31. PROBABILITY

31.1. General [8–9]

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

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

(31.1)

From this definition and using the fact that $A \cap B = A \cap \emptyset = \emptyset$, one obtains Bayes' theorem,

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

(31.2)

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

$$ P(B) = \sum_i P(B|A_i)P(A_i) , $$

(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

$$ P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} , $$

(31.4)

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

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 803 and 805 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

$$ P(\text{theory}|\text{data}) = P(\text{data}|\text{theory}) P(\text{theory}) / P(\text{data}) , $$

(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 $P(\text{data|theory})$ is the probability of having 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 $f(x|\theta)$ 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 $\theta$. 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 $\theta$ may have multiple components and are often written as vectors. If $\theta$ 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 $F(x)$ is the probability that $x \leq a$:

$$ F(x) = \int_{-\infty}^{x} f(u) \, du . $$

(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 a step function with values $x_i$ 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

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

(31.7)

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

The $n$th moment of a random variable is

$$ m_n = E[x^n] = \int_{-\infty}^{\infty} x^n f(x) \, dx , $$

(31.8a)

and the $n$th central moment of $x$ (or moment about the mean, $\mu$) is

$$ m_n = E[(x - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) \, dx . $$

(31.8b)

The most commonly used moments are the mean $\mu$ and variance $\sigma^2$:

$$ \mu = E[x] = m_1 = \mu_1 = \mu , $$

$$ \sigma^2 = E[(x - \mu)^2] = m_2 = \mu_2 - \mu^2 . $$

(31.8c)

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 $V=y^2 + \sigma^2 = Z\,\cosh^2 y$. It is often convenient to use the standard deviation of $x$, $\sigma$, 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, $x_{\text{med}}$, defined by $F(x_{\text{med}}) = 1/2$, i.e., half the probability lies above and half lies below $x_{\text{med}}$. (More rigorously, $x_{\text{med}}$ 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

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y) \, dy \quad \text{and similarly for the marginal p.d.f. $f_2(y)$}.$$  

(31.10)

The mean of $x$ is

$$\mu_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \, f(x,y) \, dx \, dy = \int_{-\infty}^{\infty} x \, f_1(x) \, dx \quad \text{and similarly for $y$.}$$

(31.12)

The covariance of $x$ and $y$ is

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

(31.13)

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

$$\rho_{xy} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y},$$

(31.14)

where $\sigma_x$ and $\sigma_y$ 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

$$f(x,y) = f_1(x)f_2(y).$$

(31.15)

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

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

$$g(y) = f(x(y))|J|,$$

(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 $x$ to $y$ is not one-to-one, the $x$-space must be broken in to regions where the function $g(y)$ can be inverted and the contributions to $g(y)$ from each region summed.

Given a set of functions $y = (y_1, \ldots, y_m)$ with $m < n$, one can construct $n - m$ additional independent functions, apply the procedure above, then integrate the resulting $g(y)$ 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 $\Theta$, a change to a different parameter set $\eta(\Theta)$ 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 $e^{iu x}$:

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

(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. (31.18a) and (31.17) that the $n$th moment of a random variable $x$ that follows $f(x)$ is given by

$$\int_{-\infty}^{\infty} x^n f(x) \, dx = \phi^{(n)}(0).$$

(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 $\phi_1(ax)\phi_2(by)$. 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. $f_2(x|y)$ be $\phi_2(u|y)$, and the p.d.f. of $x$ be $f_1(x)$. The characteristic function after integration over the conditional variables is

$$\phi(u) = \int \phi_2(u|y) f_1(x) \, dx.$$  

(31.19)

Suppose we can write $\phi_2$ in the form

$$\phi_2(u|y) = A(u) e^{i\mu_y y}.$$  

(31.20)

Then

$$\phi(u) = A(u) \phi_1(i\mu_y).$$  

(31.21)

The semi-invariants $\kappa_n$ are defined by

$$\phi(u) = \exp \sum_{n=0}^\infty \frac{\kappa_n}{n!} (i u)^n = \exp \left( i \mu_y u - \frac{1}{2} \sigma^2 u^2 + \cdots \right).$$

(31.22)

The values $\kappa_n$ are related to the moments $\mu_n$ and $\sigma^2_n$. The first few relations are

$$\begin{align*}
\kappa_1 & = \mu_1 \quad (\text{the mean}) \\
\kappa_2 & = \sigma_1^2 + \mu_2 \quad (\text{the variance}) \\
\kappa_3 & = 3 \sigma_1^3 + 3 \mu_1 \sigma_1^2 \\
\kappa_4 & = 15 \sigma_1^4 + 6 \mu_1 \sigma_1^3 + 3 \mu_2 \sigma_1^2 \quad (\text{the skewness}) \\
\kappa_5 & = 105 \sigma_1^5 + 105 \mu_1 \sigma_1^4 + 15 \mu_2 \sigma_1^3 + 5 \mu_3 \sigma_1^2 \quad (\text{the excess kurtosis}) \end{align*}.$$  

