give the same numerical answers, even though they are based on fundamentally different interpretations of probability. For small data samples, however, and for measurements of a parameter near a physical boundary, the different approaches may yield different results, so we are forced to make a choice. For a discussion of Bayesian vs. non-Bayesian methods, see References written by a statistician[1], by a physicist[2], or the more detailed comparison in Ref. [3].

Following common usage in physics, the word "error" is often used in this chapter to mean "uncertainty". More specifically it can indicate the size of an interval as in "the standard error" or "error propagation", where the term refers to the standard deviation of an estimator.

32.1. Parameter estimation

Here we review *point estimation* of parameters. An *estimator* $\widehat{\theta}$ (written with a hat) is a function of the data whose value, the *estimate*, is intended as a meaningful guess for the value of the parameter θ.

There is no fundamental rule dictating how an estimator must be constructed. One tries therefore to choose that estimator which has the best properties. The most important of these are (a) *consistency*, (b) *bias*, (c) *efficiency*, and (d) *robustness*.

(a) An estimator is said to be *consistent* if the estimate $\widehat{\theta}$ converges to the true value θ as the amount of data increases. This property is so important that it is possessed by all commonly used estimators.

(b) The *bias*, $b = E[\,\widehat{\theta}\,] - \theta$, is the difference between the expectation value of the estimator and the true value of the parameter. The expectation value is taken over a hypothetical set of similar experiments in which $\widehat{\theta}$ is constructed in the same way. When $b = 0$ the estimator is said to be unbiased. The bias depends on the chosen metric, *i.e.*, if $\widehat{\theta}$ is an unbiased estimator of θ, then $\widehat{\theta}^{\,2}$ is not in general an unbiased estimator for θ^2. If we have an estimate $\widehat{b}$ for the bias we can subtract it from $\widehat{\theta}$ to obtain a new $\widehat{\theta}^{\,\prime} = \widehat{\theta} - \widehat{b}$. The estimate $\widehat{b}$ may, however, be subject to statistical or systematic uncertainties that are larger than the bias itself, so that the new estimator may not be better than the original.

(c) *Efficiency* is the inverse of the ratio of the variance $V[\,\widehat{\theta}\,]$ to its minimum possible value. Under rather general conditions, the minimum variance is given by the Rao-Cramér-Frechet bound,

$$\sigma^2_{\rm min} = \left(1 + \frac{\partial b}{\partial \theta}\right)^2 / I(\theta) \; , \tag{32.1}$$

where

$$I(\theta) = E\left[\left(\frac{\partial}{\partial \theta}\sum_i \ln f(x_i;\theta)\right)^2\right] \tag{32.2}$$

is the *Fisher information*. The sum is over all data, assumed independent and distributed according to the p.d.f. $f(x;\theta)$, b is the bias, if any, and the allowed range of x must not depend on θ.

The *mean-squared error*,

$$\mathrm{MSE} = E[(\widehat{\theta} - \theta)^2] = V[\widehat{\theta}] + b^2 \; , \tag{32.3}$$

is a convenient quantity which combines the uncertainties in an estimate due to bias and variance.

(d) *Robustness* is the property of being insensitive to departures from assumptions in the p.d.f. owing to factors such as noise.

For some common estimators the properties above are known exactly. More generally, it is possible to evaluate them by Monte Carlo simulation. Note that they will often depend on the unknown θ.

32.1.1. *Estimators for mean, variance and median* :

Suppose we have a set of N independent measurements x_i assumed to be unbiased measurements of the same unknown quantity μ with a common, but unknown, variance σ^2. Then

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{32.4}$$

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \widehat{\mu})^2 \tag{32.5}$$

are unbiased estimators of μ and σ^2. The variance of $\widehat{\mu}$ is σ^2/N and the variance of $\widehat{\sigma^2}$ is

$$V\left[\widehat{\sigma^2}\right] = \frac{1}{N}\left(m_4 - \frac{N-3}{N-1}\sigma^4\right) \; , \tag{32.6}$$

where m_4 is the 4th central moment of x. For Gaussian distributed x_i this becomes $2\sigma^4/(N-1)$ for any $N \geq 2$, and for large N the standard deviation of $\widehat{\sigma}$ (the "error of the error") is $\sigma/\sqrt{2N}$. Again if the x_i are Gaussian, $\widehat{\mu}$ is an efficient estimator for μ and the estimators $\widehat{\mu}$ and $\widehat{\sigma^2}$ are uncorrelated. Otherwise the arithmetic mean (32.4) is not necessarily the most efficient estimator; this is discussed in more detail in [4] Sec. 8.7

If σ^2 is known, it does not improve the estimate $\widehat{\mu}$, as can be seen from Eq. (32.4); however, if μ is known, substitute it for $\widehat{\mu}$ in Eq. (32.5) and replace $N-1$ by N to obtain a somewhat better estimator of σ^2. If the x_i have different, known variances σ_i^2, then the weighted average

$$\widehat{\mu} = \frac{1}{w}\sum_{i=1}^{N} w_i x_i \tag{32.7}$$

is an unbiased estimator for μ with a smaller variance than an unweighted average; here $w_i = 1/\sigma_i^2$ and $w = \sum_i w_i$. The standard deviation of $\widehat{\mu}$ is $1/\sqrt{w}$.

As an estimator for the median $x_{\rm med}$ one can use the value $\widehat{x}_{\rm med}$ such that half the x_i are below and half above (the sample median). If the sample median lies between two observed values, it is set by convention halfway between them. If the p.d.f. of x has the form $f(x-\mu)$ and μ is both mean and median, then for large N the variance of the sample median approaches $1/[4Nf^2(0)]$, provided $f(0) > 0$. Although estimating the median can often be more difficult computationally than the mean, the resulting estimator is generally more robust, as it is insensitive to the exact shape of the tails of a distribution.

32.1.2. *The method of maximum likelihood* :

"From a theoretical point of view, the most important general method of estimation so far known is the *method of maximum likelihood*" [5]. We suppose that a set of N independently measured quantities x_i came from a p.d.f. $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ is set of n parameters whose values are unknown. The method of maximum likelihood takes the estimators $\widehat{\boldsymbol{\theta}}$ to be those values of $\boldsymbol{\theta}$ that maximize the *likelihood function*,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i;\boldsymbol{\theta}) \; . \qquad (32.8)$$

The likelihood function is the joint p.d.f. for the data, evaluated with the data obtained in the experiment and regarded as a function of the parameters. Note that the likelihood function is *not* a p.d.f. for the parameters $\boldsymbol{\theta}$; in frequentist statistics this is not defined. In Bayesian statistics one can obtain from the likelihood the posterior p.d.f. for $\boldsymbol{\theta}$, but this requires multiplying by a prior p.d.f. (see Sec. 32.3.1).

It is usually easier to work with $\ln L$, and since both are maximized for the same parameter values $\boldsymbol{\theta}$, the maximum likelihood (ML) estimators can be found by solving the *likelihood equations*,

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \, , \qquad i = 1, \ldots, n \; . \qquad (32.9)$$

Maximum likelihood estimators are important because they are approximately unbiased and efficient for large data samples, under quite general conditions, and the method has a wide range of applicability.

In evaluating the likelihood function, it is important that any normalization factors in the p.d.f. that involve $\boldsymbol{\theta}$ be included. However, we will only be interested in the maximum of L and in ratios of L at different values of the parameters; hence any multiplicative factors that do not involve the parameters that we want to estimate may be dropped, including factors that depend on the data but not on $\boldsymbol{\theta}$.

Under a one-to-one change of parameters from $\boldsymbol{\theta}$ to $\boldsymbol{\eta}$, the ML estimators $\widehat{\boldsymbol{\theta}}$ transform to $\boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})$. That is, the ML solution is invariant under change of parameter. However, other properties of ML estimators, in particular the bias, are not invariant under change of parameter.

The inverse V^{-1} of the covariance matrix $V_{ij} = {\rm cov}[\widehat{\theta}_i, \widehat{\theta}_j]$ for a set of ML estimators can be estimated by using

$$(\widehat{V}^{-1})_{ij} = - \left. \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right|_{\widehat{\theta}} \; . \qquad (32.10)$$

For finite samples, however, Eq. (32.10) can result in an underestimate of the variances. In the large sample limit (or in a linear model with Gaussian errors), L has a Gaussian form and $\ln L$ is (hyper)parabolic. In this case it can be seen that a numerically equivalent way of determining s-standard-deviation errors is from the contour given by the $\boldsymbol{\theta}'$ such that

$$\ln L(\boldsymbol{\theta}') = \ln L_{\rm max} - s^2/2 \, , \qquad (32.11)$$

where $\ln L_{\rm max}$ is the value of $\ln L$ at the solution point (compare with Eq. (32.46)). The extreme limits of this contour on the θ_i axis give an approximate s-standard-deviation confidence interval for θ_i (see Section 32.3.2.3).

In the case where the size n of the data sample $x_1, \ldots, x_n$ is small, the unbinned maximum likelihood method, i.e., use of equation (32.8), is preferred since binning can only result in a loss of information and hence larger statistical errors for the parameter estimates. The sample size n can be regarded as fixed or the user can choose to treat it as a Poisson-distributed variable; this latter option is sometimes called "extended maximum likelihood" (see, *e.g.*, [6, 7, 8]). If the sample is large it can be convenient to bin the values in a histogram, so that one obtains a vector of data $\boldsymbol{n} = (n_1, \ldots, n_N)$ with expectation values $\boldsymbol{\nu} = E[\boldsymbol{n}]$ and probabilities $f(\boldsymbol{n};\boldsymbol{\nu})$. Then one may maximize the likelihood function based on the contents of the bins (so i labels bins). This is equivalent to maximizing the likelihood ratio $\lambda(\boldsymbol{\theta}) = f(\boldsymbol{n};\boldsymbol{\nu}(\boldsymbol{\theta}))/f(\boldsymbol{n};\boldsymbol{n})$, or to minimizing the quantity [9]

$$-2 \ln \lambda(\boldsymbol{\theta}) = 2 \sum_{i=1}^{N} \left[\nu_i(\boldsymbol{\theta}) - n_i + n_i \ln \frac{n_i}{\nu_i(\boldsymbol{\theta})} \right] \, , \qquad (32.12)$$

where in bins where $n_i = 0$, the last term in (32.12) is zero. In the limit of zero bin width, maximizing (32.12) is equivalent to maximizing the unbinned likelihood function (32.8).

A benefit of binning is that it allows for a goodness-of-fit test (see Sec. 32.2.2). The minimum of $-2\ln\lambda$ as defined by Eq. (32.12) follows a χ^2 distribution in the large sample limit. If there are N bins and m fitted parameters, then the number of degrees of freedom for the χ^2 distribution is $N - m - 1$ if the data are treated as multinomially distributed and $N - m$ if the n_i are Poisson variables with $\nu_{\rm tot} = \sum_i \nu_i$ fixed. If the n_i are Poisson distributed and $\nu_{\rm tot}$ is also fitted, then by minimizing Eq. (32.12) one obtains that the area under the fitted function is equal to the sum of the histogram contents, *i.e.*, $\sum_i \nu_i = \sum_i n_i$. This is not the case for parameter estimation methods based on a least-squares procedure with traditional weights (see, *e.g.*, Ref. [8]).

32.1.3. *The method of least squares* :

The *method of least squares* (LS) coincides with the method of maximum likelihood in the following special case. Consider a set of N independent measurements y_i at known points x_i. The measurement y_i is assumed to be Gaussian distributed with mean $F(x_i;\boldsymbol{\theta})$ and known variance σ_i^2. The goal is to construct estimators for the unknown parameters $\boldsymbol{\theta}$. The likelihood function contains the sum of squares

$$\chi^2(\boldsymbol{\theta}) = -2 \ln L(\boldsymbol{\theta}) + \text{ constant } = \sum_{i=1}^{N} \frac{(y_i - F(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} \; . \qquad (32.13)$$

The set of parameters $\boldsymbol{\theta}$ which maximize L is the same as those which minimize χ^2.

The minimum of Equation (32.13) defines the least-squares estimators $\widehat{\boldsymbol{\theta}}$ for the more general case where the y_i are not Gaussian distributed as long as they are independent. If they are not independent but rather have a covariance matrix $V_{ij} = {\rm cov}[y_i, y_j]$, then the LS estimators are determined by the minimum of

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta}))^T V^{-1} (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta})) \, , \qquad (32.14)$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)$ is the vector of measurements, $\boldsymbol{F}(\boldsymbol{\theta})$ is the corresponding vector of predicted values (understood as a column vector in (32.14)), and the superscript T denotes transposed (*i.e.*, row) vector.

In many practical cases one further restricts the problem to the situation where $F(x_i;\boldsymbol{\theta})$ is a linear function of the parameters, *i.e.*,

$$F(x_i;\boldsymbol{\theta}) = \sum_{j=1}^{m} \theta_j h_j(x_i) \; . \qquad (32.15)$$

Here the $h_j(x)$ are m linearly independent functions, *e.g.*, $1, x, x^2, \ldots, x^{m-1}$, or Legendre polynomials. We require $m < N$ and at least m of the x_i must be distinct.

Minimizing χ^2 in this case with m parameters reduces to solving a system of m linear equations. Defining $H_{ij} = h_j(x_i)$ and minimizing χ^2 by setting its derivatives with respect to the θ_i equal to zero gives the LS estimators,

$$\widehat{\boldsymbol{\theta}} = (H^T V^{-1} H)^{-1} H^T V^{-1} \boldsymbol{y} \equiv D\boldsymbol{y} \ . \tag{32.16}$$

The covariance matrix for the estimators $U_{ij} = \text{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$ is given by

$$U = DVD^T = (H^T V^{-1} H)^{-1} \ , \tag{32.17}$$

or equivalently, its inverse U^{-1} can be found from

$$(U^{-1})_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j}\bigg|_{\theta=\widehat{\theta}} = \sum_{k,l=1}^{N} h_i(x_k)(V^{-1})_{kl} h_j(x_l) \ . \tag{32.18}$$

The LS estimators can also be found from the expression

$$\widehat{\boldsymbol{\theta}} = U\boldsymbol{g} \ , \tag{32.19}$$

where the vector $\boldsymbol{g}$ is defined by

$$g_i = \sum_{j,k=1}^{N} y_j h_i(x_k)(V^{-1})_{jk} \ . \tag{32.20}$$

For the case of uncorrelated y_i, for example, one can use (32.19) with

$$(U^{-1})_{ij} = \sum_{k=1}^{N} \frac{h_i(x_k) h_j(x_k)}{\sigma_k^2} \ , \tag{32.21}$$

$$g_i = \sum_{k=1}^{N} \frac{y_k h_i(x_k)}{\sigma_k^2} \ . \tag{32.22}$$

Expanding $\chi^2(\boldsymbol{\theta})$ about $\widehat{\boldsymbol{\theta}}$, one finds that the contour in parameter space defined by

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\widehat{\boldsymbol{\theta}}) + 1 = \chi^2_{\min} + 1 \tag{32.23}$$

has tangent planes located at plus or minus one standard deviation $\sigma_{\widehat{\theta}}$ from the LS estimates $\widehat{\boldsymbol{\theta}}$.

In constructing the quantity $\chi^2(\boldsymbol{\theta})$, one requires the variances or, in the case of correlated measurements, the covariance matrix. Often these quantities are not known *a priori* and must be estimated from the data; an important example is where the measured value y_i represents a counted number of events in the bin of a histogram. If, for example, y_i represents a Poisson variable, for which the variance is equal to the mean, then one can either estimate the variance from the predicted value, $F(x_i; \boldsymbol{\theta})$, or from the observed number itself, y_i. In the first option, the variances become functions of the fitted parameters, which may lead to calculational difficulties. The second option can be undefined if y_i is zero, and in both cases for small y_i the variance will be poorly estimated. In either case one should constrain the normalization of the fitted curve to the correct value, *e.g.*, one should determine the area under the fitted curve directly from the number of entries in the histogram (see [8] Section 7.4). A further alternative is to use the method of maximum likelihood; for binned data this can be done by minimizing Eq. (32.12)

As the minimum value of the χ^2 represents the level of agreement between the measurements and the fitted function, it can be used for assessing the goodness-of-fit; this is discussed further in Section 32.2.2.

32.1.4. ***Propagation of errors*** :

Consider a set of n quantities $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and a set of m functions $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \ldots, \eta_m(\boldsymbol{\theta}))$. Suppose we have estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$, using, say, maximum likelihood or least squares, and we also know or have estimated the covariance matrix $V_{ij} = \text{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$. The goal of *error propagation* is to determine the covariance matrix for the functions, $U_{ij} = \text{cov}[\widehat{\eta}_i, \widehat{\eta}_j]$, where $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\widehat{\boldsymbol{\theta}})$. In particular, the diagonal elements $U_{ii} = V[\widehat{\eta}_i]$ give the variances. The new covariance matrix can be found by expanding the functions $\boldsymbol{\eta}(\boldsymbol{\theta})$ about the estimates $\widehat{\boldsymbol{\theta}}$ to first order in a Taylor series. Using this one finds

$$U_{ij} \approx \sum_{k,l} \frac{\partial \eta_i}{\partial \theta_k} \frac{\partial \eta_j}{\partial \theta_l}\bigg|_{\widehat{\theta}} V_{kl} \ . \tag{32.24}$$

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

$$A_{ij} = \frac{\partial \eta_i}{\partial \theta_j}\bigg|_{\widehat{\theta}} \tag{32.25}$$

and A^T is its transpose. The approximation is exact if $\boldsymbol{\eta}(\boldsymbol{\theta})$ is linear (it holds, for example, in equation (32.17)). If this is not the case the approximation can break down if, for example, $\boldsymbol{\eta}(\boldsymbol{\theta})$ is significantly nonlinear close to $\widehat{\boldsymbol{\theta}}$ in a region of a size comparable to the standard deviations of $\widehat{\boldsymbol{\theta}}$.

32.2. Statistical tests

In addition to estimating parameters, one often wants to assess the validity of certain statements concerning the data's underlying distribution. *Hypothesis tests* provide a rule for accepting or rejecting hypotheses depending on the outcome of a measurement. In *goodness-of-fit tests* one gives the probability to obtain a level of incompatibility with a certain hypothesis that is greater than or equal to the level observed with the actual data.

32.2.1. ***Hypothesis tests*** :

Consider an experiment whose outcome is characterized by a vector of data $\boldsymbol{x}$. A *hypothesis* is a statement about the distribution of $\boldsymbol{x}$. It could, for example, define completely the p.d.f. for the data (a simple hypothesis) or it could specify only the functional form of the p.d.f., with the values of one or more parameters left open (a composite hypothesis).

A *statistical test* is a rule that states for which values of $\boldsymbol{x}$ a given hypothesis (often called the null hypothesis, H_0) should be rejected. This is done by defining a region of $\boldsymbol{x}$-space called the critical region; if the outcome if the experiment lands in this region, H_0 is rejected. Equivalently one can say that the hypothesis is accepted if $\boldsymbol{x}$ is observed in the acceptance region, *i.e.*, the complement of the critical region. Here 'accepted' is understood to mean simply that the test did not reject H_0.

Rejecting H_0 if it is true is called an error of the first kind. The probability for this to occur is called the *significance level* of the test, α, which is often chosen to be equal to some pre-specified value. It can also happen that H_0 is false and the true hypothesis is given by some alternative, H_1. If H_0 is accepted in such a case, this is called an error of the second kind. The probability for this to occur, β, depends on the alternative hypothesis, say, H_1, and $1 - \beta$ is called the *power* of the test to reject H_1.

In High Energy Physics the components of $\boldsymbol{x}$ might represent the measured properties of candidate events, and the acceptance region is defined by the cuts that one imposes in order to select events of a certain desired type. That is, H_0 could represent the signal hypothesis, and various alternatives, H_1, H_2, etc., could represent background processes.

Often rather than using the full data sample $\boldsymbol{x}$ it is convenient to define a *test statistic*, t, which can be a single number or in any case a vector with fewer components than $\boldsymbol{x}$. Each hypothesis for the distribution of $\boldsymbol{x}$ will determine a distribution for t, and the acceptance region in $\boldsymbol{x}$-space will correspond to a specific range of values of t.

In constructing t one attempts to reduce the volume of data without losing the ability to discriminate between different hypotheses.

In particle physics terminology, the probability to accept the signal hypothesis, H_0, is the selection efficiency, *i.e.*, one minus the significance level. The efficiencies for the various background processes are given by one minus the power. Often one tries to construct a test to minimize the background efficiency for a given signal efficiency. The *Neyman–Pearson lemma* states that this is done by defining the acceptance region such that, for $\boldsymbol{x}$ in that region, the ratio of p.d.f.s for the hypotheses H_0 and H_1,

$$\lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|H_0)}{f(\boldsymbol{x}|H_1)} , \qquad (32.26)$$

is greater than a given constant, the value of which is chosen to give the desired signal efficiency. This is equivalent to the statement that (32.26) represents the test statistic with which one may obtain the highest purity sample for a given signal efficiency. It can be difficult in practice, however, to determine $\lambda(\boldsymbol{x})$, since this requires knowledge of the joint p.d.f.s $f(\boldsymbol{x}|H_0)$ and $f(\boldsymbol{x}|H_1)$. Instead, test statistics based on *neural networks* or *Fisher discriminants* are often used (see [10]).

32.2.2. *Goodness-of-fit tests* :

Often one wants to quantify the level of agreement between the data and a hypothesis without explicit reference to alternative hypotheses. This can be done by defining a *goodness-of-fit statistic*, t, which is a function of the data whose value reflects in some way the level of agreement between the data and the hypothesis. The user must decide what values of the statistic correspond to better or worse levels of agreement with the hypothesis in question; for many goodness-of-fit statistics there is an obvious choice.

The hypothesis in question, say, H_0, will determine the p.d.f. $g(t|H_0)$ for the statistic. The goodness-of-fit is quantified by giving the p-value, defined as the probability to find t in the region of equal or lesser compatibility with H_0 than the level of compatibility observed with the actual data. For example, if t is defined such that large values correspond to poor agreement with the hypothesis, then the p-value would be

$$p = \int_{t_{\rm obs}}^{\infty} g(t|H_0)\, dt , \qquad (32.27)$$

where $t_{\rm obs}$ is the value of the statistic obtained in the actual experiment. The p-value should not be confused with the significance level of a test or the confidence level of a confidence interval (Section 32.3), both of which are pre-specified constants.

The p-value is a function of the data and is therefore itself a random variable. If the hypothesis used to compute the p-value is true, then for continuous data, p will be uniformly distributed between zero and one. Note that the p-value is not the probability for the hypothesis; in frequentist statistics this is not defined. Rather, the p-value is the probability, under the assumption of a hypothesis H_0, of obtaining data at least as incompatible with H_0 as the data actually observed.

When estimating parameters using the method of least squares, one obtains the minimum value of the quantity χ^2 (32.13), which can be used as a goodness-of-fit statistic. It may also happen that no parameters are estimated from the data, but that one simply wants to compare a histogram, *e.g.*, a vector of Poisson distributed numbers $\boldsymbol{n} = (n_1, \ldots, n_N)$, with a hypothesis for their expectation values $\nu_i = E[n_i]$. As the distribution is Poisson with variances $\sigma_i^2 = \nu_i$, the χ^2 (32.13) becomes *Pearson's* χ^2 *statistic*,

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i} . \qquad (32.28)$$

If the hypothesis $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ is correct and if the measured values n_i in (32.28) are sufficiently large (in practice, this will be a good approximation if all $n_i > 5$), then the χ^2 statistic will follow the χ^2 p.d.f. with the number of degrees of freedom equal to the number of measurements N minus the number of fitted parameters. The same holds for the minimized χ^2 from Eq. (32.13) if the y_i are Gaussian.

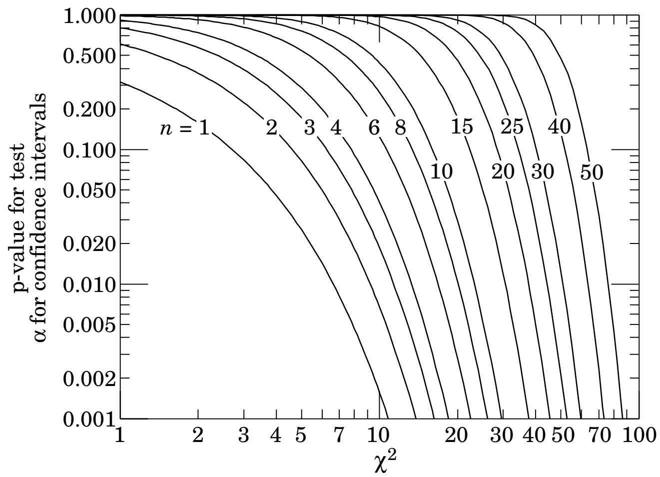


Figure 32.1: One minus the χ^2 cumulative distribution, $1 - F(\chi^2; n)$, for n degrees of freedom. This gives the p-value for the χ^2 goodness-of-fit test as well as one minus the coverage probability for confidence regions (see Sec. 32.3.2.3).

Alternatively one may fit parameters and evaluate goodness-of-fit by minimizing $-2\ln\lambda$ from Eq. (32.12). One finds that the distribution of this statistic approaches the asymptotic limit faster than does Pearson's χ^2 and thus computing the p-value with the χ^2 p.d.f. will in general be better justified (see [9] and references therein).

Assuming the goodness-of-fit statistic follows a χ^2 p.d.f., the p-value for the hypothesis is then

$$p = \int_{\chi^2}^{\infty} f(z; n_{\rm d})\, dz , \qquad (32.29)$$

where $f(z; n_{\rm d})$ is the χ^2 p.d.f. and $n_{\rm d}$ is the appropriate number of degrees of freedom. Values can be obtained from Fig. 32.1 or from the CERNLIB routine `PROB`. If the conditions for using the χ^2 p.d.f. do not hold, the statistic can still be defined as before, but its p.d.f. must be determined by other means in order to obtain the p-value, *e.g.*, using a Monte Carlo calculation.

If one finds a χ^2 value much greater than $n_{\rm d}$ and a correspondingly small p-value, one may be tempted to expect a high degree of uncertainty for any fitted parameters. Although this may be true for systematic errors in the parameters, it is not in general the case for statistical uncertainties. If, for example, the error bars (or covariance matrix) used in constructing the χ^2 are underestimated, then this will lead to underestimated statistical errors for the fitted parameters. But in such a case an estimate $\hat{\theta}$ can differ from the true value θ by an amount much greater than its estimated statistical error. The standard deviations of estimators that one finds from, say, equation (32.11) reflect how widely the estimates would be distributed if one were to repeat the measurement many times, assuming that the measurement errors used in the χ^2 are also correct. They do not include the systematic error which may result from an incorrect hypothesis or incorrectly estimated measurement errors in the χ^2.

Since the mean of the χ^2 distribution is equal to $n_{\rm d}$, one expects in a "reasonable" experiment to obtain $\chi^2 \approx n_{\rm d}$. Hence the quantity $\chi^2/n_{\rm d}$ is sometimes reported. Since the p.d.f. of $\chi^2/n_{\rm d}$ depends on $n_{\rm d}$, however, one must report $n_{\rm d}$ as well in order to make a meaningful statement. The p-values obtained for different values of $\chi^2/n_{\rm d}$ are shown in Fig. 32.2.

32.3. Confidence intervals and limits

When the goal of an experiment is to determine a parameter θ, the result is usually expressed by quoting, in addition to the point estimate, some sort of interval which reflects the statistical precision of the measurement. In the simplest case this can be given by the parameter's estimated value $\widehat{\theta}$ plus or minus an estimate of the standard deviation of $\widehat{\theta}$, $\sigma_{\widehat{\theta}}$. If, however, the p.d.f. of the estimator is not Gaussian or if there are physical boundaries on the possible values of the parameter, then one usually quotes instead an interval according to one of the procedures described below.

In reporting an interval or limit, the experimenter may wish to

- communicate as objectively as possible the result of the experiment;
- provide an interval that is constructed to cover the true value of the parameter with a specified probability;
- provide the information needed by the consumer of the result to draw conclusions about the parameter or to make a particular decision;
- draw conclusions about the parameter that incorporate stated prior beliefs.

With a sufficiently large data sample, the point estimate and standard deviation (or for the multiparameter case, the parameter estimates and covariance matrix) satisfy essentially all of these goals. For finite data samples, no single method for quoting an interval will achieve all of them. In particular, drawing conclusions about the parameter in the framework of Bayesian statistics necessarily requires subjective input.

In addition to the goals listed above, the choice of method may be influenced by practical considerations such as ease of producing an interval from the results of several measurements. Of course the experimenter is not restricted to quoting a single interval or limit; one may choose, for example, first to communicate the result with a confidence interval having certain frequentist properties, and then in addition to draw conclusions about a parameter using Bayesian statistics. It is recommended, however, that there be a clear separation between these two aspects of reporting a result. In the remainder of this section we assess the extent to which various types of intervals achieve the goals stated here.

32.3.1. *The Bayesian approach* :

Suppose the outcome of the experiment is characterized by a vector of data $\boldsymbol{x}$, whose probability distribution depends on an unknown parameter (or parameters) $\boldsymbol{\theta}$ that we wish to determine. In Bayesian statistics, all knowledge about $\boldsymbol{\theta}$ is summarized by the posterior p.d.f. $p(\boldsymbol{\theta}|\boldsymbol{x})$, which gives the degree of belief for $\boldsymbol{\theta}$ to take on values in a certain region given the data $\boldsymbol{x}$. It is obtained by using Bayes' theorem,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')\,d\boldsymbol{\theta}'} , \quad (32.30)$$

where $L(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function, *i.e.*, the joint p.d.f. for the data given a certain value of $\boldsymbol{\theta}$, evaluated with the data actually obtained in the experiment, and $\pi(\boldsymbol{\theta})$ is the prior p.d.f. for $\boldsymbol{\theta}$. Note that the denominator in (32.30) serves simply to normalize the posterior p.d.f. to unity.

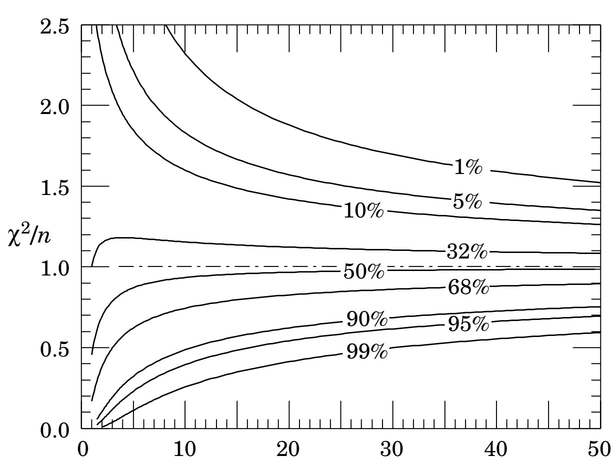


Figure 32.2: The 'reduced' χ^2, equal to χ^2/n, for n degrees of freedom. The curves show as a function of n the χ^2/n that corresponds to a given p-value.

Bayesian statistics supplies no fundamental rule for determining $\pi(\boldsymbol{\theta})$; this reflects the experimenter's subjective degree of belief about $\boldsymbol{\theta}$ before the measurement was carried out. By itself, therefore, the posterior p.d.f. is not a good way to report objectively the result of an observation, since it contains both the result (through the likelihood function) and the experimenter's prior beliefs. Without the likelihood function, someone with different prior beliefs would be unable to substitute these to determine his or her own posterior p.d.f. This is an important reason, therefore, to publish wherever possible the likelihood function or an appropriate summary of it. Often this can be achieved by reporting the ML estimate and one or several low order derivatives of L evaluated at the estimate.

In the single parameter case, for example, an interval (called a Bayesian or credible interval) $[\theta_{\rm lo}, \theta_{\rm up}]$ can be determined which contains a given fraction $1-\alpha$ of the probability, *i.e.*,

$$1-\alpha = \int_{\theta_{\rm lo}}^{\theta_{\rm up}} p(\theta|\boldsymbol{x})\,d\theta . \quad (32.31)$$

Sometimes an upper or lower limit is desired, *i.e.*, $\theta_{\rm lo}$ can be set to zero or $\theta_{\rm up}$ to infinity. In other cases one might choose $\theta_{\rm lo}$ and $\theta_{\rm up}$ such that $p(\theta|\boldsymbol{x})$ is higher everywhere inside the interval than outside; these are called *highest posterior density* (HPD) intervals. Note that HPD intervals are not invariant under a nonlinear transformation of the parameter.

The main difficulty with Bayesian intervals is in quantifying the prior beliefs. Sometimes one attempts to construct $\pi(\boldsymbol{\theta})$ to represent complete ignorance about the parameters by setting it equal to a constant. A problem here is that if the prior p.d.f. is flat in $\boldsymbol{\theta}$, then it is not flat for a nonlinear function of $\boldsymbol{\theta}$, and so a different parametrization of the problem would lead in general to a different posterior p.d.f. In fact, one rarely chooses a flat prior as a true expression of degree of belief about a parameter; rather, it is used as a recipe to construct an interval, which in the end will have certain frequentist properties.

If a parameter is constrained to be non-negative, then the prior p.d.f. can simply be set to zero for negative values. An important example is the case of a Poisson variable n which counts signal events with unknown mean s as well as background with mean b, assumed known. For the signal mean s one often uses the prior

$$\pi(s) = \begin{cases} 0 & s < 0 \\ 1 & s \geq 0 \end{cases} . \quad (32.32)$$

As mentioned above, this is regarded as providing an interval whose frequentist properties can be studied, rather than as representing a degree of belief. In the absence of a clear discovery, (*e.g.*, if $n = 0$ or if in any case n is compatible with the expected background), one usually wishes to place an upper limit on s. Using the likelihood function for Poisson distributed n,

$$L(n|s) = \frac{(s+b)^n}{n!} e^{-(s+b)} , \quad (32.33)$$

along with the prior (32.32) in (32.30) gives the posterior density for s. An upper limit $s_{\rm up}$ at confidence level $1-\alpha$ can be obtained by requiring

$$1-\alpha = \int_{-\infty}^{s_{\rm up}} p(s|n)ds = \frac{\int_{-\infty}^{s_{\rm up}} L(n|s)\,\pi(s)\,ds}{\int_{-\infty}^{\infty} L(n|s)\,\pi(s)\,ds} , \quad (32.34)$$

where the lower limit of integration is effectively zero because of the cut-off in $\pi(s)$. By relating the integrals in Eq. (32.34) to incomplete gamma functions, the equation reduces to

$$\alpha = e^{-s_{\rm up}} \frac{\sum_{m=0}^{n} (s_{\rm up}+b)^m/m!}{\sum_{m=0}^{n} b^m/m!} . \quad (32.35)$$

This must be solved numerically for the limit $s_{\rm up}$. For the special case of $b = 0$, the sums can be related to the *quantile* $F^{-1}_{\chi^2}$ of the χ^2 distribution (inverse of the cumulative distribution) to give

$$s_{\rm up} = \tfrac{1}{2} F^{-1}_{\chi^2}(1-\alpha; n_{\rm d}) \; , \tag{32.36}$$

where the number of degrees of freedom is $n_{\rm d} = 2(n+1)$. The quantile of the χ^2 distribution can be obtained using the CERNLIB routine `CHISIN`. It so happens that for the case of $b = 0$, the upper limits from Eq. (32.36) coincide numerically with the values of the frequentist upper limits discussed in Section 32.3.2.4. Values for $1-\alpha = 0.9$ and 0.95 are given by the values $\nu_{\rm up}$ in Table 32.3. The frequentist properties of confidence intervals for the Poisson mean obtained in this way are discussed in Refs. [2] and [11].