(31.23)
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. Some probability distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Probability density function $f$ (variable; parameters)</th>
<th>Characteristic function $\phi(u)$</th>
<th>Mean $\mu$</th>
<th>Variance $\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>$f(x; a, b) = \begin{cases} 1/(b-a) &amp; \text{if } a \leq x \leq b \ 0 &amp; \text{otherwise} \end{cases}$</td>
<td>$\phi(u) = \frac{e^{iu(b-a)} - e^{iu}}{i(b-a)}$</td>
<td>$\mu = a + b/2$</td>
<td>$\sigma^2 = (b-a)^2/12$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$f(r; N, p) = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$ for $r = 0, 1, 2, \ldots, N$; $0 \leq p \leq 1$; $q = 1-p$</td>
<td>$(q + p^2v)^N$</td>
<td>$Np$</td>
<td>$Npq$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$f(n; \mu) = \frac{e^{-\mu} \mu^n}{n!}$ for $n = 0, 1, 2, \ldots$ and $\nu &gt; 0$.</td>
<td>$\exp(\nu e^{\mu} - 1)$</td>
<td>$\nu$</td>
<td>$\nu$</td>
</tr>
<tr>
<td>Normal (Gaussian)</td>
<td>$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ for $-\infty &lt; x &lt; \infty$; $-\infty &lt; \mu &lt; \infty$; $\sigma &gt; 0$</td>
<td>$\exp\left(\nu \mu - \frac{\nu^2}{2}\right)$</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Multivariate Gaussian</td>
<td>$f(x; \mu, V) = \frac{1}{(2\pi)^{n/2}</td>
<td>V</td>
<td>^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T V^{-1} (x - \mu)\right)$ for $-\infty &lt; x_j &lt; \infty$; $-\infty &lt; \mu_j &lt; \infty$; $</td>
<td>\det V</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>$f(z; n) = \frac{z^{(n-2)/2} e^{-z/2}}{2^{(n/2)-1} \Gamma(n/2)}$ for $z \geq 0$</td>
<td>$(1-2u)^{-n/2}$</td>
<td>$n$</td>
<td>$2n$</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td>$f(t; n) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$ for $-\infty &lt; t &lt; \infty$; $n$ not required to be integer</td>
<td>$\frac{n+1}{2}$</td>
<td>$0$ for $n \geq 2$</td>
<td>$n(n-2)$ for $n \geq 3$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$f(x; \lambda, k) = \frac{e^{-x} x^{k-1}}{\Gamma(k)}$ for $0 &lt; x &lt; \infty$; $k$ not required to be integer</td>
<td>$(1-iu/\lambda)^{-k}$</td>
<td>$k/\lambda$</td>
<td></td>
</tr>
</tbody>
</table>

### 31.4.2. Poisson distribution

The Poisson distribution $f(n; \nu)$ gives the probability of finding exactly $n$ events in a given interval (e.g., space and time) when the events occur independently of one another and of $x$ at an average rate of $\nu$ per the given interval. The variance $\sigma^2$ equals $\nu$. 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 $\nu$.

### 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 $\mu$ and $\sigma$ and variance $V = \sigma^2$. Comparison of the characteristic function $\phi(u)$ given in Table 31.1 with Eq. (31.22) shows that all semi-invariants $\text{c}_n$ beyond $\text{c}_2$ vanish; this is a unique property of the Gaussian distribution. Some other properties are:

- $P(x \text{ in range } \mu \pm \sigma) = 0.6827$;
- $P(x \text{ in range } \mu \pm 2\sigma) = 0.9545$;
- $E[|x - \mu|] = \sqrt{2/\pi} \sigma = 0.39\lambda \sigma$;
- half-width at half maximum $= \sqrt{2\ln 2} \sigma = 1.17\sigma$.

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

$$F(x; 0, 1) = \frac{1}{2} \left[ 1 + \text{erf}(x/\sqrt{2}) \right].$$  \hspace{1cm} (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 $\mu$ and variance $\sigma^2$, 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 pdf $f$ with finite means and variances, then the sum $y = \sum x_i$ will have a pdf 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], [9] for details.)

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

$$\phi(u|\mu,V) = \exp \left[ i \mu \cdot u - \frac{1}{2} u^T V u \right].$$  \hspace{1cm} (31.25)

From Eq. (31.18), the covariance of $x_1$ and $x_2$ is

$$E[(x_1 - \mu_1)(x_2 - \mu_2)] = V_{12}.$$  \hspace{1cm} (31.26)

If the components of $x$ 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.