Bayesian statistics provides a framework for incorporating systematic uncertainties into a result. Suppose, for example, that a model depends not only on parameters of interest $\boldsymbol{\theta}$ but on *nuisance parameters* $\boldsymbol{\nu}$, whose values are known with some limited accuracy. For a single nuisance parameter ν, for example, one might have a p.d.f. centered about its nominal value with a certain standard deviation σ_ν. Often a Gaussian p.d.f. provides a reasonable model for one's degree of belief about a nuisance parameter; in other cases more complicated shapes may be appropriate. The likelihood function, prior and posterior p.d.f.s then all depend on both $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ and are related by Bayes' theorem as usual. One can obtain the posterior p.d.f. for $\boldsymbol{\theta}$ alone by integrating over the nuisance parameters, *i.e.*,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \int p(\boldsymbol{\theta},\boldsymbol{\nu}|\boldsymbol{x})\, d\boldsymbol{\nu} \; . \tag{32.37}$$

If the prior joint p.d.f. for $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ factorizes, then integrating the posterior p.d.f. over $\boldsymbol{\nu}$ is equivalent to replacing the likelihood function by (see Ref. [12]),

$$L'(\boldsymbol{x}|\boldsymbol{\theta}) = \int L(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu})\pi(\boldsymbol{\nu})\, d\boldsymbol{\nu} \; . \tag{32.38}$$

The function $L'(\boldsymbol{x}|\boldsymbol{\theta})$ can also be used together with frequentist methods that employ the likelihood function such as ML estimation of parameters. The results then have a mixed frequentist/Bayesian character, where the systematic uncertainty due to limited knowledge of the nuisance parameters is built in. Although this may make it more difficult to disentangle statistical from systematic effects, such a hybrid approach may satisfy the objective of reporting the result in a convenient way.

Even if the subjective Bayesian approach is not used explicitly, Bayes' theorem represents the way that people evaluate the impact of a new result on their beliefs. One of the criteria in choosing a method for reporting a measurement, therefore, should be the ease and convenience with which the consumer of the result can carry out this exercise.

32.3.2. *Frequentist confidence intervals* :

The unqualified phrase "confidence intervals" refers to frequentist intervals obtained with a procedure due to Neyman [13], described below. These are intervals (or in the multiparameter case, regions) constructed so as to include the true value of the parameter with a probability greater than or equal to a specified level, called the *coverage probability*. In this section we discuss several techniques for producing intervals that have, at least approximately, this property.

32.3.2.1. *The Neyman construction for confidence intervals:*

Consider a p.d.f. $f(x;\theta)$ where x represents the outcome of the experiment and θ is the unknown parameter for which we want to construct a confidence interval. The variable x could (and often does) represent an estimator for θ. Using $f(x;\theta)$ we can find for a pre-specified probability $1-\alpha$ and for every value of θ a set of values $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ such that

$$P(x_1 < x < x_2;\theta) = 1-\alpha = \int_{x_1}^{x_2} f(x;\theta)\, dx \; . \tag{32.39}$$

This is illustrated in Fig. 32.3: a horizontal line segment $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ is drawn for representative values of θ. The union of such intervals for all values of θ, designated in the figure as $D(\alpha)$, is known as the *confidence belt*. Typically the curves $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ are monotonic functions of θ, which we assume for this discussion.

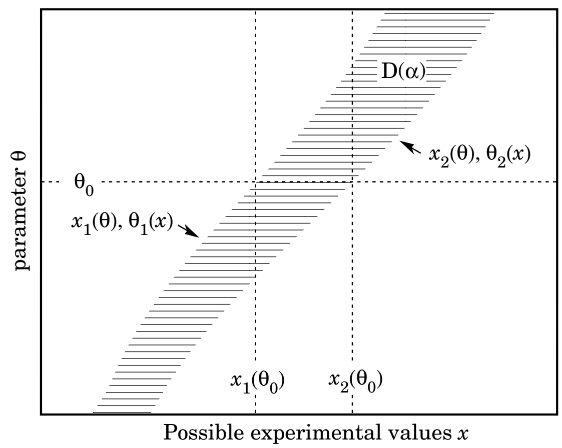


Figure 32.3: Construction of the confidence belt (see text).

Upon performing an experiment to measure x and obtaining a value x_0, one draws a vertical line through x_0. The confidence interval for θ is the set of all values of θ for which the corresponding line segment $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ is intercepted by this vertical line. Such confidence intervals are said to have a *confidence level* (CL) equal to $1-\alpha$.

Now suppose that the true value of θ is θ_0, indicated in the figure. We see from the figure that θ_0 lies between $\theta_1(x)$ and $\theta_2(x)$ if and only if x lies between $x_1(\theta_0)$ and $x_2(\theta_0)$. The two events thus have the same probability, and since this is true for any value θ_0, we can drop the subscript 0 and obtain

$$1-\alpha = P(x_1(\theta) < x < x_2(\theta)) = P(\theta_2(x) < \theta < \theta_1(x)) \; . \tag{32.40}$$

In this probability statement $\theta_1(x)$ and $\theta_2(x)$, *i.e.*, the endpoints of the interval, are the random variables and θ is an unknown constant. If the experiment were to be repeated a large number of times, the interval $[\theta_1, \theta_2]$ would vary, covering the fixed value θ in a fraction $1-\alpha$ of the experiments.

The condition of coverage Eq. (32.39) does not determine x_1 and x_2 uniquely and additional criteria are needed. The most common criterion is to choose *central intervals* such that the probabilities excluded below x_1 and above x_2 are each $\alpha/2$. In other cases one may want to report only an upper or lower limit, in which case the probability excluded below x_1 or above x_2 can be set to zero. Another principle based on *likelihood ratio ordering* for determining which values of x should be included in the confidence belt is discussed in Sec. 32.3.2.2

When the observed random variable x is continuous, the coverage probability obtained with the Neyman construction is $1-\alpha$, regardless of the true value of the parameter. If x is discrete, however, it is not possible to find segments $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ that satisfy (32.39) exactly for all values of θ. By convention one constructs the confidence belt requiring the probability $P(x_1 < x < x_2)$ to be *greater than or equal to* $1-\alpha$. This gives confidence intervals that include the true parameter with a probability greater than or equal to $1-\alpha$.

32.3.2.2. *Relationship between intervals and tests:*

An equivalent method of constructing confidence intervals is to consider a test (see Sec. 32.2) of the hypothesis that the parameter's true value is θ. One then excludes all values of θ where the hypothesis would be rejected at a significance level less than α. The remaining values constitute the confidence interval at confidence level $1-\alpha$.

In this procedure one is still free to choose the test to be used; this corresponds to the freedom in the Neyman construction as to which values of the data are included in the confidence belt. One possibility is use a test statistic based on the *likelihood ratio*,

$$\lambda = \frac{f(x;\theta)}{f(x;\widehat{\theta}\,)} , \qquad (32.41)$$

where $\widehat{\theta}$ is the value of the parameter which, out of all allowed values, maximizes $f(x;\theta)$. This results in the intervals described in [14] by Feldman and Cousins. The same intervals can be obtained from the Neyman construction described in the previous section by including in the confidence belt those values of x which give the greatest values of λ.

Another technique that can be formulated in the language of statistical tests has been used to set limits on the Higgs mass from measurements at LEP [15,16]. For each value of the Higgs mass, a statistic called $\mathrm{CL_s}$ is determined from the ratio

$$\mathrm{CL_s} = \frac{p\text{-value of signal plus background hypothesis}}{1 - p\text{-value of hypothesis of background only}} . \qquad (32.42)$$

The p-values in (32.42) are themselves based on a goodness-of-fit statistic which depends in general on the signal being tested, *i.e.*, on the hypothesized Higgs mass. Smaller $\mathrm{CL_s}$ corresponds to a lesser level of agreement with the signal hypothesis.

In the usual procedure for constructing confidence intervals, one would exclude the signal hypothesis if the probability to obtain a value of $\mathrm{CL_s}$ less than the one actually observed is less than α. The LEP Higgs group has in fact followed a more conservative approach and excludes the signal at a confidence level $1-\alpha$ if $\mathrm{CL_s}$ itself (not the probability to obtain a lower $\mathrm{CL_s}$ value) is less than α. This results in a coverage probability that is in general greater than $1-\alpha$. The interpretation of such intervals is discussed in [15,16].

32.3.2.3. *Gaussian distributed measurements:*

An important example of constructing a confidence interval is when the data consists of a single random variable x that follows a Gaussian distribution; this is often the case when x represents an estimator for a parameter and one has a sufficiently large data sample. If there is more than one parameter being estimated, the multivariate Gaussian is used. For the univariate case with known σ,

$$1-\alpha = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mu-\delta}^{\mu+\delta} e^{-(x-\mu)^2/2\sigma^2}\, dx = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\,\sigma}\right) \qquad (32.43)$$

is the probability that the measured value x will fall within $\pm\delta$ of the true value μ. From the symmetry of the Gaussian with respect to x and μ, this is also the probability for the interval $x \pm \delta$ to include μ. Fig. 32.4 shows a $\delta = 1.64\sigma$ confidence interval unshaded. The choice $\delta = \sigma$ gives an interval called the *standard error* which has $1-\alpha = 68.27\%$ if σ is known. Values of α for other frequently used choices of δ are given in Table 32.1.

We can set a one-sided (upper or lower) limit by excluding above $x+\delta$ (or below $x-\delta$). The values of α for such limits are half the values in Table 32.1.

In addition to Eq. (32.43), α and δ are also related by the cumulative distribution function for the χ^2 distribution,

$$\alpha = 1 - F(\chi^2; n) , \qquad (32.44)$$

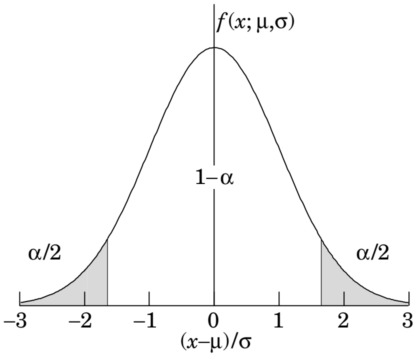


Figure 32.4: Illustration of a symmetric 90% confidence interval (unshaded) for a measurement of a single quantity with Gaussian errors. Integrated probabilities, defined by α, are as shown.

Table 32.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
4.55×10^{-2}	2σ	0.1	1.64σ
2.7×10^{-3}	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ

for $\chi^2 = (\delta/\sigma)^2$ and $n = 1$ degree of freedom. This can be obtained from Fig. 32.1 on the $n = 1$ curve or by using the CERNLIB routine `PROB`.

For multivariate measurements of, say, n parameter estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$, one requires the full covariance matrix $V_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]$, which can be estimated as described in Sections 32.1.2 and 32.1.3. Under fairly general conditions with the methods of maximum-likelihood or least-squares in the large sample limit, the estimators will be distributed according to a multivariate Gaussian centered about the true (unknown) values $\boldsymbol{\theta}$, and furthermore the likelihood function itself takes on a Gaussian shape.

The standard error ellipse for the pair $(\widehat{\theta}_i, \widehat{\theta}_j)$ is shown in Fig. 32.5, corresponding to a contour $\chi^2 = \chi^2_{\mathrm{min}} + 1$ or $\ln L = \ln L_{\mathrm{max}} - 1/2$. The ellipse is centered about the estimated values $\widehat{\boldsymbol{\theta}}$, and the tangents to the ellipse give the standard deviations of the estimators, σ_i and σ_j. The angle of the major axis of the ellipse is given by

$$\tan 2\phi = \frac{2\rho_{ij}\sigma_i\sigma_j}{\sigma_i^2 - \sigma_j^2} , \qquad (32.45)$$

where $\rho_{ij} = \mathrm{cov}[\widehat{\theta}_i, \widehat{\theta}_j]/\sigma_i\sigma_j$ is the correlation coefficient.

The correlation coefficient can be visualized as the fraction of the distance σ_i from the ellipse's horizontal centerline at which the ellipse becomes tangent to vertical, *i.e.* at the distance $\rho_{ij}\sigma_i$ below the centerline as shown. As ρ_{ij} goes to $+1$ or -1, the ellipse thins to a diagonal line.

It could happen that one of the parameters, say, θ_j, is known from previous measurements to a precision much better than σ_j so that the current measurement contributes almost nothing to the knowledge of θ_j. However, the current measurement of of θ_i and its dependence on θ_j may still be important. In this case, instead of quoting both parameter estimates and their correlation, one sometimes reports the value of θ_i which minimizes χ^2 at a fixed value of θ_j, such as the PDG best value. This θ_i value lies along the dotted line between the points where the ellipse becomes tangent to vertical, and has statistical error σ_{inner} as shown on the figure, where $\sigma_{\mathrm{inner}} = (1-\rho_{ij}^2)^{1/2}\sigma_i$. Instead of the correlation ρ_{ij}, one reports the dependency $d\widehat{\theta}_i/d\theta_j$ which is the slope of the dotted line. This slope is related to the correlation coefficient by $d\widehat{\theta}_i/d\theta_j = \rho_{ij} \times \frac{\sigma_i}{\sigma_j}$.

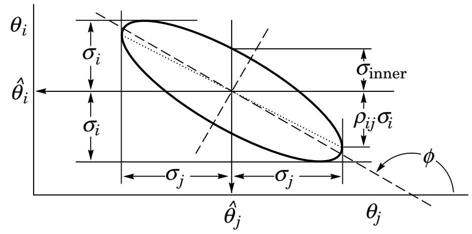


Figure 32.5: Standard error ellipse for the estimators $\widehat{\theta}_i$ and $\widehat{\theta}_j$. In this case the correlation is negative.

Table 32.2: $\Delta\chi^2$ or $2\Delta \ln L$ corresponding to a coverage probability $1-\alpha$ in the large data sample limit, for joint estimation of m parameters.

$(1-\alpha)$ (%)	$m=1$	$m=2$	$m=3$
68.27	1.00	2.30	3.53
90.	2.71	4.61	6.25
95.	3.84	5.99	7.82
95.45	4.00	6.18	8.03
99.	6.63	9.21	11.34
99.73	9.00	11.83	14.16

As in the single-variable case, because of the symmetry of the Gaussian function between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$, one finds that contours of constant $\ln L$ or χ^2 cover the true values with a certain, fixed probability. That is, the confidence region is determined by

$$\ln L(\boldsymbol{\theta}) \geq \ln L_{\rm max} - \Delta \ln L \ , \tag{32.46}$$

or where a χ^2 has been defined for use with the method of least squares,

$$\chi^2(\boldsymbol{\theta}) \leq \chi^2_{\rm min} + \Delta\chi^2 \ . \tag{32.47}$$

Values of $\Delta\chi^2$ or $2\Delta \ln L$ are given in Table 32.2 for several values of the coverage probability and number of fitted parameters.

For finite data samples, the probability for the regions determined by Equations (32.46) or (32.47) to cover the true value of $\boldsymbol{\theta}$ will depend on $\boldsymbol{\theta}$, so these are not exact confidence regions according to our previous definition. Nevertheless, they can still have a coverage probability only weakly dependent on the true parameter and approximately as given in Table 32.2. In any case the coverage probability of the intervals or regions obtained according to this procedure can in principle be determined as a function of the true parameter(s), for example, using a Monte Carlo calculation.

One of the practical advantages of intervals that can be constructed from the log-likelihood function or χ^2 is that it is relatively simple to produce the interval for the combination of several experiments. If N independent measurements result in log-likelihood functions $\ln L_i(\boldsymbol{\theta})$, then the combined log-likelihood function is simply the sum,

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln L_i(\boldsymbol{\theta}) \ . \tag{32.48}$$

This can then be used to determine an approximate confidence interval or region with Equation (32.46), just as with a single experiment.

32.3.2.4. *Poisson or binomial data:*

Another important class of measurements consists of counting a certain number of events n. In this section we will assume these are all events of the desired type, *i.e.*, there is no background. If n represents the number of events produced in a reaction with cross section σ, say, in a fixed integrated luminosity $\mathcal{L}$, then it follows a Poisson distribution with mean $\nu = \sigma\mathcal{L}$. If, on the other hand, one has selected a larger sample of N events and found n of them to have a particular property, then n follows a binomial distribution where the parameter p gives the probability for the event to possess the property in question. This is appropriate, *e.g.*, for estimates of branching ratios or selection efficiencies based on a given total number of events.

For the case of Poisson distributed n, the upper and lower limits on the mean value ν can be found from the Neyman procedure to be

$$\nu_{\rm lo} = \tfrac{1}{2} F^{-1}_{\chi^2}(\alpha_{\rm lo}; 2n) \ , \tag{32.49a}$$

$$\nu_{\rm up} = \tfrac{1}{2} F^{-1}_{\chi^2}(1-\alpha_{\rm up}; 2(n+1)) \ , \tag{32.49b}$$

where the upper and lower limits are at confidence levels of $1-\alpha_{\rm lo}$ and $1-\alpha_{\rm up}$, respectively, and $F^{-1}_{\chi^2}$ is the *quantile* of the χ^2 distribution (inverse of the cumulative distribution). The quantiles $F^{-1}_{\chi^2}$ can be obtained from standard tables or from the CERNLIB routine `CHISIN`. For central confidence intervals at confidence level $1-\alpha$, set $\alpha_{\rm lo} = \alpha_{\rm up} = \alpha/2$.

It happens that the upper limit from (32.49a) coincides numerically with the Bayesian upper limit for a Poisson parameter using a uniform prior p.d.f. for ν. Values for confidence levels of 90% and 95% are shown in Table 32.3.

Table 32.3: Lower and upper (one-sided) limits for the mean ν of a Poisson variable given n observed events in the absence of background, for confidence levels of 90% and 95%.

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	$\nu_{\rm lo}$	$\nu_{\rm up}$	$\nu_{\rm lo}$	$\nu_{\rm up}$
0	–	2.30	–	3.00
1	0.105	3.89	0.051	4.74
2	0.532	5.32	0.355	6.30
3	1.10	6.68	0.818	7.75
4	1.74	7.99	1.37	9.15
5	2.43	9.27	1.97	10.51
6	3.15	10.53	2.61	11.84
7	3.89	11.77	3.29	13.15
8	4.66	12.99	3.98	14.43
9	5.43	14.21	4.70	15.71
10	6.22	15.41	5.43	16.96

For the case of binomially distributed n successes out of N trials with probability of success p, the upper and lower limits on p are found to be

$$p_{\rm lo} = \frac{nF_F^{-1}[\alpha_{\rm lo}; 2n, 2(N-n+1)]}{N-n+1 \ + \ nF_F^{-1}[\alpha_{\rm lo}; 2n, 2(N-n+1)]} \ , \tag{32.50a}$$

$$p_{\rm up} = \frac{(n+1)F_F^{-1}[1-\alpha_{\rm up}; 2(n+1), 2(N-n)]}{(N-n) \ + \ (n+1)F_F^{-1}[1-\alpha_{\rm up}; 2(n+1), 2(N-n)]} \ . \tag{32.50b}$$

Here F_F^{-1} is the quantile of the F distribution (also called the Fisher–Snedecor distribution; see Ref. [4]).

32.3.2.5. *Difficulties with intervals near a boundary:*

A number of issues arise in the construction and interpretation of confidence intervals when the parameter can only take on values in a restricted range. An important example is where the mean of a Gaussian variable is constrained on physical grounds to be non-negative. This arises, for example, when the square of the neutrino mass is estimated from $\widehat{m}^2 = \widehat{E}^2 - \widehat{p}^2$, where $\widehat{E}$ and $\widehat{p}$ are independent, Gaussian distributed estimates of the energy and momentum. Although the true m^2 is constrained to be positive, random errors in $\widehat{E}$ and $\widehat{p}$ can easily lead to negative values for the estimate $\widehat{m}^2$.

If one uses the prescription given above for Gaussian distributed measurements, which says to construct the interval by taking the estimate plus or minus one standard deviation, then this can give intervals that are partially or entirely in the unphysical region. In fact, by following strictly the Neyman construction for the central confidence interval, one finds that the interval is truncated below zero; nevertheless an extremely small or even a zero-length interval can result.

An additional important example is where the experiment consists of counting a certain number of events, n, which is assumed to be Poisson distributed. Suppose the expectation value $E[n] = \nu$ is equal to $s + b$, where s and b are the means for signal and background processes, and assume further that b is a known constant. Then $\widehat{s} = n - b$ is an unbiased estimator for s. Depending on true magnitudes of s and b, the estimate $\widehat{s}$ can easily fall in the negative region. Similar to the Gaussian case with the positive mean, the central confidence interval or even the upper limit for s may be of zero length.

The confidence interval is in fact designed not to cover the parameter with a probability of at most α, and if a zero-length interval results, then this is evidently one of those experiments. So although the construction is behaving as it should, a null interval is an unsatisfying result to report and several solutions to this type of problem are possible.

An additional difficulty arises when a parameter estimate is not significantly far away from the boundary, in which case it is natural to report a one-sided confidence interval (often an upper limit). It is straightforward to force the Neyman prescription to produce only an upper limit by setting $x_2 = \infty$ in Eq. 32.39. Then x_1 is uniquely determined and the upper limit can be obtained. If, however, the data come out such that the parameter estimate is not so close to the boundary, one might wish to report a central (*i.e.*, two-sided) confidence interval. As pointed out by Feldman and Cousins [14], however, if the decision to report an upper limit or two-sided interval is made by looking at the data ("flip-flopping"), then the resulting intervals will not in general cover the parameter with the probability $1 - \alpha$.

With the confidence intervals suggested in [14], the prescription determines whether the interval is one- or two-sided in a way which preserves the coverage probability. Intervals with this property are said to be *unified*. Furthermore, the Feldman–Cousins prescription is such that null intervals do not occur. For a given choice of $1 - \alpha$, if the parameter estimate is sufficiently close to the boundary, then the method gives a one-sided limit. In the case of a Poisson variable in the presence of background, for example, this would occur if the number of observed events is compatible with the expected background. For parameter estimates increasingly far away from the boundary, i.e., for increasing signal significance, the interval makes a smooth transition from one- to two-sided, and far away from the boundary one obtains a central interval.

The intervals according to this method for the mean of Poisson variable in the absence of background are given in Table 32.4. (Note that α in [14] is defined following Neyman [13] as the coverage probability; this is opposite the modern convention used here in which the coverage probability is $1 - \alpha$.) The values of $1 - \alpha$ given here refer to the coverage of the true parameter by the whole interval $[\nu_1, \nu_2]$. In Table 32.3 for the one-sided upper and lower limits, however, $1 - \alpha$ refers to the probability to have individually $\nu_{\rm up} \geq \nu$ or $\nu_{\rm lo} \leq \nu$.

A potential difficulty with unified intervals arises if, for example, one constructs such an interval for a Poisson parameter s of some yet to be discovered signal process with, say, $1 - \alpha = 0.9$. If the true signal parameter is zero, or in any case much less than the expected background, one will usually obtain a one-sided upper limit on s. In a certain fraction of the experiments, however, a two-sided interval for s will result. Since, however, one typically chooses $1 - \alpha$ to be only 0.9 or 0.95 when searching for a new effect, the value $s = 0$ may be excluded from the interval before the existence of the effect is well established. It must then be communicated carefully that in excluding $s = 0$ from the interval, one is not necessarily claiming to have discovered the effect.

Table 32.4: Unified confidence intervals $[\nu_1, \nu_2]$ for a the mean of a Poisson variable given n observed events in the absence of background, for confidence levels of 90% and 95%.

	$1-\alpha$ =90%		$1-\alpha$ =95%	
n	ν_1	ν_2	ν_1	ν_2
0	0.00	2.44	0.00	3.09
1	0.11	4.36	0.05	5.14
2	0.53	5.91	0.36	6.72
3	1.10	7.42	0.82	8.25
4	1.47	8.60	1.37	9.76
5	1.84	9.99	1.84	11.26
6	2.21	11.47	2.21	12.75
7	3.56	12.53	2.58	13.81
8	3.96	13.99	2.94	15.29
9	4.36	15.30	4.36	16.77
10	5.50	16.50	4.75	17.82

The intervals constructed according to the unified procedure in [14] for a Poisson variable n consisting of signal and background have the property that for $n = 0$ observed events, the upper limit decreases for increasing expected background. This is counter-intuitive, since it is known that if $n = 0$ for the experiment in question, then no background was observed, and therefore one may argue that the expected background should not be relevant. The extent to which one should regard this feature as a drawback is a subject of some controversy (see, *e.g.*, Ref. [18]).

Another possibility is to construct a Bayesian interval as described in Section 32.3.1. The presence of the boundary can be incorporated simply by setting the prior density to zero in the unphysical region. Priors based on invariance principles (rather than subjective degree of belief) for the Poisson mean are rarely used in high energy physics; they diverge for the case of zero events observed, and they give upper limits which undercover when evaluated by the frequentist definition of coverage [2]. Rather, priors uniform in the Poisson mean have been used, although as previously mentioned, this is generally not done to reflect the experimenter's degree of belief but rather as a procedure for obtaining an interval with certain frequentist properties. The resulting upper limits have a coverage probability that depends on the true value of the Poisson parameter and is everywhere greater than the stated probability content. Lower limits and two-sided intervals for the Poisson mean based on flat priors undercover, however, for some values of the parameter, although to an extent that in practical cases may not be too severe [2, 11]. Intervals constructed in this way have the advantage of being easy to derive; if several independent measurements are to be combined then one simply multiplies the likelihood functions (cf. Eq. (32.48)).

An additional alternative is presented by the intervals found from the likelihood function or χ^2 using the prescription of Equations (32.46) or (32.47). As in the case of the Bayesian intervals, the coverage probability is not, in general, independent of the true parameter. Furthermore, these intervals can for some parameter values undercover. The coverage probability can of course be determined with some extra effort and reported with the result.

Also as in the Bayesian case, intervals derived from the value of the likelihood function from a combination of independent experiments can be determined simply by multiplying the likelihood functions. These intervals are also invariant under transformation of the parameter; this is not true for Bayesian intervals with a conventional flat prior, because a uniform distribution in, say, θ will not be uniform if transformed to θ^2. Use of the likelihood function to determine approximate confidence intervals is discussed further in [17].

In any case it is important always to report sufficient information so that the result can be combined with other measurements. Often this means giving an unbiased estimator and its standard deviation, even if the estimated value is in the unphysical region.

Regardless of the type of interval reported, the consumer of that result will almost certainly use it to derive some impression about the value of the parameter. This will inevitably be done, either explicitly or intuitively, with Bayes' theorem,

$$p(\theta|\text{result}) \propto L(\text{result}|\theta)\pi(\theta) \; , \qquad (32.51)$$

where the reader supplies his or her own prior beliefs $\pi(\theta)$ about the parameter, and the 'result' is whatever sort of interval or other information the author has reported. For all of the intervals discussed, therefore, it is not sufficient to know the result; one must also know the probability to have obtained this result as a function of the parameter, *i.e.*, the likelihood. Contours of constant likelihood, for example, provide this information, and so an interval obtained from $\ln L = \ln L_{\max} - \Delta \ln L$ already takes one step in this direction.

It can also be useful with a frequentist interval to calculate its subjective probability content using the posterior p.d.f. based on one or several reasonable guesses for the prior p.d.f. If it turns out to be significantly less than the stated confidence level, this warns that it would be particularly misleading to draw conclusions about the parameter's value without further information from the likelihood.

References:

1. B. Efron, Am. Stat. **40**, 11 (1986).
2. R.D. Cousins, Am. J. Phys. **63**, 398 (1995).
3. A. Stuart, A.K. Ord, and Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2 *Classical Inference and Relationship* 6th Ed., (Oxford Univ. Press, 1998), and earlier editions by Kendall and Stuart. The likelihood-ratio ordering principle is described at the beginning of Ch. 23. Chapter 26 compares different schools of statistical inference.
4. W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam and London, 1971).
5. H. Cramér, *Mathematical Methods of Statistics,* Princeton Univ. Press, New Jersey (1958).
6. L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, New York, 1986).
7. R. Barlow, Nucl. Inst. Meth. A **297**, 496 (1990).
8. G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
9. For a review, see S. Baker and R. Cousins, Nucl. Instrum. Methods **221**, 437 (1984).
10. For information on neural networks and related topics, see *e.g.* C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford (1995); C. Peterson and T. Rögnvaldsson, An Introduction to Artificial Neural Networks, in *Proceedings of the 1991 CERN School of Computing*, C. Verkerk (ed.), CERN 92-02 (1992).
11. Byron P. Roe and Michael B. Woodroofe, Phys. Rev. **D63**, 13009 (2001).
12. Paul H. Garthwaite, Ian T. Jolliffe and Byron Jones, *Statistical Inference* (Prentice Hall, 1995).
13. J. Neyman, Phil. Trans. Royal Soc. London, Series A, **236**, 333 (1937), reprinted in *A Selection of Early Statistical Papers on J. Neyman* (University of California Press, Berkeley, 1967).
14. G.J. Feldman and R.D. Cousins, Phys. Rev. **D57**, 3873 (1998). This paper does not specify what to do if the ordering principle gives equal rank to some values of x. Eq. 23.6 of Ref. 3 gives the rule: all such points are included in the acceptance region (the domain $D(\alpha)$). Some authors have assumed the contrary, and shown that one can then obtain null intervals.
15. T. Junk, Nucl. Inst. Meth. A **434**, 435 (1999).
16. A.L. Read, *Modified frequentist analysis of search results (the CL_s method)*, in F. James, L. Lyons and Y. Perrin (eds.), *Workshop on Confidence Limits*, CERN Yellow Report 2000-005, available through `weblib.cern.ch`.
17. F. Porter, Nucl. Inst. Meth. A **368**, 793 (1996).
18. Workshop on Confidence Limits, CERN, 17-18 Jan. 2000, `www.cern.ch/CERN/Divisions/EP/Events/CLW/`. The proceedings, F. James, L. Lyons, and Y. Perrin (eds.), CERN Yellow Report 2000-005, are available through `weblib.cern.ch`. See also the later Fermilab workshop linked to the CERN web page.

33. MONTE CARLO TECHNIQUES

Revised July 1995 by S. Youssef (SCRI, Florida State University).Updated February 2000 by R. Cousins (UCLA) in consultation with F. James (CERN); October 2003 by G. Cowan (RHUL) and R. Miquel (LBNL), and September 2005 by G. Cowan (RHUL).

Monte Carlo techniques are often the only practical way to evaluate difficult integrals or to sample random variables governed by complicated probability density functions. Here we describe an assortment of methods for sampling some commonly occurring probability density functions.

33.1. Sampling the uniform distribution

Most Monte Carlo sampling or integration techniques assume a "random number generator" which generates uniform statistically independent values on the half open interval $[0, 1)$. There is a long history of problems with various generators on a finite digital computer, but recently, the RANLUX generator [1] has emerged with a solid theoretical basis in chaos theory. Based on the method of Lüscher, it allows the user to select different quality levels, trading off quality with speed.

Other generators are also available which pass extensive batteries of tests for statistical independence and which have periods which are so long that, for practical purposes, values from these generators can be considered to be uniform and statistically independent. In particular, the lagged-Fibonacci based generator introduced by Marsaglia, Zaman, and Tsang [2] is efficient, has a period of approximately 10^{43}, produces identical sequences on a wide variety of computers and, passes the extensive "DIEHARD" battery of tests [3]. Many commonly available congruential generators fail these tests and often have sequences (typically with periods less than 2^{32}) which can be easily exhausted on modern computers and should therefore be avoided [4].

33.2. Inverse transform method

If the desired probability density function is $f(x)$ on the range $-\infty < x < \infty$, its cumulative distribution function (expressing the probability that $x \leq a$) is given by Eq. (31.6). If a is chosen with probability density $f(a)$, then the integrated probability up to point a, $F(a)$, is itself a random variable which will occur with uniform probability density on $[0, 1]$. If x can take on any value, and ignoring the endpoints, we can then find a unique x chosen from the p.d.f. $f(s)$ for a given u if we set

$$u = F(x) \ , \tag{33.1}$$

provided we can find an inverse of F, defined by

$$x = F^{-1}(u) \ . \tag{33.2}$$

This method is shown in Fig. 33.1a. It is most convenient when one can calculate by hand the inverse function of the indefinite integral of f. This is the case for some common functions $f(x)$ such as $\exp(x)$, $(1-x)^n$, and $1/(1+x^2)$ (Cauchy or Breit-Wigner), although it does not necessarily produce the fastest generator. CERNLIB contains routines to implement this method numerically, working from functions or histograms.

For a discrete distribution, $F(x)$ will have a discontinuous jump of size $f(x_k)$ at each allowed $x_k, k = 1, 2, \cdots$. Choose u from a uniform distribution on (0,1) as before. Find x_k such that

$$F(x_{k-1}) < u \leq F(x_k) \equiv \text{Prob}\ (x \leq x_k) = \sum_{i=1}^{k} f(x_i) \ ; \tag{33.3}$$

then x_k is the value we seek (note: $F(x_0) \equiv 0$). This algorithm is illustrated in Fig. 33.1b.

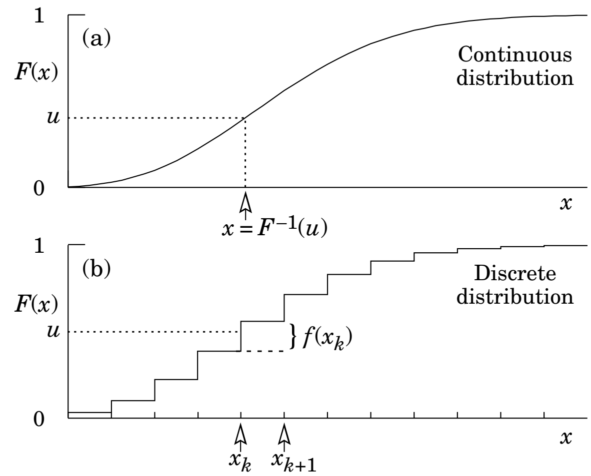


Figure 33.1: Use of a random number u chosen from a uniform distribution (0,1) to find a random number x from a distribution with cumulative distribution function $F(x)$.

33.3. Acceptance-rejection method (Von Neumann)

Very commonly an analytic form for $F(x)$ is unknown or too complex to work with, so that obtaining an inverse as in Eq. (33.2) is impractical. We suppose that for any given value of x the probability density function $f(x)$ can be computed and further that enough is known about $f(x)$ that we can enclose it entirely inside a shape which is C times an easily generated distribution $h(x)$ as illustrated in Fig. 33.2.

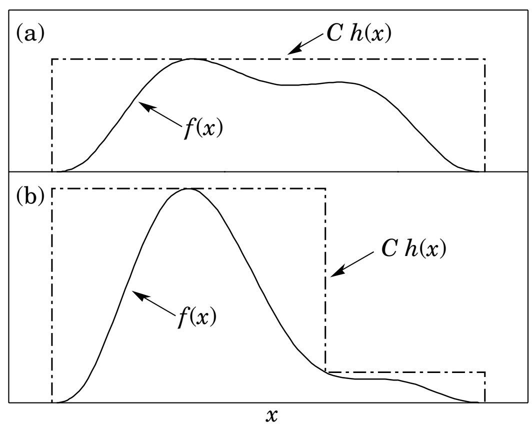


Figure 33.2: Illustration of the acceptance-rejection method. Random points are chosen inside the upper bounding figure, and rejected if the ordinate exceeds $f(x)$. Lower figure illustrates importance sampling.