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

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

For $n = 2$, $f(x|\mu,V)$ is

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

The marginal distribution of any $x_i$ is a Gaussian with mean $\mu_i$ and variance $V_{ii}$. $V$ is $n \times n$, symmetric, and positive definite. Therefore for any vector $X$, the quadratic form $X^T V^{-1} X = C$, where $C$ is any positive number, traces an $n$-dimensional ellipsoid as $X$ varies. If $X_1 = x_1 - \mu_1$, then $C$ is a random variable obeying the $\chi^2$ distribution with $n$ degrees of freedom, discussed in the following section. The probability that $X$ 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 $F_{\chi^2}(C;n)$ is the cumulative $\chi^2$ distribution. This may be read from Fig. 32.1. For example, the “out-of-standard-deviation ellipsoid” occurs at $C = n$. For the two-variable case ($n = 2$), the point $X$ lies outside the one-standard-deviation ellipsoid with $63\%$ probability. The use of these ellipsoids as indicators of probability error is discussed in Sec. 32.3.2; the validity of these indicators assumes that $V$ and $x$ are correct.

31.4.4. $\chi^2$ 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 $\chi^2$ 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 $z = X^T V^{-1} X$ as in the previous section. For a set of $z_i$ each of which follows $\chi^2(n_i)$, $z = \sum z_i$ follows $\chi^2(\sum n_i)$. For large $n_i$, the $\chi^2$ p.d.f. approaches a Gaussian with mean $n$ and variance $n$. The $\chi^2$ 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 discussion is found further in Sec. 322.2 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

$$z = \frac{\sum_{i=1}^n x_i^2}{n} \text{ and } t = \frac{z}{\sqrt{z/n}}.$$  \hspace{1cm} (31.29)

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

$$t = \frac{(\bar{x} - \mu)\sqrt{n-1}/s}{\sqrt{(n-1)/n}} = \frac{\bar{x} - \mu}{s}$$  \hspace{1cm} (31.30)

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

In Table 31.1 $n$ in $f(t;n)$ is not required to be an integer. A Student’s $t$ distribution with non-integer $n > 0$ 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 is $\lambda$ per unit $x$. The special case $k = 1$ (i.e., $f(x;\lambda,1) = \exp(-\lambda x)$ is called the exponential distribution. A sum of $\lambda$ exponential random variables $x_i$ is distributed as $f(x;\lambda,k)$.

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:

Revised April 1988 by F. James (CERN); February 2000 by B. Cousins (UCLA); October 2001 and October 2003 by G. Cowan (Rutherford Inst.).

This chapter gives an overview of statistical methods used in High Energy Physics. In statistics we are interested in using a given set 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, as representing a 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. (31.14), 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, by a physicist, or the more detailed comparison in Ref. 36.

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 “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 is a function of the data whose value, the estimate, is intended as a meaningful guess for the value of the parameter. 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 converges to the true value as the amount of data increases. This property is so important that it is possessed by all commonly used estimators.

(b) The bias, $b = E(\hat{\theta}) - \theta$, 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 $\theta$ 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 $\hat{\theta}$ is an unbiased estimator of $\theta$, then $\hat{\theta}^2$ is not in general an unbiased estimator for $\theta^2$. If we have an estimate $b$ for the bias we can subtract it from $\theta$ to obtain a new $\theta' = \theta - b$. The estimate $b$, however, is 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 $V[\hat{\theta}]$ to its minimum possible value. Under rather general conditions, the minimum variance is given by the Hsu-Cramer-Rao bound,

$$\sigma^2_{\text{min}} = \left(1 + \frac{\partial b}{\partial \theta}ight)^2 / I(\theta),$$

where

$$I(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \sum f(x; \theta) \right)^2 \right]$$

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 $\theta$.

The mean-squared error,

$$\text{MSE} = E(\hat{\theta} - \theta)^2 = V[\hat{\theta}] + b^2,$$

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

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

#### 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 $\mu$ with a common, but unknown, variance $\sigma^2$. Then

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

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

$$\frac{1}{N} \left( m_4 - \frac{N-3}{N-1} \right),$$

where $m_4$ is the 4th central moment of $x$. For Gaussian distributed $x$ this becomes $2\sigma^4 / (N-1)$ for any $N \geq 2$, and for large $N$ the standard deviation of $\sigma$ (the "error of the error") is $\sigma / \sqrt{2N}$. Again if the $x_i$ are Gaussian, $\mu$ is an efficient estimator for $\mu$ and the estimators $\mu$ and $\sigma^2$ 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 $\sigma^2$ is known, it does not improve the estimate $\bar{x}$, as can be seen from Eq. (32.4); however, if $\mu$ is known, substitute it for $\bar{x}$ in Eq. (32.5) and replace $N - 1$ by $N$ to obtain a somewhat better estimator of $\sigma^2$. 