Frequently $h(x)$ is uniform or is a normalized sum of uniform distributions. Note that both $f(x)$ and $h(x)$ must be normalized to unit area and therefore the proportionality constant $C > 1$. To generate $f(x)$, first generate a candidate x according to $h(x)$. Calculate $f(x)$ and the height of the envelope $C\,h(x)$; generate u and test if $uC\,h(x) \leq f(x)$. If so, accept x; if not reject x and try again. If we regard x and $uC\,h(x)$ as the abscissa and ordinate of a point in a two-dimensional plot, these points will populate the entire area $C\,h(x)$ in a smooth manner; then we accept those which fall under $f(x)$. The efficiency is the ratio of areas, which must equal $1/C$; therefore we must keep C as close as possible to 1.0. Therefore we try to choose $C\,h(x)$ to be as close to $f(x)$ as convenience dictates, as in the lower part of Fig. 33.2. This practice is called importance sampling, because we generate more trial values of x in the region where $f(x)$ is most important.

33.4. Algorithms

Algorithms for generating random numbers belonging to many different distributions are given by Press [5], Ahrens and Dieter [6], Rubinstein [7], Everett and Cashwell [8], Devroye [9], and Walck [10]. For many distributions alternative algorithms exist, varying in complexity, speed, and accuracy. For time-critical applications, these algorithms may be coded in-line to remove the significant overhead often encountered in making function calls. Variables named "u" are assumed to be independent and uniform on [0,1). (Hence, u must be verified to be non-zero where relevant.)

In the examples given below, we use the notation for the variables and parameters given in Table 31.1.

33.4.1. ***Exponential decay*** :

This is a common application of the inverse transform method and uses the fact that if u is uniformly distributed in $[0,1]$ then $(1-u)$ is as well. Consider an exponential p.d.f. $f(t) = (1/\tau)\exp(-t/\tau)$ that is truncated so as to lie between two values, a and b, and renormalized to unit area. To generate decay times t according to this p.d.f., first let $\alpha = \exp(-a/\tau)$ and $\beta = \exp(-b/\tau)$; then generate u and let

$$t = -\tau \ln(\beta + u(\alpha - \beta)). \tag{33.4}$$

For $(a,b) = (0,\infty)$, we have simply $t = -\tau \ln u$. (See also Sec. 33.4.6.)

33.4.2. ***Isotropic direction in 3D*** :

Isotropy means the density is proportional to solid angle, the differential element of which is $d\Omega = d(\cos\theta)d\phi$. Hence $\cos\theta$ is uniform $(2u_1 - 1)$ and ϕ is uniform $(2\pi u_2)$. For alternative generation of $\sin\phi$ and $\cos\phi$, see the next subsection.

33.4.3. ***Sine and cosine of random angle in 2D*** :

Generate u_1 and u_2. Then $v_1 = 2u_1 - 1$ is uniform on (−1,1), and $v_2 = u_2$ is uniform on (0,1). Calculate $r^2 = v_1^2 + v_2^2$. If $r^2 > 1$, start over. Otherwise, the sine (S) and cosine (C) of a random angle (*i.e.*, uniformly distributed between zero and 2π) are given by

$$S = 2v_1 v_2/r^2 \quad \text{and} \quad C = (v_1^2 - v_2^2)/r^2 . \tag{33.5}$$

33.4.4. ***Gaussian distribution*** :

If u_1 and u_2 are uniform on (0,1), then

$$z_1 = \sin 2\pi u_1 \sqrt{-2\ln u_2} \quad \text{and} \quad z_2 = \cos 2\pi u_1 \sqrt{-2\ln u_2} \tag{33.6}$$

are independent and Gaussian distributed with mean 0 and $\sigma = 1$.

There are many faster variants of this basic algorithm. For example, construct $v_1 = 2u_1 - 1$ and $v_2 = 2u_2 - 1$, which are uniform on (−1,1). Calculate $r^2 = v_1^2 + v_2^2$, and if $r^2 > 1$ start over. If $r^2 < 1$, it is uniform on (0,1). Then

$$z_1 = v_1\sqrt{\frac{-2\ln r^2}{r^2}} \quad \text{and} \quad z_2 = v_2\sqrt{\frac{-2\ln r^2}{r^2}} \tag{33.7}$$

are independent numbers chosen from a normal distribution with mean 0 and variance 1. $z_i' = \mu + \sigma z_i$ distributes with mean μ and variance σ^2. A recent implementation of the fast algorithm of Leva Ref. 11 is in CERNLIB.

For a multivariate Gaussian with an $n \times n$ covariance matrix V, one can start by generating n independent Gaussian variables, $\{\eta_j\}$, with mean 0 and variance 1 as above. Then the new set $\{x_i\}$ is obtained as $x_i = \mu_i + \sum_j L_{ij}\eta_j$, where μ_i is the mean of x_i, and L_{ij} are the components of L, the unique lower triangular matrix that fulfils $V = LL^T$. The matrix L can be easily computed by the following recursive relation (Cholesky's method):

$$L_{jj} = \left(V_{jj} - \sum_{k=1}^{j-1} L_{jk}^2 \right)^{1/2} , \tag{33.8a}$$

$$L_{ij} = \frac{V_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk}}{L_{jj}} , \; j = 1,...,n \, ; \; i = j+1,...,n, \tag{33.8b}$$

where $V_{ij} = \rho_{ij}\sigma_i\sigma_j$ are the components of V. For $n = 2$ one has

$$L = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\,\sigma_2 \end{pmatrix} , \tag{33.9}$$

and therefore the correlated Gaussian variables are generated as $x_1 = \mu_1 + \sigma_1\eta_1$, $x_2 = \mu_2 + \rho\sigma_2\eta_1 + \sqrt{1-\rho^2}\,\sigma_2\eta_2$.

33.4.5. $\chi^2(n)$ **distribution** :

For n even, generate $n/2$ uniform numbers u_i; then

$$y = -2\ln\left(\prod_{i=1}^{n/2} u_i\right) \quad \text{is} \quad \chi^2(n) \; . \tag{33.10}$$

For n odd, generate $(n-1)/2$ uniform numbers u_i and one Gaussian z as in Sec. 33.4.4; then

$$y = -2\ln\left(\prod_{i=1}^{(n-1)/2} u_i\right) + z^2 \quad \text{is} \quad \chi^2(n) \; . \tag{33.11}$$

For $n \gtrsim 30$ the much faster Gaussian approximation for the χ^2 may be preferable: generate z as in Sec. 33.4.4. If $z \geq -\sqrt{2n-1}$ use $y = \left[z + \sqrt{2n-1}\right]^2/2$; otherwise reject.

33.4.6. ***Gamma distribution*** :

All of the following algorithms are given for $\lambda = 1$. For $\lambda \neq 1$, divide the resulting random number x by λ.

- If $k = 1$ (the *exponential* distribution), accept $x = -\ln u$. (See also Sec. 33.4.1.)
- If $0 < k < 1$, initialize with $v_1 = (e+k)/e$ (with $e = 2.71828...$ being the natural log base). Generate u_1, u_2. Define $v_2 = v_1 u_1$.

  Case 1: $v_2 \leq 1$. Define $x = v_2^{1/k}$. If $u_2 \leq e^{-x}$, accept x and stop, else restart by generating new u_1, u_2.
  Case 2: $v_2 > 1$. Define $x = -\ln([v_1 - v_2]/k)$. If $u_2 \leq x^{k-1}$, accept x and stop, else restart by generating new u_1, u_2. Note that, for $k < 1$, the probability density has a pole at $x = 0$, so that return values of zero due to underflow must be accepted or otherwise dealt with.

- Otherwise, if $k > 1$, initialize with $c = 3k - 0.75$. Generate u_1 and compute $v_1 = u_1(1-u_1)$ and $v_2 = (u_1 - 0.5)\sqrt{c/v_1}$. If $x = k + v_2 - 1 \leq 0$, go back and generate new u_1; otherwise generate u_2 and compute $v_3 = 64v_1^3 u_2^2$. If $v_3 \leq 1 - 2v_2^2/x$ **or** if $\ln v_3 \leq 2\{[k-1]\ln[x/(k-1)] - v_2\}$, accept x and stop; otherwise go back and generate new u_1.

33.4.7. ***Binomial distribution*** :

Begin with $k = 0$ and generate u uniform in $[0,1)$. Compute $P_k = (1-p)^n$ and store P_k into B. If $u \leq B$ accept $r_k = k$ and stop. Otherwise, increment k by one; compute the next P_k as $P_k \cdot (p/(1-p)) \cdot (n-k)/(k+1)$; add this to B. Again if $u \leq B$ accept $r_k = k$ and stop, otherwise iterate until a value is accepted. If $p > 1/2$ it will be more efficient to generate r from $f(r;n,q)$, *i.e.*, with p and q interchanged, and then set $r_k = n - r$.

33.4.8. ***Poisson distribution*** :

Iterate until a successful choice is made: Begin with $k = 1$ and set $A = 1$ to start. Generate u. Replace A with uA; if now $A < \exp(-\mu)$, where μ is the Poisson parameter, accept $n_k = k - 1$ and stop. Otherwise increment k by 1, generate a new u and repeat, always starting with the value of A left from the previous try. For large $\mu(\gtrsim 10)$ it may be satisfactory (and much faster) to approximate the Poisson distribution by a Gaussian distribution (see our Probability chapter, Sec. 31.4.3). Generate z from a Gaussian with zero mean and unit standard deviation; then use $x = \max(0, [\mu + z\sqrt{\mu} + 0.5])$ where [ ] signifies the greatest integer $\leq$ the expression. [12]

33.4.9. ***Student's t* distribution** :

For $n > 0$ degrees of freedom (n not necessarily integer), generate x from a Gaussian with mean 0 and $\sigma^2 = 1$ according to the method of 33.4.4. Next generate y, an independent gamma random variate, according to 33.4.6 with $\lambda = 1/2$ and $k = n/2$. Then $z = x/(\sqrt{y}/n)$ is distributed as a t with n degrees of freedom.

For the special case $n = 1$, the Breit-Wigner distribution, generate u_1 and u_2; set $v_1 = 2u_1 - 1$ and $v_2 = 2u_2 - 1$. If $v_1^2 + v_2^2 \leq 1$ accept $z = v_1/v_2$ as a Breit-Wigner distribution with unit area, center at 0.0, and FWHM 2.0. Otherwise start over. For center M_0 and FWHM Γ, use $W = z\Gamma/2 + M_0$.

References:

1. F. James, Comp. Phys. Comm. **79** 111 (1994), based on M. Lüscher, Comp. Phys. Comm. **79** 100 (1994). This generator is available as the CERNLIB routine V115, RANLUX.
2. G. Marsaglia, A. Zaman, and W.W. Tsang, *Towards a Universal Random Number Generator*, Supercomputer Computations Research Institute, Florida State University technical report FSU-SCRI-87-50 (1987). This generator is available as the CERNLIB routine V113, RANMAR, by F. Carminati and F. James.
3. Much of DIEHARD is described in: G. Marsaglia, *A Current View of Random Number Generators*, keynote address, *Computer Science and Statistics: 16th Symposium on the Interface*, Elsevier (1985).
4. Newer generators with periods even longer than the lagged-Fibonacci based generator are described in G. Marsaglia and A. Zaman, *Some Portable Very-Long-Period Random Number Generators*, Compt. Phys. **8**, 117 (1994). The Numerical Recipes generator **ran2** [W.H. Press and S.A. Teukolsky, *Portable Random Number Generators*, Compt. Phys. **6**, 521 (1992)] is also known to pass the DIEHARD tests.
5. W.H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, New York, 1986).
6. J.H. Ahrens and U. Dieter, Computing **12**, 223 (1974).
7. R.Y. Rubinstein, *Simulation and the Monte Carlo Method* (John Wiley and Sons, Inc., New York, 1981).
8. C.J. Everett and E.D. Cashwell, *A Third Monte Carlo Sampler*, Los Alamos report LA-9721-MS (1983).
9. L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).
10. Ch. Walck, *Random Number Generation*, University of Stockholm Physics Department Report 1987-10-20 (Vers. 3.0).
11. J.L. Leva, ACM Trans. Math. Softw. **18** 449 (1992). This generator has been implemented by F. James in the CERNLIB routine V120, RNORML.
12. This generator has been implemented by D. Drijard and K. Kölbig in the CERNLIB routine V136, RNPSSN.

34. MONTE CARLO PARTICLE NUMBERING SCHEME

Revised June 2006 by L. Garren (Fermilab), I.G. Knowles (Edinburgh U.), S. Navas (U. Granada), P. Richardson (Durham U.), T. Sjöstrand (Lund U.), and T. Trippe (LBNL).

The Monte Carlo particle numbering scheme presented here is intended to facilitate interfacing between event generators, detector simulators, and analysis packages used in particle physics. The numbering scheme was introduced in 1988 [1] and a revised version [2,3] was adopted in 1998 in order to allow systematic inclusion of quark model states which are as yet undiscovered and hypothetical particles such as SUSY particles. The numbering scheme is used in several event generators, *e.g.* HERWIG and PYTHIA/JETSET, and in the /HEPEVT/ [4] standard interface.

The general form is a 7–digit number:

$$\pm n\, n_r\, n_L\, n_{q_1}\, n_{q_2}\, n_{q_3}\, n_J \ .$$

This encodes information about the particle's spin, flavor content, and internal quantum numbers. The details are as follows:

1. Particles are given positive numbers, antiparticles negative numbers. The PDG convention for mesons is used, so that K^+ and B^+ are particles.
2. Quarks and leptons are numbered consecutively starting from 1 and 11 respectively; to do this they are first ordered by family and within families by weak isospin.
3. In composite quark systems (diquarks, mesons, and baryons) $n_{q_{1-3}}$ are quark numbers used to specify the quark content, while the rightmost digit $n_J = 2J + 1$ gives the system's spin (except for the K^0_S and K^0_L). The scheme does not cover particles of spin $J > 4$.
4. Diquarks have 4-digit numbers with $n_{q_1} \geq n_{q_2}$ and $n_{q_3} = 0$.
5. The numbering of mesons is guided by the nonrelativistic (L–S decoupled) quark model, as listed in Tables 14.2 and 14.3.
   a. The numbers specifying the meson's quark content conform to the convention $n_{q_1} = 0$ and $n_{q_2} \geq n_{q_3}$. The special case K^0_L is the sole exception to this rule.
   b. The quark numbers of flavorless, light (u, d, s) mesons are: 11 for the member of the isotriplet $(\pi^0, \rho^0, \ldots)$, 22 for the lighter isosinglet $(\eta, \omega, \ldots)$, and 33 for the heavier isosinglet $(\eta', \phi, \ldots)$. Since isosinglet mesons are often large mixtures of $u\overline{u} + d\overline{d}$ and $s\overline{s}$ states, 22 and 33 are assigned by mass and do not necessarily specify the dominant quark composition.
   c. The special numbers 310 and 130 are given to the K^0_S and K^0_L respectively.
   d. The fifth digit n_L is reserved to distinguish mesons of the same total (J) but different spin (S) and orbital (L) angular momentum quantum numbers. For $J > 0$ the numbers are: $(L, S) = (J-1, 1)$ $n_L = 0$, $(J, 0)$ $n_L = 1$, $(J, 1)$ $n_L = 2$ and $(J+1, 1)$ $n_L = 3$. For the exceptional case $J = 0$ the numbers are $(0, 0)$ $n_L = 0$ and $(1, 1)$ $n_L = 1$ (*i.e.* $n_L = L$). See Table 34.1.

Table 34.1: Meson numbering logic. Here qq stands for $n_{q2}\, n_{q3}$.

	$L = J-1, S = 1$			$L = J, S = 0$			$L = J, S = 1$			$L = J+1, S = 1$		
J	code	J^{PC}	L	code	J^{PC}	L	code	J^{PC}	L	code	J^{PC}	L
0	—	—	—	$00qq1$	0^{-+}	0	—	—	—	$10qq1$	0^{++}	1
1	$00qq3$	1^{--}	0	$10qq3$	1^{+-}	1	$20qq3$	1^{++}	1	$30qq3$	1^{--}	2
2	$00qq5$	2^{++}	1	$10qq5$	2^{-+}	2	$20qq5$	2^{--}	2	$30qq5$	2^{++}	3
3	$00qq7$	3^{--}	2	$10qq7$	3^{+-}	3	$20qq7$	3^{++}	3	$30qq7$	3^{--}	4
4	$00qq9$	4^{++}	3	$10qq9$	4^{-+}	4	$20qq9$	4^{--}	4	$30qq9$	4^{++}	5

   e. If a set of physical mesons correspond to a (non-negligible) mixture of basis states, differing in their internal quantum numbers, then the lightest physical state gets the smallest basis state number. For example the $K_1(1270)$ is numbered 10313 (1^1P_1 K_{1B}) and the $K_1(1400)$ is numbered 20313 (1^3P_1 K_{1A}).
   f. The sixth digit n_r is used to label mesons radially excited above the ground state.
   g. Numbers have been assigned for complete $n_r = 0$ S- and P-wave multiplets, even where states remain to be identified.
   h. In some instances assignments within the $q\bar{q}$ meson model are only tentative; here best guess assignments are made.
   i. Many states appearing in the Meson Listings are not yet assigned within the $q\bar{q}$ model. Here $n_{q_{2-3}}$ and n_J are assigned according to the state's likely flavors and spin; all such unassigned light isoscalar states are given the flavor code 22. Within these groups $n_L = 0, 1, 2, \ldots$ is used to distinguish states of increasing mass. These states are flagged using $n = 9$. It is to be expected that these numbers will evolve as the nature of the states are elucidated. Codes are assigned to all mesons which are listed in the one-page table at the end of the Meson Summary Table as long as they have a prefered or established spin. Additional heavy meson states expected from heavy quark spectroscopy are also assigned codes.
6. The numbering of baryons is again guided by the nonrelativistic quark model, see Table 14.6.
   a. The numbers specifying a baryon's quark content are such that in general $n_{q_1} \geq n_{q_2} \geq n_{q_3}$.
   b. Two states exist for $J = 1/2$ baryons containing 3 different types of quarks. In the lighter baryon $(\Lambda, \Xi, \Omega, \ldots)$ the light quarks are in an antisymmetric ($J = 0$) state while for the heavier baryon $(\Sigma^0, \Xi', \Omega', \ldots)$ they are in a symmetric ($J = 1$) state. In this situation n_{q_2} and n_{q_3} are reversed for the lighter state, so that the smaller number corresponds to the lighter baryon.
   c. At present most Monte Carlos do not include excited baryons and no systematic scheme has been developed to denote them, though one is foreseen. In the meantime, use of the PDG 96 [5] numbers for excited baryons is recommended.
   d. For pentaquark states $n = 9$, $n_r n_L n_{q_1} n_{q_2}$ gives the four quark numbers in order $n_r \geq n_L \geq n_{q_1} \geq n_{q_2}$, n_{q_3} gives the antiquark number, and $n_J = 2J + 1$, with the assumption that $J = 1/2$ for the states currently reported.
7. The gluon, when considered as a gauge boson, has official number 21. In codes for glueballs, however, 9 is used to allow a notation in close analogy with that of hadrons.
8. The pomeron and odderon trajectories and a generic reggeon trajectory of states in QCD are assigned codes 990, 9990, and 110 respectively, where the final 0 indicates the indeterminate nature of the spin, and the other digits reflect the expected "valence" flavor content. We do not attempt a complete classification of all reggeon trajectories, since there is currently no need to distinguish a specific such trajectory from its lowest-lying member.
9. Two-digit numbers in the range 21–30 are provided for the Standard Model gauge bosons and Higgs.
10. Codes 81–100 are reserved for generator-specific pseudoparticles and concepts.
11. The search for physics beyond the Standard Model is an active area, so these codes are also standardized as far as possible.
    a. A standard fourth generation of fermions is included by analogy with the first three.
    b. The graviton and the boson content of a two-Higgs-doublet scenario and of additional SU(2)×U(1) groups are found in the range 31–40.
    c. "One-of-a-kind" exotic particles are assigned numbers in the range 41–80.
    d. Fundamental supersymmetric particles are identified by adding a nonzero n to the particle number. The superpartner of a boson or a left-handed fermion has $n = 1$ while the superpartner of a right-handed fermion has $n = 2$. When mixing occurs, such as between the winos and charged Higgsinos to give charginos, or between left and right sfermions, the lighter physical state is given the smaller basis state number.
    e. Technicolor states have $n = 3$, with technifermions treated like ordinary fermions. States which are ordinary color singlets have $n_r = 0$. Color octets have $n_r = 1$. If a state has non-trivial quantum numbers under the topcolor groups

SU(3)$_1$ × SU(3)$_2$, the quantum numbers are specified by tech,ij, where i and j are 1 or 2. n_L is then $2i + j$. The coloron, V_8, is a heavy gluon color octet and thus is 3100021.

f. Excited (composite) quarks and leptons are identified by setting $n = 4$.

g. Within several scenarios of new physics, it is possible to have colored particles sufficiently long-lived for color-singlet hadronic states to form around them. In the context of supersymmetric scenarios, these states are called R-hadrons, since they carry odd R-parity. R-hadron codes, defined here, should be viewed as templates for corresponding codes also in other scenarios, for any long-lived particle that is either an unflavored color octet or a flavored color triplet. The R-hadron code is obtained by combining the SUSY particle code with a code for the light degrees of freedom, with as many intermediate zeros removed from the former as required to make place for the latter at the end. (To exemplify, a sparticle $n00000n_{\tilde{q}}$ combined with quarks q_1 and q_2 obtains code $n00n_{\tilde{q}}n_{q_1}n_{q_2}n_J$.) Specifically, the new-particle spin decouples in the limit of large masses, so that the final n_J digit is defined by the spin state of the light-quark system alone. An appropriate number of n_q digits is used to define the ordinary-quark content. As usual, 9 rather than 21 is used to denote a gluon/gluino in composite states. The sign of the hadron agrees with that of the constituent new particle (a color triplet) where there is a distinct new antiparticle, and else is defined as for normal hadrons. Particle names are R with the flavor content as lower index. A non-exhaustive list of R-hadron codes is given below.

12. Occasionally program authors add their own states. To avoid confusion, these should be flagged by setting $nn_r = 99$.
13. Concerning the non-99 numbers, it may be noted that only quarks, excited quarks, squarks, and diquarks have $n_{q_3} = 0$; only diquarks, baryons (including pentaquarks), and the odderon have $n_{q_1} \neq 0$; and only mesons, the reggeon, and the pomeron have $n_{q_1} = 0$ and $n_{q_2} \neq 0$. Concerning mesons (not antimesons), if n_{q_1} is odd then it labels a quark and an antiquark if even.
14. Nuclear codes are given as 10-digit numbers $\pm 10LZZZAAAI$. For a (hyper)nucleus consisting of n_p protons, n_n neutrons and n_Λ Λ's, $A = n_p + n_n + n_\Lambda$ gives the total baryon number, $Z = n_p$ the total charge and $L = n_\Lambda$ the total number of strange quarks. I gives the isomer level, with $I = 0$ corresponding to the ground state and $I > 0$ to excitations, see [9], where states denoted m, n, p, q translate to $I = 1 - 4$. As examples, the deuteron is 1000010020 and ^{235}U is 1000922350. To avoid ambiguities, nuclear codes should not be applied to a single hadron, like p, n or Λ^0, where quark-contents-based codes already exist.

This text and lists of particle numbers can be found on the WWW [6]. The StdHep Monte Carlo standardization project [7] maintains the list of PDG particle numbers, as well as numbering schemes from most event generators and software to convert between the different schemes.

References:

1. G.P. Yost *et al.*, Particle Data Group, Phys. Lett. **B204**, 1 (1988).
2. I. G. Knowles *et al.*, in *"Physics at LEP2"*, CERN 96-01, vol. 2, p. 103.
3. C. Caso *et al.*, Particle Data Group, Eur. Phys. J. **C3**, 1 (1998).
4. T. Sjöstrand *et al.*, in *"Z physics at LEP1"*, CERN 89-08, vol. 3, p. 327.
5. R.M. Barnett *et al.*, PDG, Phys. Rev. **D54**, 1 (1996).
6. `pdg.lbl.gov/2006/mcdata/mc_particle_id_contents.html`.
7. L. Garren, StdHep, *Monte Carlo Standardization at FNAL*, Fermilab PM0091 and StdHep WWW site: `http://cepa.fnal.gov/psm/stdhep/`.
8. S. Eidelman *et al.*, PDG, Phys. Lett. **B592**, 1 (2004).
9. G. Audi *et al.*, Nucl. Phys. **A729**, 3 (2003) See also `http://www.nndc.bnl.gov/amdc/web/nubase_en.html`.

QUARKS

d	1
u	2
s	3
c	4
b	5
t	6
b'	7
t'	8

LEPTONS

e^-	11
ν_e	12
μ^-	13
ν_μ	14
τ^-	15
ν_τ	16
τ'^-	17
$\nu_{\tau'}$	18

EXCITED PARTICLES

d^*	4000001
u^*	4000002
e^*	4000011
ν_e^*	4000012

GAUGE AND HIGGS BOSONS

g	(9) 21
γ	22
Z^0	23
W^+	24
h^0/H_1^0	25
Z'/Z_2^0	32
Z''/Z_3^0	33
W'/W_2^+	34
H^0/H_2^0	35
A^0/H_3^0	36
H^+	37

DIQUARKS

$(dd)_1$	1103
$(ud)_0$	2101
$(ud)_1$	2103
$(uu)_1$	2203
$(sd)_0$	3101
$(sd)_1$	3103
$(su)_0$	3201
$(su)_1$	3203
$(ss)_1$	3303
$(cd)_0$	4101
$(cd)_1$	4103
$(cu)_0$	4201
$(cu)_1$	4203
$(cs)_0$	4301
$(cs)_1$	4303
$(cc)_1$	4403
$(bd)_0$	5101
$(bd)_1$	5103
$(bu)_0$	5201
$(bu)_1$	5203
$(bs)_0$	5301
$(bs)_1$	5303
$(bc)_0$	5401
$(bc)_1$	5403
$(bb)_1$	5503

TECHNICOLOR PARTICLES

$\pi^0_{\rm tech}$	3000111
$\pi^+_{\rm tech}$	3000211
$\pi'^0_{\rm tech}$	3000221
$\eta^0_{\rm tech}$	3100221
$\rho^0_{\rm tech}$	3000113
$\rho^+_{\rm tech}$	3000213
$\omega^0_{\rm tech}$	3000223
V_8	3100021
$\pi^1_{\rm tech,22}$	3060111
$\pi^8_{\rm tech,22}$	3160111
$\rho_{\rm tech,11}$	3130113
$\rho_{\rm tech,12}$	3140113
$\rho_{\rm tech,21}$	3150113
$\rho_{\rm tech,22}$	3160113

R-HADRONS

$R^0_{\widetilde{g}g}$	1000993
$R^0_{\widetilde{g}d\overline{d}}$	1009113
$R^+_{\widetilde{g}u\overline{d}}$	1009213
$R^0_{\widetilde{g}u\overline{u}}$	1009223
$R^0_{\widetilde{g}d\overline{s}}$	1009313
$R^+_{\widetilde{g}u\overline{s}}$	1009323
$R^0_{\widetilde{g}s\overline{s}}$	1009333
$R^-_{\widetilde{g}ddd}$	1091114
$R^0_{\widetilde{g}udd}$	1092114
$R^+_{\widetilde{g}uud}$	1092214
$R^{++}_{\widetilde{g}uuu}$	1092224
$R^-_{\widetilde{g}sdd}$	1093114
$R^0_{\widetilde{g}sud}$	1093214
$R^+_{\widetilde{g}suu}$	1093224
$R^-_{\widetilde{g}ssd}$	1093314
$R^0_{\widetilde{g}ssu}$	1093324
$R^-_{\widetilde{g}sss}$	1093334
$R^+_{\widetilde{t}_1\overline{d}}$	1000612
$R^0_{\widetilde{t}_1\overline{u}}$	1000622
$R^+_{\widetilde{t}_1\overline{s}}$	1000632
$R^0_{\widetilde{t}_1\overline{c}}$	1000642
$R^+_{\widetilde{t}_1\overline{b}}$	1000652
$R^0_{\widetilde{t}_1 dd_1}$	1006113
$R^+_{\widetilde{t}_1 ud_0}$	1006211
$R^+_{\widetilde{t}_1 ud_1}$	1006213
$R^{++}_{\widetilde{t}_1 uu_1}$	1006223
$R^0_{\widetilde{t}_1 sd_0}$	1006311
$R^0_{\widetilde{t}_1 sd_1}$	1006313
$R^+_{\widetilde{t}_1 su_0}$	1006321
$R^+_{\widetilde{t}_1 su_1}$	1006323
$R^0_{\widetilde{t}_1 ss_1}$	1006333

SUSY PARTICLES

$\widetilde{d}_L$	1000001
$\widetilde{u}_L$	1000002
$\widetilde{s}_L$	1000003
$\widetilde{c}_L$	1000004
$\widetilde{b}_1$	1000005[a]
$\widetilde{t}_1$	1000006[a]
$\widetilde{e}^-_L$	1000011
$\widetilde{\nu}_{eL}$	1000012
$\widetilde{\mu}^-_L$	1000013
$\widetilde{\nu}_{\mu L}$	1000014
$\widetilde{\tau}^-_1$	1000015[a]
$\widetilde{\nu}_{\tau L}$	1000016
$\widetilde{d}_R$	2000001
$\widetilde{u}_R$	2000002
$\widetilde{s}_R$	2000003
$\widetilde{c}_R$	2000004
$\widetilde{b}_2$	2000005[a]
$\widetilde{t}_2$	2000006[a]
$\widetilde{e}^-_R$	2000011
$\widetilde{\mu}^-_R$	2000013
$\widetilde{\tau}^-_2$	2000015[a]
$\widetilde{g}$	1000021
$\widetilde{\chi}^0_1$	1000022[b]
$\widetilde{\chi}^0_2$	1000023[b]
$\widetilde{\chi}^+_1$	1000024[b]
$\widetilde{\chi}^0_3$	1000025[b]
$\widetilde{\chi}^0_4$	1000035[b]
$\widetilde{\chi}^+_2$	1000037[b]
$\widetilde{G}$	1000039

SPECIAL PARTICLES

G (graviton)	39
R^0	41
LQ[c]	42
reggeon	110
pomeron	990
odderon	9990

for MC internal use 81–100

LIGHT $I = 1$ MESONS

π^0	111
π^+	211
$a_0(980)^0$	9000111
$a_0(980)^+$	9000211
$\pi(1300)^0$	100111
$\pi(1300)^+$	100211
$a_0(1450)^0$	10111
$a_0(1450)^+$	10211
$\pi(1800)^0$	9010111
$\pi(1800)^+$	9010211
$\rho(770)^0$	113
$\rho(770)^+$	213
$b_1(1235)^0$	10113
$b_1(1235)^+$	10213
$a_1(1260)^0$	20113
$a_1(1260)^+$	20213
$\pi_1(1400)^0$	9000113
$\pi_1(1400)^+$	9000213
$\rho(1450)^0$	100113
$\rho(1450)^+$	100213
$\pi_1(1600)^0$	9010113
$\pi_1(1600)^+$	9010213
$\boldsymbol{a_1(1640)^0}$	**9020113***
$\boldsymbol{a_1(1640)^+}$	**9020213***
$\rho(1700)^0$	30113
$\rho(1700)^+$	30213
$\rho(1900)^0$	**9030113***
$\rho(1900)^+$	**9030213***
$\rho(2150)^0$	**9040113***
$\rho(2150)^+$	**9040213***
$a_2(1320)^0$	115
$a_2(1320)^+$	215
$\pi_2(1670)^0$	10115
$\pi_2(1670)^+$	10215
$\boldsymbol{a_2(1700)^0}$	**9000115***
$\boldsymbol{a_2(1700)^+}$	**9000215***
$\pi_2(2100)^0$	**9010115***
$\pi_2(2100)^+$	**9010215***
$\rho_3(1690)^0$	117
$\rho_3(1690)^+$	217
$\rho_3(1990)^0$	9000117
$\rho_3(1990)^+$	9000217
$\rho_3(2250)^0$	9010117
$\rho_3(2250)^+$	9010217
$a_4(2040)^0$	119
$a_4(2040)^+$	219

LIGHT $I = 0$ MESONS
($u\overline{u}$, $d\overline{d}$, and $s\overline{s}$ Admixtures)

η	221
$\eta'(958)$	331
$f_0(600)$	9000221
$f_0(980)$	9010221
$\eta(1295)$	100221
$f_0(1370)$	10221
$\eta(1405)$	9020221
$\eta(1475)$	100331
$f_0(1500)$	9030221
$f_0(1710)$	10331
$\boldsymbol{\eta(1760)}$	**9040221***
$f_0(2020)$	**9050221***
$f_0(2100)$	**9060221***
$f_0(2200)$	**9070221***
$\boldsymbol{\eta(2225)}$	**9080221***
$\omega(782)$	223
$\phi(1020)$	333
$h_1(1170)$	10223
$f_1(1285)$	20223
$h_1(1380)$	10333
$f_1(1420)$	20333
$\omega(1420)$	100223
$f_1(1510)$	9000223
$\boldsymbol{h_1(1595)}$	**9010223***
$\omega(1650)$	30223
$\phi(1680)$	100333
$f_2(1270)$	225
$f_2(1430)$	9000225
$f'_2(1525)$	335
$f_2(1565)$	9010225
$f_2(1640)$	9020225
$\eta_2(1645)$	10225
$f_2(1810)$	9030225
$\eta_2(1870)$	10335
$f_2(1910)$	9040225
$f_2(1950)$	9050225
$f_2(2010)$	9060225
$f_2(2150)$	9070225
$f_2(2300)$	9080225
$f_2(2340)$	9090225
$\omega_3(1670)$	227
$\phi_3(1850)$	337
$f_4(2050)$	229
$f_J(2220)$	9000229
$f_4(2300)$	9010229

STRANGE MESONS

K^0_L	130
K^0_S	310
K^0	311
K^+	321
$\boldsymbol{K^*_0(800)^0}$	**9000311***
$\boldsymbol{K^*_0(800)^+}$	**9000321***
$K^*_0(1430)^0$	10311
$K^*_0(1430)^+$	10321
$K(1460)^0$	100311
$K(1460)^+$	100321
$\boldsymbol{K(1830)^0}$	**9010311***
$\boldsymbol{K(1830)^+}$	**9010321***
$K^*_0(1950)^0$	**9020311***
$K^*_0(1950)^+$	**9020321***
$K^*(892)^0$	313
$K^*(892)^+$	323
$K_1(1270)^0$	10313
$K_1(1270)^+$	10323
$K_1(1400)^0$	20313
$K_1(1400)^+$	20323
$K^*(1410)^0$	100313
$K^*(1410)^+$	100323
$\boldsymbol{K_1(1650)^0}$	**9000313***
$\boldsymbol{K_1(1650)^+}$	**9000323***
$K^*(1680)^0$	30313
$K^*(1680)^+$	30323
$K^*_2(1430)^0$	315
$K^*_2(1430)^+$	325
$K_2(1580)^0$	9000315
$K_2(1580)^+$	9000325
$K_2(1770)^0$	10315
$K_2(1770)^+$	10325
$K_2(1820)^0$	20315
$K_2(1820)^+$	20325
$\boldsymbol{K^*_2(1980)^0}$	**9010315***
$\boldsymbol{K^*_2(1980)^+}$	**9010325***
$K_2(2250)^0$	**9020315***
$K_2(2250)^+$	**9020325***
$K^*_3(1780)^0$	317
$K^*_3(1780)^+$	327
$K_3(2320)^0$	9010317
$K_3(2320)^+$	9010327
$K^*_4(2045)^0$	319
$K^*_4(2045)^+$	329
$K_4(2500)^0$	9000319
$K_4(2500)^+$	9000329