32. Statistics
If the \( x_i \) have different, known variances \( \sigma_i^2 \), then the weighted average

\[
\bar{\mu} = \frac{1}{w} \sum_{i=1}^{N} w_i x_i
\]

is an unbiased estimator for \( \mu \) 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 \( \bar{\mu} \) is \( 1/\sqrt{w} \).

As an estimator for the median \( x_{\text{med}} \) one can use the value \( \bar{x}_{\text{med}} \) such that half the \( x_i \) are greater and half are smaller (the sample median).

If the sample median lies between two observed values, it is set by (8.1) and since both are maximized at different \( \theta_i \), known variances \( \sigma_i^2 \) such that half the \( x_i \) have deviated of \( \sigma_i^2 \).

In the large sample limit (or in a linear model with

\[
\text{deviation of } \bar{\mu} \text{ is } \frac{1}{\sqrt{w}}.
\]

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 \( \theta \) such that

\[
\ln L(\theta') = \ln L_{\text{max}} - s^2/2,
\]

where \( L_{\text{max}} \) is the value of \( L \) at the solution point (compare with Eq. (32.46)). The extreme limits of this contour on the \( \theta_i \) axis give an approximate \( s \)-standard-deviation confidence interval for \( \theta_i \) (see Section 32.5.2.3).

In the case where the size \( n \) of the data sample \( x_1, \ldots, x_n \) is small, the unbiased 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., [17, 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 \( n_i = (n_1, \ldots, n_n) \) with expectation values \( \mu = E[n] \) and probabilities \( f_i = E[n_i] \). 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 \( \lambda(\theta) = f_i(\mu(\theta))/f_i(n_i) \) or to minimizing the quantity \( H \)

\[
-2 \ln \lambda(\theta) = 2 \sum_{i=1}^{N} n_i \ln \frac{n_i}{f_i(\mu(\theta))} + 2 \sum_{i=1}^{N} n_i \ln f_i(\mu(i))
\]

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 unbiased 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 \( \chi^2 \) distribution in the large sample limit. If there are \( N \) bins and \( m \) fitted parameters, then the number of degrees of freedom for the \( \chi^2 \) distribution is \( N - m - 1 \) if the data are treated as multinomially distributed and \( N - m \) if the \( n_i \) are Poisson variables with \( N \mu_i = \sum_i n_i \) fixed. If the \( n_i \) are Poisson distributed and \( n_{\text{corr}} \) 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., \( \sum_i n_i = \sum_i n_{\text{corr}} \). This is not the case for parameter estimation methods based on a least-squares procedure with traditional weights (see, e.g., Ref. [22]).

### 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. \( f(x; \theta) \), where \( \theta = (\theta_1, \ldots, \theta_n) \) is set of \( n \) parameters whose values are unknown. The method of maximum likelihood takes the estimator \( \theta \) to be those values of \( \theta \) that maximize the likelihood function,

\[
L(\theta) = \prod_{i=1}^{N} f(x_i; \theta).
\]

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; in frequentist statistics this is not defined. In Bayesian statistics one can obtain from the likelihood the posterior p.d.f. for \( \theta \), 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 \( \theta \), the maximum likelihood (ML) estimators can be found by solving the likelihood equations,

\[
\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, \ldots, n.
\]

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 \( \theta \) 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 \( \theta \).

Under a one-to-one change of parameters from \( \theta \) to \( \eta \), the ML estimators \( \theta \) transform to \( \eta(\theta) \). 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 \( V_{ij} = \text{cov}(\theta_i, \theta_j) \) for a set of ML estimators can be estimated by using

\[
(V^{-1})_{ij} = -\sum_{k=1}^{N} \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}.
\]

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)parsbolic. 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 \( \theta \) such that

\[
\ln L(\theta') = \ln L_{\text{max}} - s^2/2,
\]

where \( L_{\text{max}} \) is the value of \( L \) at the solution point (compare with Eq. (32.46)). The extreme limits of this contour on the \( \theta_i \) axis give an approximate \( s \)-standard-deviation confidence interval for \( \theta_i \) (see Section 32.5.2.3).
In many practical cases one further restricts the problem to the situation where $F(x_i; \theta)$ is a linear function of the parameters, i.e.,

$$F(x_i; \theta) = \sum_{j=1}^{m} \theta_j h_j(x_i).$$  \hfill (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 $x_j$ must be distinct.

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

$$\bar{\theta} = (H^TV^{-1}H)^{-1}H^TV^{-1}y = Dy.$$  \hfill (32.16)

The covariance matrix for the estimators $\bar{U}_j = \text{cov} \hat{\theta}_j$ is given by

$$U = DV^TD^{-1} = (H^TV^{-1}H)^{-1},$$  \hfill (32.17)

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

$$U^{-1} = \frac{1}{2} \left[ \frac{\partial^2\chi^2}{\partial \theta_i \partial \theta_j} \right]_{\bar{\theta} = \bar{y}} = -\sum_{k=1}^{N} h_k(x_k)(U^{-1})_{kl} h_l(x_l).$$  \hfill (32.18)

The LS estimators can also be found from the expression

$$\bar{\theta} = U \hat{y},$$  \hfill (32.19)

where the vector $\hat{y}$ is defined by

$$\hat{y} = \sum_{k=1}^{N} y_k h_k(x_k)(U^{-1})_{ik}.$$  \hfill (32.20)

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

$$U^{-1}_{ij} = \sum_{k=1}^{N} \frac{\bar{h}_i(x_k) \bar{h}_j(x_k)}{\sigma_k^2},$$  \hfill (32.21)

$$\hat{y} = \sum_{k=1}^{N} \frac{y_k \bar{h}_k(x_k)}{\sigma_k}.$$  \hfill (32.22)

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

$$\chi^2(\theta) = \chi^2(\bar{\theta}) + 1 = \chi^2_{\text{min}} + 1$$  \hfill (32.23)

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

32.2.1. Hypothesis tests

Consider an experiment whose outcome is characterized by a vector of data $x$. A hypothesis is a statement about the distribution of $x$. 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 $x$ a given hypothesis (often called the null hypothesis, $H_0$) should be rejected. This is done by defining a region of a-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 $x$ is observed in the acceptance region, i.e., the complement of the critical region. Here ‘accept’ 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, $\alpha$, which is often chosen to be equal to some proscribed 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, $\beta$, 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 $x$ 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 $x$ 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 $x$. Each hypothesis for the distribution of $x$ will determine a distribution for $t$, and the acceptance region in $t$-space will correspond to a specific range of values of $t$.  

32.1.4. Propagation of errors

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

$$U_{ij} \approx \sum_{kl} \frac{\partial \eta_i}{\partial \theta_k} \frac{\partial \eta_j}{\partial \theta_l} V_{kl}.$$  \hfill (32.24)

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

$$A_{ij} = \frac{\partial \eta_i}{\partial \theta_j}.$$

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

32.2. Statistical tests
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 $x$ in that region, the ratio of p.d.f.s for the hypotheses $H_0$ and $H_1$,

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

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 (32.20) 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 the joint p.d.f.s $f(x|H_0)$ and $f(x|H_1)$. Instead, test statistics based on neural network 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 is question, $H_0$, will determine the p.d.f. $g(H_0)$ 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

$$
p = \int_{-\infty}^{\infty} g(H_0) dt,
$$

where $t_{obs}$ 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 proscribed 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 $H_0$, 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 $\chi^2$ (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 $n = (n_1, \ldots, n_m)$, with a hypothesis for their expectation values $\nu_1, \ldots, \nu_n$. As the distribution is Poisson with variances $\nu_j = \nu$, the $\chi^2$ (32.13) becomes Pearson's $\chi^2$ statistic,

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

If the hypothesis $\nu = (\nu_1, \ldots, \nu_N)$ 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 $\chi^2$ statistic will follow the $\chi^2$ 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 $\chi^2$ from Eq. (32.13) if the $y_i$ are Gaussian.
32.3. Confidence intervals and limits

When the goal of an experiment is to determine a parameter \( \theta \), 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 \( \hat{\theta} \) plus or minus an estimate of the standard deviation of \( \theta \), \( \sigma \). 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 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 above.

32.3.1. The Bayesian approach

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

\[
p(\theta|\mathbf{x}) = \frac{L(\mathbf{x} | \theta) \pi(\theta)}{\int L(\mathbf{x} | \theta) \pi(\theta) d\theta} \tag{32.30}
\]

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

Bayesian statistics supplies no fundamental rule for determining \( \pi(\theta) \); this reflects the experimenter's subjective degree of belief about \( \theta \) 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) \( \theta_a, \theta_b \) can be determined which contains a given function \( 1 - \alpha \) of the probability, i.e.,

\[
1 - \alpha = \int_{\theta_a}^{\theta_b} p(\theta | \mathbf{x}) d\theta . \tag{32.31}
\]

Sometimes an upper or lower limit is desired, i.e., \( \theta_a \) can be set to zero or \( \theta_b \) to infinity. In other cases one might choose \( \theta_a \) and \( \theta_b \) such that \( p(\theta | \mathbf{x}) \) 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 non-linear transformation of the parameter.