CHARMED MESONS

D^+	411
D^0	421
$\boldsymbol{D^*_0(2400)^+}$	10411
$\boldsymbol{D^*_0(2400)^0}$	10421
$D^*(2010)^+$	413
$D^*(2007)^0$	423
$D_1(2420)^+$	10413
$D_1(2420)^0$	10423
$D_1(H)^+$	20413
$\boldsymbol{D_1(2430)^0}$	20423
$D^*_2(2460)^+$	415
$D^*_2(2460)^0$	425
D^+_s	431
$D^*_{s0}(2317)^+$	10431
D^{*+}_s	433
$D_{s1}(2536)^+$	10433
$D_{s1}(2460)^+$	20433
$D^*_{s2}(2573)^+$	435

BOTTOM MESONS

B^0	511
B^+	521
B^{*0}_0	10511
B^{*+}_0	10521
B^{*0}	513
B^{*+}	523
$B_1(L)^0$	10513
$B_1(L)^+$	10523
$B_1(H)^0$	20513
$B_1(H)^+$	20523
B^{*0}_2	515
B^{*+}_2	525
B^0_s	531
B^{*0}_{s0}	10531
B^{*0}_s	533
$B_{s1}(L)^0$	10533
$B_{s1}(H)^0$	20533
B^{*0}_{s2}	535
B^+_c	541
B^{*+}_{c0}	10541
B^{*+}_c	543
$B_{c1}(L)^+$	10543
$B_{c1}(H)^+$	20543
B^{*+}_{c2}	545

$c\overline{c}$ MESONS

$\eta_c(1S)$	441
$\chi_{c0}(1P)$	10441
$\eta_c(2S)$	100441
$J/\psi(1S)$	443
$h_c(1P)$	10443
$\chi_{c1}(1P)$	20443
$\psi(2S)$	100443
$\psi(3770)$	30443
$\psi(4040)$	9000443
$\psi(4160)$	9010443
$\psi(4415)$	9020443
$\chi_{c2}(1P)$	445
$\boldsymbol{\chi_{c2}(2P)}$	**100445***

$b\overline{b}$ MESONS

$\eta_b(1S)$	551
$\chi_{b0}(1P)$	10551
$\eta_b(2S)$	100551
$\chi_{b0}(2P)$	110551
$\eta_b(3S)$	200551
$\chi_{b0}(3P)$	210551
$\Upsilon(1S)$	553
$h_b(1P)$	10553
$\chi_{b1}(1P)$	20553
$\Upsilon_1(1D)$	30553
$\Upsilon(2S)$	100553
$h_b(2P)$	110553
$\chi_{b1}(2P)$	120553
$\Upsilon_1(2D)$	130553
$\Upsilon(3S)$	200553
$h_b(3P)$	210553
$\chi_{b1}(3P)$	220553
$\Upsilon(4S)$	300553
$\Upsilon(10860)$	9000553
$\Upsilon(11020)$	9010553
$\chi_{b2}(1P)$	555
$\eta_{b2}(1D)$	10555
$\Upsilon_2(1D)$	20555
$\chi_{b2}(2P)$	100555
$\eta_{b2}(2D)$	110555
$\Upsilon_2(2D)$	120555
$\chi_{b2}(3P)$	200555
$\Upsilon_3(1D)$	557
$\Upsilon_3(2D)$	100557

LIGHT BARYONS

p	2212
n	2112
Δ^{++}	2224
Δ^+	2214
Δ^0	2114
Δ^-	1114

STRANGE BARYONS

Λ	3122
Σ^+	3222
Σ^0	3212
Σ^-	3112
Σ^{*+}	3224[d]
Σ^{*0}	3214[d]
Σ^{*-}	3114[d]
Ξ^0	3322
Ξ^-	3312
Ξ^{*0}	3324[d]
Ξ^{*-}	3314[d]
Ω^-	3334

CHARMED BARYONS

Λ^+_c	4122
Σ^{++}_c	4222
Σ^+_c	4212
Σ^0_c	4112
Σ^{*++}_c	4224
Σ^{*+}_c	4214
Σ^{*0}_c	4114
Ξ^+_c	4232
Ξ^0_c	4132
Ξ'^+_c	4322
Ξ'^0_c	4312
Ξ^{*+}_c	4324
Ξ^{*0}_c	4314
Ω^0_c	4332
Ω^{*0}_c	4334
Ξ^+_{cc}	4412
Ξ^{++}_{cc}	4422
Ξ^{*+}_{cc}	4414
Ξ^{*++}_{cc}	4424
Ω^+_{cc}	4432
Ω^{*+}_{cc}	4434
Ω^{++}_{ccc}	4444

PENTAQUARKS

Θ^+	9221132
Φ^{--}	9331122

BOTTOM BARYONS

Λ^0_b	5122
Σ^-_b	5112
Σ^0_b	5212
Σ^+_b	5222
Σ^{*-}_b	5114
Σ^{*0}_b	5214
Σ^{*+}_b	5224
Ξ^-_b	5132
Ξ^0_b	5232
Ξ'^-_b	5312
Ξ'^0_b	5322
Ξ^{*-}_b	5314
Ξ^{*0}_b	5324
Ω^-_b	5332
Ω^{*-}_b	5334
Ξ^0_{bc}	5142
Ξ^+_{bc}	5242
Ξ'^0_{bc}	5412
Ξ'^+_{bc}	5422
Ξ^{*0}_{bc}	5414
Ξ^{*+}_{bc}	5424
Ω^0_{bc}	5342
Ω'^0_{bc}	5432
Ω^{*0}_{bc}	5434
Ω^+_{bcc}	5442
Ω^{*+}_{bcc}	5444
Ξ^-_{bb}	5512
Ξ^0_{bb}	5522
Ξ^{*-}_{bb}	5514
Ξ^{*0}_{bb}	5524
Ω^-_{bb}	5532
Ω^{*-}_{bb}	5534
Ω^0_{bbc}	5542
Ω^{*0}_{bbc}	5544
Ω^-_{bbb}	5554

Footnotes to the Tables:

*) Numbers or names in bold face are new or have changed since the 2004 *Review* [8].

a) Particulary in the third generation, the left and right sfermion states may mix, as shown. The lighter mixed state is given the smaller number.

b) The physical $\widetilde{\chi}$ states are admixtures of the pure $\widetilde{\gamma}$, $\widetilde{Z}^0$, $\widetilde{W}^+$, $\widetilde{H}^0_1$, $\widetilde{H}^0_2$, and $\widetilde{H}^+$ states.

c) In this draft we have only provided one generic leptoquark code. More general classifications according to spin, weak isospin and flavor content would lead to a host of states, that could be added as the need arises.

d) Σ^* and Ξ^* are alternate names for $\Sigma(1385)$ and $\Xi(1530)$.

35. CLEBSCH-GORDAN COEFFICIENTS, SPHERICAL HARMONICS, AND d FUNCTIONS

Note: A square-root sign is to be understood over *every* coefficient, *e.g.*, for −8/15 read $-\sqrt{8/15}$.

Notation:

	J	J	...
	M	M	...
m_1 m_2			
m_1 m_2	Coefficients		
⋮ ⋮			

$$Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta$$

$$Y_1^1 = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\phi}$$

$$Y_2^0 = \sqrt{\frac{5}{4\pi}}\Big(\frac{3}{2}\cos^2\theta - \frac{1}{2}\Big)$$

$$Y_2^1 = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{i\phi}$$

$$Y_2^2 = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\, e^{2i\phi}$$

$$Y_\ell^{-m} = (-1)^m Y_\ell^{m*}$$

$$d^{\,\ell}_{m,0} = \sqrt{\frac{4\pi}{2\ell+1}}\, Y_\ell^m\, e^{-im\phi}$$

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 J M\rangle = (-1)^{J-j_1-j_2}\langle j_2 j_1 m_2 m_1 | j_2 j_1 J M\rangle$$

1/2×1/2

m_1 m_2	1, +1
+1/2 +1/2	1

m_1 m_2	1, 0	0, 0
+1/2 −1/2	1/2	1/2
−1/2 +1/2	1/2	−1/2

m_1 m_2	1, −1
−1/2 −1/2	1

1×1/2

m_1 m_2	3/2, +3/2
+1 +1/2	1

m_1 m_2	3/2, +1/2	1/2, +1/2
+1 −1/2	1/3	2/3
0 +1/2	2/3	−1/3

m_1 m_2	3/2, −1/2	1/2, −1/2
0 −1/2	2/3	1/3
−1 +1/2	1/3	−2/3

m_1 m_2	3/2, −3/2
−1 −1/2	1

2×1

m_1 m_2	3, +3
+2 +1	1

m_1 m_2	3, +2	2, +2
+2 0	1/3	2/3
+1 +1	2/3	−1/3

m_1 m_2	3, +1	2, +1	1, +1
+2 −1	1/15	1/3	3/5
+1 0	8/15	1/6	−3/10
0 +1	2/5	−1/2	1/10

m_1 m_2	3, 0	2, 0	1, 0
+1 −1	1/5	1/2	3/10
0 0	3/5	0	−2/5
−1 +1	1/5	−1/2	3/10

m_1 m_2	3, −1	2, −1	1, −1
0 −1	2/5	1/2	1/10
−1 0	8/15	−1/6	−3/10
−2 +1	1/15	−1/3	3/5

m_1 m_2	3, −2	2, −2
−1 −1	2/3	1/3
−2 0	1/3	−2/3

m_1 m_2	3, −3
−2 −1	1

1×1

m_1 m_2	2, +2
+1 +1	1

m_1 m_2	2, +1	1, +1
+1 0	1/2	1/2
0 +1	1/2	−1/2

m_1 m_2	2, 0	1, 0	0, 0
+1 −1	1/6	1/2	1/3
0 0	2/3	0	−1/3
−1 +1	1/6	−1/2	1/3

m_1 m_2	2, −1	1, −1
0 −1	1/2	1/2
−1 0	1/2	−1/2

m_1 m_2	2, −2
−1 −1	1

2×1/2

m_1 m_2	5/2, +5/2
+2 +1/2	1

m_1 m_2	5/2, +3/2	3/2, +3/2
+2 −1/2	1/5	4/5
+1 +1/2	4/5	−1/5

m_1 m_2	5/2, +1/2	3/2, +1/2
+1 −1/2	2/5	3/5
0 +1/2	3/5	−2/5

m_1 m_2	5/2, −1/2	3/2, −1/2
0 −1/2	3/5	2/5
−1 +1/2	2/5	−3/5

m_1 m_2	5/2, −3/2	3/2, −3/2
−1 −1/2	4/5	1/5
−2 +1/2	1/5	−4/5

m_1 m_2	5/2, −5/2
−2 −1/2	1

3/2×1/2

m_1 m_2	2, +2
+3/2 +1/2	1

m_1 m_2	2, +1	1, +1
+3/2 −1/2	1/4	3/4
+1/2 +1/2	3/4	−1/4

m_1 m_2	2, 0	1, 0
+1/2 −1/2	1/2	1/2
−1/2 +1/2	1/2	−1/2

m_1 m_2	2, −1	1, −1
−1/2 −1/2	3/4	1/4
−3/2 +1/2	1/4	−3/4

m_1 m_2	2, −2
−3/2 −1/2	1

3/2×1

m_1 m_2	5/2, +5/2
+3/2 +1	1

m_1 m_2	5/2, +3/2	3/2, +3/2
+3/2 0	2/5	3/5
+1/2 +1	3/5	−2/5

m_1 m_2	5/2, +1/2	3/2, +1/2	1/2, +1/2
+3/2 −1	1/10	2/5	1/2
+1/2 0	3/5	1/15	−1/3
−1/2 +1	3/10	−8/15	1/6

m_1 m_2	5/2, −1/2	3/2, −1/2	1/2, −1/2
+1/2 −1	3/10	8/15	1/6
−1/2 0	3/5	−1/15	−1/3
−3/2 +1	1/10	−2/5	1/2

m_1 m_2	5/2, −3/2	3/2, −3/2
−1/2 −1	3/5	2/5
−3/2 0	2/5	−3/5

m_1 m_2	5/2, −5/2
−3/2 −1	1

$$d^{\,j}_{m',m} = (-1)^{m-m'}d^{\,j}_{m,m'} = d^{\,j}_{-m,-m'}$$

3/2×3/2

m_1 m_2	3, +3
+3/2 +3/2	1

m_1 m_2	3, +2	2, +2
+3/2 +1/2	1/2	1/2
+1/2 +3/2	1/2	−1/2

m_1 m_2	3, +1	2, +1	1, +1
+3/2 −1/2	1/5	1/2	3/10
+1/2 +1/2	3/5	0	−2/5
−1/2 +3/2	1/5	−1/2	3/10

m_1 m_2	3, 0	2, 0	1, 0	0, 0
+3/2 −3/2	1/20	1/4	9/20	1/4
+1/2 −1/2	9/20	1/4	−1/20	−1/4
−1/2 +1/2	9/20	−1/4	−1/20	1/4
−3/2 +3/2	1/20	−1/4	9/20	−1/4

m_1 m_2	3, −1	2, −1	1, −1
+1/2 −3/2	1/5	1/2	3/10
−1/2 −1/2	3/5	0	−2/5
−3/2 +1/2	1/5	−1/2	3/10

m_1 m_2	3, −2	2, −2
−1/2 −3/2	1/2	1/2
−3/2 −1/2	1/2	−1/2

m_1 m_2	3, −3
−3/2 −3/2	1

2×3/2

m_1 m_2	7/2, +7/2
+2 +3/2	1

m_1 m_2	7/2, +5/2	5/2, +5/2
+2 +1/2	3/7	4/7
+1 +3/2	4/7	−3/7

m_1 m_2	7/2, +3/2	5/2, +3/2	3/2, +3/2
+2 −1/2	1/7	16/35	2/5
+1 +1/2	4/7	1/35	−2/5
0 +3/2	2/7	−18/35	1/5

m_1 m_2	7/2, +1/2	5/2, +1/2	3/2, +1/2	1/2, +1/2
+2 −3/2	1/35	6/35	2/5	2/5
+1 −1/2	12/35	5/14	0	−3/10
0 +1/2	18/35	−3/35	−1/5	1/5
−1 +3/2	4/35	−27/70	2/5	−1/10

m_1 m_2	7/2, −1/2	5/2, −1/2	3/2, −1/2	1/2, −1/2
+1 −3/2	4/35	27/70	2/5	1/10
0 −1/2	18/35	3/35	−1/5	−1/5
−1 +1/2	12/35	−5/14	0	3/10
−2 +3/2	1/35	−6/35	2/5	−2/5

m_1 m_2	7/2, −3/2	5/2, −3/2	3/2, −3/2
0 −3/2	2/7	18/35	1/5
−1 −1/2	4/7	−1/35	−2/5
−2 +1/2	1/7	−16/35	2/5

m_1 m_2	7/2, −5/2	5/2, −5/2
−1 −3/2	4/7	3/7
−2 −1/2	3/7	−4/7

m_1 m_2	7/2, −7/2
−2 −3/2	1

2×2

m_1 m_2	4, +4
+2 +2	1

m_1 m_2	4, +3	3, +3
+2 +1	1/2	1/2
+1 +2	1/2	−1/2

m_1 m_2	4, +2	3, +2	2, +2
+2 0	3/14	1/2	2/7
+1 +1	4/7	0	−3/7
0 +2	3/14	−1/2	2/7

m_1 m_2	4, +1	3, +1	2, +1	1, +1
+2 −1	1/14	3/10	3/7	1/5
+1 0	3/7	1/5	−1/14	−3/10
0 +1	3/7	−1/5	−1/14	3/10
−1 +2	1/14	−3/10	3/7	−1/5

m_1 m_2	4, 0	3, 0	2, 0	1, 0	0, 0
+2 −2	1/70	1/10	2/7	2/5	1/5
+1 −1	8/35	2/5	1/14	−1/10	−1/5
0 0	18/35	0	−2/7	0	1/5
−1 +1	8/35	−2/5	1/14	1/10	−1/5
−2 +2	1/70	−1/10	2/7	−2/5	1/5

m_1 m_2	4, −1	3, −1	2, −1	1, −1
+1 −2	1/14	3/10	3/7	1/5
0 −1	3/7	1/5	−1/14	−3/10
−1 0	3/7	−1/5	−1/14	3/10
−2 +1	1/14	−3/10	3/7	−1/5

m_1 m_2	4, −2	3, −2	2, −2
0 −2	3/14	1/2	2/7
−1 −1	4/7	0	−3/7
−2 0	3/14	−1/2	2/7

m_1 m_2	4, −3	3, −3
−1 −2	1/2	1/2
−2 −1	1/2	−1/2

m_1 m_2	4, −4
−2 −2	1

$$d^{\,1}_{0,0} = \cos\theta \qquad d^{\,1/2}_{1/2,1/2} = \cos\frac{\theta}{2} \qquad d^{\,1/2}_{1/2,-1/2} = -\sin\frac{\theta}{2}$$

$$d^{\,1}_{1,1} = \frac{1+\cos\theta}{2} \qquad d^{\,1}_{1,0} = -\frac{\sin\theta}{\sqrt{2}} \qquad d^{\,1}_{1,-1} = \frac{1-\cos\theta}{2}$$

$$d^{\,3/2}_{3/2,3/2} = \frac{1+\cos\theta}{2}\cos\frac{\theta}{2}$$

$$d^{\,3/2}_{3/2,1/2} = -\sqrt{3}\frac{1+\cos\theta}{2}\sin\frac{\theta}{2}$$

$$d^{\,3/2}_{3/2,-1/2} = \sqrt{3}\frac{1-\cos\theta}{2}\cos\frac{\theta}{2}$$

$$d^{\,3/2}_{3/2,-3/2} = -\frac{1-\cos\theta}{2}\sin\frac{\theta}{2}$$

$$d^{\,3/2}_{1/2,1/2} = \frac{3\cos\theta-1}{2}\cos\frac{\theta}{2}$$

$$d^{\,3/2}_{1/2,-1/2} = -\frac{3\cos\theta+1}{2}\sin\frac{\theta}{2}$$

$$d^{\,2}_{2,2} = \Big(\frac{1+\cos\theta}{2}\Big)^2$$

$$d^{\,2}_{2,1} = -\frac{1+\cos\theta}{2}\sin\theta$$

$$d^{\,2}_{2,0} = \frac{\sqrt{6}}{4}\sin^2\theta$$

$$d^{\,2}_{2,-1} = -\frac{1-\cos\theta}{2}\sin\theta$$

$$d^{\,2}_{2,-2} = \Big(\frac{1-\cos\theta}{2}\Big)^2$$

$$d^{\,2}_{1,1} = \frac{1+\cos\theta}{2}(2\cos\theta-1)$$

$$d^{\,2}_{1,0} = -\sqrt{\frac{3}{2}}\sin\theta\cos\theta$$

$$d^{\,2}_{1,-1} = \frac{1-\cos\theta}{2}(2\cos\theta+1)$$

$$d^{\,2}_{0,0} = \Big(\frac{3}{2}\cos^2\theta - \frac{1}{2}\Big)$$

Figure 35.1: The sign convention is that of Wigner (*Group Theory*, Academic Press, New York, 1959), also used by Condon and Shortley (*The Theory of Atomic Spectra*, Cambridge Univ. Press, New York, 1953), Rose (*Elementary Theory of Angular Momentum*, Wiley, New York, 1957), and Cohen (*Tables of the Clebsch-Gordan Coefficients*, North American Rockwell Science Center, Thousand Oaks, Calif., 1974). The coefficients here have been calculated using computer programs written independently by Cohen and at LBNL.