The main difficulty with Bayesian intervals is in quantifying the prior beliefs. Sometimes one attempts to construct \( \pi(\theta) \) 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 \( \theta \), then it is not flat for a nonlinear function of \( \theta \) 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 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 \( \mu \) as well as background with mean \( b \), assumed known. For the signal mean \( \mu \) one often uses the prior

\[
\pi(s) = \begin{cases} 
0 & s < 0 \\
1 & s \geq 0 
\end{cases} \tag{32.32}
\]

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 \( n \). Using the likelihood function for Poisson distributed \( n \),

\[
L(n| \mu) = \frac{(s+b)^n e^{-s}}{n!} , \tag{32.33}
\]

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

\[
1 - \alpha = \int_{s_{\alpha}}^{\infty} p(n | \mu) ds = \int_{s_{\alpha}}^{\infty} L(n | \mu) \pi(s) ds = \int_{s_{\alpha}}^{\infty} \frac{L(n | \mu) \pi(s)}{\int L(n | \mu) \pi(s) ds} ds , \tag{32.34}
\]

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

\[
\alpha = e^{-s_{\alpha}} \sum_{m=0}^{\infty} \frac{(s+b)^m}{m!} \left( \frac{1}{\sum_{m=0}^{\infty} \frac{(s+b)^m}{m!}} \right) . \tag{32.35}
\]

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

\[
s_{\alpha} = Q_{1-\alpha}^{\chi^2} / \sqrt{2} \left( 1 - 0.11 \mu \right) , \tag{32.36}
\]

where the number of degrees of freedom is \( \mu = 2(n+1) \). The quantile of the \( \chi^2 \) distribution can be obtained using the CERNLIB routine CERCIJ. 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 32.32.4. Values for \( 1 - \alpha = 0.9 \) and 0.05 are given by the values \( s_{\alpha} \) 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 \( \theta \) but on nuisance parameters \( \nu \), whose values are known with some limited accuracy. For a single nuisance parameter \( \nu \), for example, one might have a p.d.f. centered about its nominal value with a certain standard deviation \( \sigma_\nu \). 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 \( \theta \) and \( \nu \) and are related by Bayes’ theorem as usual. One can obtain the posterior p.d.f. for \( \theta \) alone by integrating over the nuisance parameter, i.e.,

\[
y(\theta|x) = \int p(\theta, \nu|x) d\nu.
\]

(32.37)

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

\[
L^*(x|\theta) = \int L(x|\theta, \nu) p(\nu) d\nu.
\]

(32.38)

The function \( L^*(x|\theta) \) 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 have 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 such 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 \( \theta \) is the unknown parameter for which we want to construct a confidence interval. The variable \( x \) could (and often does) represent an estimator for \( \theta \). Using \( f(x; \theta) \) we can find for a pre-specified probability \( 1 - \alpha \) and for every value of \( \theta \) a set of values \( x_1(\theta, \alpha) \) and \( x_2(\theta, \alpha) \) such that

\[
P(x_1 < x < x_2; \theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x; \theta) dx.
\]

(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 \( \theta \). The union of such intervals for all values of \( \theta \), 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 \( \theta \), which we assume for this discussion.

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 \( \theta \) is the set of all values of \( \theta \) 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 \( 1 - \alpha \).

#### 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 \( \theta \). One then excludes all values of \( \theta \) where the hypothesis would be rejected at a significance level less than \( \alpha \). 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 to use a test statistic based on the likelihood ratio,

\[
\lambda = \frac{f(x; \theta)}{f(x; \emptyset)}
\]

(32.41)

where \( \emptyset \) is the value of the parameter which, out of all allowed values, maximizes \( f(x; \emptyset) \). 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 \( \lambda \).
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 $\text{CL}_{\text{u}}$ is determined from the ratio

$$\text{CL}_{\text{u}} = \frac{p\text{-value of signal plus background hypothesis}}{1 - p\text{-value of background only}}.$$  

(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 $\text{CL}_{\text{u}}$ 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 $\text{CL}_{\text{u}}$ less than the one actually observed is less than $\alpha$. The LEP Higgs group has in fact followed a more conservative approach and excluded the signal at a confidence level $1 - \alpha$ if $\text{CL}_{\text{u}}$ itself (not the probability to obtain a lower $\text{CL}_{\text{u}}$ value) is less than $\alpha$. 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 $\sigma$,

$$1 - \alpha = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\delta}^{\delta} e^{-z^2/2\sigma^2} dz = \text{erf} \left( \frac{\delta}{\sqrt{2}} \right)$$  

(32.43)

is the probability that the measured value $x$ will fall within $\pm \delta$ of the true value $\mu$. From the symmetry of the Gaussian with respect to $x$ and $\mu$, this is also the probability for the interval $x \pm \delta$ to include $\mu$. Fig. 32.4 shows a $\delta = 1.645$ confidence interval unshaded. The choice $\delta = \sigma$ gives an interval called the standard error which has $1 - \alpha = 68.27\%$ if $\sigma$ is known. Values of $\alpha$ for other frequently used choices of $\delta$ are given in Table 32.1.