36. SU(3) ISOSCALAR FACTORS AND REPRESENTATION MATRICES

Written by R.L. Kelly (LBNL).

The most commonly used SU(3) isoscalar factors, corresponding to the singlet, octet, and decuplet content of $8 \otimes 8$ and $10 \otimes 8$, are shown at the right. The notation uses particle names to identify the coefficients, so that the pattern of relative couplings may be seen at a glance. We illustrate the use of the coefficients below. See J.J de Swart, Rev. Mod. Phys. **35**, 916 (1963) for detailed explanations and phase conventions.

A $\sqrt{\ }$ *is to be understood over every integer in the matrices;* the exponent 1/2 on each matrix is a reminder of this. For example, the $\Xi \to \Omega K$ element of the $10 \to 10 \otimes 8$ matrix is $-\sqrt{6}/\sqrt{24} = -1/2$.

Intramultiplet relative decay strengths may be read directly from the matrices. For example, in decuplet → octet + octet decays, the ratio of $\Omega^* \to \Xi\overline{K}$ and $\Delta \to N\pi$ partial widths is, from the $10 \to 8 \times 8$ matrix,

$$\frac{\Gamma\,(\Omega^* \to \Xi\overline{K})}{\Gamma\,(\Delta \to N\pi)} = \frac{12}{6} \times \ (\text{phase space factors})\ . \tag{36.1}$$

Including isospin Clebsch-Gordan coefficients, we obtain, *e.g.*,

$$\frac{\Gamma(\Omega^{*-} \to \Xi^0 K^-)}{\Gamma(\Delta^+ \to p\,\pi^0)} = \frac{1/2}{2/3} \times \frac{12}{6} \times p.s.f. = \frac{3}{2} \times p.s.f. \tag{36.2}$$

Partial widths for $8 \to 8 \otimes 8$ involve a linear superposition of 8_1 (symmetric) and 8_2 (antisymmetric) couplings. For example,

$$\Gamma(\Xi^* \to \Xi\pi) \sim \left(-\sqrt{\frac{9}{20}}\, g_1 + \sqrt{\frac{3}{12}}\, g_2\right)^2 . \tag{36.3}$$

The relations between g_1 and g_2 (with de Swart's normalization) and the standard D and F couplings that appear in the interaction Lagrangian,

$$\mathscr{L} = -\sqrt{2}\, D\, Tr\,(\{\overline{B}, B\}M) + \sqrt{2}\, F\, Tr\,([\overline{B}, B]\, M)\ , \tag{36.4}$$

where $[\overline{B}, B] \equiv \overline{B}B - B\overline{B}$ and $\{\overline{B}, B\} \equiv \overline{B}B + B\overline{B}$, are

$$D = \frac{\sqrt{30}}{40}\, g_1\ , \qquad F = \frac{\sqrt{6}}{24}\, g_2\ . \tag{36.5}$$

Thus, for example,

$$\Gamma(\Xi^* \to \Xi\pi) \sim (F - D)^2 \sim (1 - 2\alpha)^2\ , \tag{36.6}$$

where $\alpha \equiv F/(D + F)$. (This definition of α is de Swart's. The alternative $D/(D + F)$, due to Gell-Mann, is also used.)

The generators of SU(3) transformations, λ_a ($a = 1, 8$), are 3×3 matrices that obey the following commutation and anticommutation relationships:

$$[\lambda_a,\ \lambda_b] \equiv \lambda_a\lambda_b - \lambda_b\lambda_a = 2i f_{abc}\lambda_c \tag{36.7}$$

$$\{\lambda_a,\ \lambda_b\} \equiv \lambda_a\lambda_b + \lambda_b\lambda_a = \frac{4}{3}\delta_{ab}I + 2d_{abc}\lambda_c\ , \tag{36.8}$$

where I is the 3×3 identity matrix, and δ_{ab} is the Kronecker delta symbol. The f_{abc} are odd under the permutation of any pair of indices, while the d_{abc} are even. The nonzero values are

$1 \to 8 \otimes 8$

$$(\Lambda) \to (N\overline{K}\ \ \Sigma\pi\ \ \Lambda\eta\ \ \Xi K) = \frac{1}{\sqrt{8}}\,(2\ \ \ 3\ \ -1\ \ -2)^{1/2}$$

$8_1 \to 8 \otimes 8$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} N\pi & N\eta & \Sigma K & \Lambda K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ N\overline{K} & \Sigma\pi & \Lambda\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \end{pmatrix} = \frac{1}{\sqrt{20}} \begin{pmatrix} 9 & -1 & -9 & -1 \\ -6 & 0 & 4 & 4 & -6 \\ 2 & -12 & -4 & -2 \\ 9 & -1 & -9 & -1 \end{pmatrix}^{1/2}$$

$8_2 \to 8 \otimes 8$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} N\pi & N\eta & \Sigma K & \Lambda K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ N\overline{K} & \Sigma\pi & \Lambda\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \end{pmatrix} = \frac{1}{\sqrt{12}} \begin{pmatrix} 3 & 3 & 3 & -3 \\ 2 & 8 & 0 & 0 & -2 \\ 6 & 0 & 0 & 6 \\ 3 & 3 & 3 & -3 \end{pmatrix}^{1/2}$$

$10 \to 8 \otimes 8$

$$\begin{pmatrix} \Delta \\ \Sigma \\ \Xi \\ \Omega \end{pmatrix} \to \begin{pmatrix} N\pi & \Sigma K \\ N\overline{K} & \Sigma\pi & \Lambda\pi & \Sigma\eta & \Xi K \\ \Sigma\overline{K} & \Lambda\overline{K} & \Xi\pi & \Xi\eta \\ \Xi\overline{K} \end{pmatrix} = \frac{1}{\sqrt{12}} \begin{pmatrix} -6 & 6 \\ -2 & 2 & -3 & 3 & 2 \\ 3 & -3 & 3 & 3 \\ 12 \end{pmatrix}^{1/2}$$

$8 \to 10 \otimes 8$

$$\begin{pmatrix} N \\ \Sigma \\ \Lambda \\ \Xi \end{pmatrix} \to \begin{pmatrix} \Delta\pi & \Sigma K \\ \Delta\overline{K} & \Sigma\pi & \Sigma\eta & \Xi K \\ \Sigma\pi & \Xi K \\ \Sigma\overline{K} & \Xi\pi & \Xi\eta & \Omega K \end{pmatrix} = \frac{1}{\sqrt{15}} \begin{pmatrix} -12 & 3 \\ 8 & -2 & -3 & 2 \\ -9 & 6 \\ 3 & -3 & -3 & 6 \end{pmatrix}^{1/2}$$

$10 \to 10 \otimes 8$

$$\begin{pmatrix} \Delta \\ \Sigma \\ \Xi \\ \Omega \end{pmatrix} \to \begin{pmatrix} \Delta\pi & \Delta\eta & \Sigma K \\ \Delta\overline{K} & \Sigma\pi & \Sigma\eta & \Xi K \\ \Sigma\overline{K} & \Xi\pi & \Xi\eta & \Omega K \\ \Xi\overline{K} & \Omega\eta \end{pmatrix} = \frac{1}{\sqrt{24}} \begin{pmatrix} 15 & 3 & -6 \\ 8 & 8 & 0 & -8 \\ 12 & 3 & -3 & -6 \\ 12 & -12 \end{pmatrix}^{1/2}$$

abc	f_{abc}	abc	d_{abc}	abc	d_{abc}
123	1	118	$1/\sqrt{3}$	355	$1/2$
147	$1/2$	146	$1/2$	366	$-1/2$
156	$-1/2$	157	$1/2$	377	$-1/2$
246	$1/2$	228	$1/\sqrt{3}$	448	$-1/(2\sqrt{3})$
257	$1/2$	247	$-1/2$	558	$-1/(2\sqrt{3})$
345	$1/2$	256	$1/2$	668	$-1/(2\sqrt{3})$
367	$-1/2$	338	$1/\sqrt{3}$	778	$-1/(2\sqrt{3})$
458	$\sqrt{3}/2$	344	$1/2$	888	$-1/\sqrt{3}$
678	$\sqrt{3}/2$				

The λ_a's are

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Equation (36.7) defines the Lie algebra of SU(3). A general d-dimensional representation is given by a set of $d \times d$ matrices satisfying Eq. (36.7) with the f_{abc} given above. Equation (36.8) is specific to the defining 3-dimensional representation.

37. SU(n) MULTIPLETS AND YOUNG DIAGRAMS

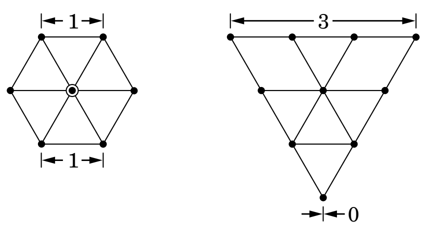
Written by C.G. Wohl (LBNL).

This note tells (1) how SU(n) particle multiplets are identified or labeled, (2) how to find the number of particles in a multiplet from its label, (3) how to draw the Young diagram for a multiplet, and (4) how to use Young diagrams to determine the overall multiplet structure of a composite system, such as a 3-quark or a meson-baryon system.

In much of the literature, the word "representation" is used where we use "multiplet," and "tableau" is used where we use "diagram."

37.1. Multiplet labels

An SU(n) multiplet is uniquely identified by a string of ($n-1$) nonnegative integers: $(\alpha, \beta, \gamma, \ldots)$. Any such set of integers specifies a multiplet. For an SU(2) multiplet such as an isospin multiplet, the single integer α is the number of *steps* from one end of the multiplet to the other (*i.e.*, it is one fewer than the number of particles in the multiplet). In SU(3), the two integers α and β are the numbers of steps across the top and bottom levels of the multiplet diagram. Thus the labels for the SU(3) octet and decuplet



are (1,1) and (3,0). For larger n, the interpretation of the integers in terms of the geometry of the multiplets, which exist in an ($n-1$)-dimensional space, is not so readily apparent.

The label for the SU(n) singlet is $(0, 0, \ldots, 0)$. In a flavor SU(n), the n quarks together form a $(1, 0, \ldots, 0)$ multiplet, and the n antiquarks belong to a $(0, \ldots, 0, 1)$ multiplet. These two multiplets are *conjugate* to one another, which means their labels are related by $(\alpha, \beta, \ldots) \leftrightarrow (\ldots, \beta, \alpha)$.

37.2. Number of particles

The number of particles in a multiplet, $N = N(\alpha, \beta, \ldots)$, is given as follows (note the pattern of the equations).

In SU(2), $N = N(\alpha)$ is

$$N = \frac{(\alpha + 1)}{1} \ . \tag{37.1}$$

In SU(3), $N = N(\alpha, \beta)$ is

$$N = \frac{(\alpha + 1)}{1} \cdot \frac{(\beta + 1)}{1} \cdot \frac{(\alpha + \beta + 2)}{2} \ . \tag{37.2}$$

In SU(4), $N = N(\alpha, \beta, \gamma)$ is

$$N = \frac{(\alpha{+}1)}{1} \cdot \frac{(\beta{+}1)}{1} \cdot \frac{(\gamma{+}1)}{1} \cdot \frac{(\alpha{+}\beta{+}2)}{2} \cdot \frac{(\beta{+}\gamma{+}2)}{2} \cdot \frac{(\alpha{+}\beta{+}\gamma{+}3)}{3} \ . \tag{37.3}$$

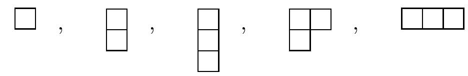
Note that in Eq. (37.3) there is no factor with $(\alpha + \gamma + 2)$: only a *consecutive* sequence of the label integers appears in any factor. One more example should make the pattern clear for any SU(n). In SU(5), $N = N(\alpha, \beta, \gamma, \delta)$ is

$$N = \frac{(\alpha{+}1)}{1} \cdot \frac{(\beta{+}1)}{1} \cdot \frac{(\gamma{+}1)}{1} \cdot \frac{(\delta{+}1)}{1} \cdot \frac{(\alpha{+}\beta{+}2)}{2} \cdot \frac{(\beta{+}\gamma{+}2)}{2} \times \frac{(\gamma{+}\delta{+}2)}{2} \cdot \frac{(\alpha{+}\beta{+}\gamma{+}3)}{3} \cdot \frac{(\beta{+}\gamma{+}\delta{+}3)}{3} \cdot \frac{(\alpha{+}\beta{+}\gamma{+}\delta{+}4)}{4} \tag{37.4}$$

From the symmetry of these equations, it is clear that multiplets that are conjugate to one another have the same number of particles, but so can other multiplets. For example, the SU(4) multiplets (3,0,0) and (1,1,0) each have 20 particles. Try the equations and see.

37.3. Young diagrams

A Young diagram consists of an array of boxes (or some other symbol) arranged in one or more *left-justified* rows, with each row being *at least as long* as the row beneath. The correspondence between a diagram and a multiplet label is: The top row juts out α boxes to the right past the end of the second row, the second row juts out β boxes to the right past the end of the third row, *etc.* A diagram in SU(n) has at most n rows. There can be any number of "completed" columns of n boxes buttressing the left of a diagram; these don't affect the label. Thus in SU(3) the diagrams

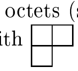
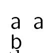


represent the multiplets (1,0), (0,1), (0,0), (1,1), and (3,0). In any SU(n), the quark multiplet is represented by a single box, the antiquark multiplet by a column of ($n-1$) boxes, and a singlet by a completed column of n boxes.

37.4. Coupling multiplets together

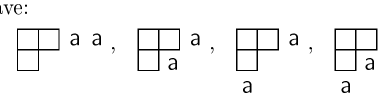
The following recipe tells how to find the multiplets that occur in coupling two multiplets together. To couple together more than two multiplets, first couple two, then couple a third with each of the multiplets obtained from the first two, *etc.*

First a definition: A sequence of the letters $a, b, c, \ldots$ is *admissible* if at any point in the sequence at least as many a's have occurred as b's, at least as many b's have occurred as c's, *etc.* Thus *abcd* and *aabcb* are admissible sequences and *abb* and *acb* are not. Now the recipe:

(a) Draw the Young diagrams for the two multiplets, but in one of the diagrams replace the boxes in the first row with a's, the boxes in the second row with b's, *etc.* Thus, to couple two SU(3) octets (such as the π-meson octet and the baryon octet), we start with  and 

. The *unlettered* diagram forms the *upper left-hand corner* of all the enlarged diagrams constructed below.

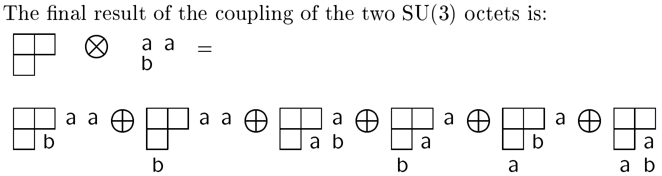
(b) Add the a's from the lettered diagram to the right-hand ends of the rows of the unlettered diagram to form all possible legitimate Young diagrams that have no more than one a per column. In general, there will be several distinct diagrams, and all the a's appear in each diagram. At this stage, for the coupling of the two SU(3) octets, we have:



(c) Use the b's to further enlarge the diagrams already obtained, subject to the same rules. Then throw away any diagram in which the full sequence of letters formed by reading *right to left* in the first row, then the second row, *etc.*, is not admissible.

(d) Proceed as in (c) with the c's (if any), *etc.*

The final result of the coupling of the two SU(3) octets is:



Here only the diagrams with admissible sequences of a's and b's and with fewer than four rows (since $n = 3$) have been kept. In terms of multiplet labels, the above may be written

$$(1,1) \otimes (1,1) = (2,2) \oplus (3,0) \oplus (0,3) \oplus (1,1) \oplus (1,1) \oplus (0,0) \ .$$

In terms of numbers of particles, it may be written

$$\mathbf{8} \otimes \mathbf{8} = \mathbf{27} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \ .$$

The product of the numbers on the left here is equal to the sum on the right, a useful check. (See also Sec. 14 on the Quark Model.)