![Figure 32.4](image)

**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 $\alpha$, are as shown.

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

In addition to Eq. (32.43), $\alpha$ and $\delta$ are also related by the cumulative distribution function for the $\chi^2$ distribution,

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

(32.44)

for $\chi^2 = (x - \mu)^2/\sigma^2$ 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 $\theta = (\theta_1, \ldots, \theta_n)$, one requires the full covariance matrix $V_{ij} = \text{cov}(|\theta_1, \theta_j|)$, 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 $\theta$, and furthermore the likelihood function itself takes on a Gaussian shape.

The standard error ellipse for the pair $((\theta_1, \theta_2))$ is shown in Fig. 32.5, corresponding to a contour $\chi^2 = \chi^2_{\text{min}} + 1$ or $L = L_{\text{max}} - 1/2$. The ellipse is centered about the estimated values $\hat{\theta}$ and the tangents to the ellipse give the standard deviations of the estimates, $\sigma_1$ and $\sigma_2$. The angle of the major axis of the ellipse is given by

$$\tan 2\phi = \frac{2 \rho_{12} \sigma_1 \sigma_2}{\sigma_1^2 - \sigma_2^2},$$  

(32.45)

where $\rho_{12} = \text{cov}(|\theta_1, \theta_2|)/\sigma_1 \sigma_2$ is the correlation coefficient.

The correlation coefficient can be visualized as the fraction of the distance $\sigma_2$ from the ellipse’s horizontal centerline at which the ellipse becomes tangent to vertical, i.e., at the distance $\rho_{12} \sigma_2$ below the centerline as shown. As $\rho_{12}$ goes to +1 or -1, the ellipse thickens to a diagonal line.

It could happen that one of the parameters, say, $\theta_2$, is known from previous measurements to a precision much better than $\sigma_2$ so that the current measurement contributes almost nothing to the knowledge of $\theta_2$. However, the current measurement of $\theta_1$ and its dependence on $\theta_2$ may still be important. In this case, instead of quoting both parameter estimates and their correlation, one sometimes reports the value of $\theta_2$ which minimizes $\chi^2$ at a fixed value of $\theta_1$ such as the FPD best value. This $\theta_1$ value lies along the dotted line between the points where the ellipse becomes tangent to vertical, and has statistical error $\sigma_{\text{error}}$ as shown on the figure, where $\sigma_{\text{error}} = (1 - \rho_{12}^2)^{1/2} \sigma_2$.

Instead of the correlation $\rho_{12}$, one reports the dependency $\partial \theta_2/\partial \theta_1$ which is the slope of the dotted line. This slope is related to the correlation coefficient by $\partial \theta_2/\partial \theta_1 = \rho_{12} \sigma_2/\sigma_1$.

![Figure 32.5](image)

**Figure 32.5**: Standard error ellipse for the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$. In this case the correlation is negative.

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

$$\ln L(\theta, \hat{\theta}) \geq \ln L_{\text{max}} - \ln 1 - \alpha,$$

(32.46)
or where a $\chi^2$ has been defined for use with the method of least squares,

$$\chi^2(\theta) \leq \chi^2_{\alpha} + \Delta \chi^2.$$  (32.47)

Values of $\Delta \chi^2$ or $2.4 \ln L$ 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 $\theta$ will depend on $\theta$, so these are not exact confidence regions according to our previous definition. Nevertheless, they 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 $\chi^2$ 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 $L_i(\theta)$, then the combined log-likelihood function is simply the sum,

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

### 32.3.4. 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 $\sigma$, say, in a fixed integrated luminosity $L$, then it follows a Poisson distribution with mean $\nu = \sigma L$. 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 $\nu$ gives the probability for the event to possess the property in question. This is approximate, $\nu = n/N$, 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 $\nu$ can be found from the Neyman procedure to be

$$\nu_{\text{up}} = \frac{n F_{\nu}^{-1}(\alpha; 2n)}{2}.$$  (32.49a)$$\nu_{\text{low}} = \frac{n F_{\nu}^{-1}(1 - \alpha; 2(n + 1))}{2}.$$  (32.49b)

where the upper and lower limits are $1 - \alpha_{\text{up}}$, and $1 - \alpha_{\text{low}}$, respectively, and $F_{\nu}^{-1}$ is the quantile of the $\chi^2$ distribution (inverse of the cumulative distribution). The quantities $F_{\nu}^{-1}$ can be obtained from standard tables or from the CERNLIB routine CFNS. For central confidence intervals at confidence level $1 - \alpha$, set $\alpha_{\text{up}} = \alpha_{\text{low}} = \alpha/2$.

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 $\nu$. 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 $\nu$ of a Poisson variable given $n$ observed events in the absence of background, for confidence levels of 90% and 95%.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$\nu_{\text{up}}$</th>
<th>$\nu_{\text{low}}$</th>
<th>$\nu_{\text{up}}$</th>
<th>$\nu_{\text{low}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>2.30</td>
<td>3.90</td>
<td>4.74</td>
<td>6.30</td>
</tr>
<tr>
<td>0.95</td>
<td>2.05</td>
<td>3.68</td>
<td>5.18</td>
<td>7.25</td>
</tr>
<tr>
<td>0.99</td>
<td>1.74</td>
<td>2.81</td>
<td>10.51</td>
<td>13.44</td>
</tr>
<tr>
<td>0.995</td>
<td>1.49</td>
<td>1.73</td>
<td>16.91</td>
<td>20.44</td>
</tr>
</tbody>
</table>

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

$$p_{\text{up}} = \frac{n F_p^{-1}(\alpha; 2n, 2(N - n + 1))}{2(N - n)} - p,$$  (32.50a)$$p_{\text{low}} = \frac{n F_p^{-1}(1 - \alpha; 2(n + 1), 2N - 2n)}{2(N - n)}.$$  (32.50b)

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

#### 32.3.5. 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 $n^2 = E^2 - b^2$, where $E$ and $b$ are independent, Gaussian distributed estimates of the energy and momentum. Although the true $m^2$ is constrained to be positive, random errors in $E$ and $b$ can easily lead to negative values for the estimate $n^2$.

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 $E \nu = \nu - n$ 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 $s = n - b$ is an unbiased estimator for $s$. Depending on true magnitudes of $s$ and $b$, the estimate $s$ 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 a may be of zero length.

The confidence interval is in fact designed not to cover the parameter with a probability of at most $\alpha$, and if a zero-length interval results, then this is evidently one of those experiments.

### Table 32.2: $\Delta \chi^2$ or $2.4 \ln L$ corresponding to a coverage probability $1 - \alpha$ in the large data sample limit, for joint estimation of $m$ parameters.

<table>
<thead>
<tr>
<th>$(1 - \alpha)$ (%)</th>
<th>$m = 1$</th>
<th>$m = 2$</th>
<th>$m = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>68.27</td>
<td>1.00</td>
<td>2.30</td>
<td>3.53</td>
</tr>
<tr>
<td>90.</td>
<td>2.71</td>
<td>4.61</td>
<td>6.25</td>
</tr>
<tr>
<td>95.</td>
<td>3.84</td>
<td>5.99</td>
<td>7.82</td>
</tr>
<tr>
<td>99.45</td>
<td>4.00</td>
<td>6.18</td>
<td>8.03</td>
</tr>
<tr>
<td>99.73</td>
<td>6.63</td>
<td>9.21</td>
<td>11.34</td>
</tr>
<tr>
<td>99.93</td>
<td>9.00</td>
<td>11.83</td>
<td>14.16</td>
</tr>
</tbody>
</table>
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. (32.20). 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 ("slip-sliding"), 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 a Poisson variable given an observed number of events is the inaccuracy of the estimate for 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))

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 alternative is presented by the intervals found from the likelihood function or $x^2$ 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, $\theta$ will not be uniform if transformed to $\theta^2$. 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,

$$p(\theta | \text{result}) \propto L(\text{result}) p(\theta)$$

where the reader supplies his or her own prior beliefs $p(\theta)$ 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 so 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 $L = L_{\text{max}} - \Delta L$ 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

| Table 32.4: Unified confidence intervals $[v_1, v_2]$ for the mean of a Poisson variable given $n$ observed events in the absence of background, for confidence levels of $90\%$ and $95\%$. |
|---|---|---|---|
| $n$ | $1 - \alpha = 90\%$ | $1 - \alpha = 95\%$ |
| $v_1$ | $v_2$ | $v_1$ | $v_2$ |
| 0 | 0.00 | 4.44 | 0.00 | 3.09 |
| 1 | 0.11 | 4.30 | 0.05 | 5.14 |
| 2 | 0.53 | 5.91 | 0.36 | 6.72 |
| 3 | 1.10 | 7.46 | 0.82 | 8.25 |
| 4 | 1.87 | 8.88 | 1.37 | 9.86 |
| 5 | 2.81 | 9.99 | 1.84 | 11.26 |
| 6 | 3.86 | 10.97 | 2.21 | 12.75 |
| 7 | 4.95 | 11.83 | 2.58 | 13.81 |
| 8 | 5.96 | 12.30 | 2.94 | 15.29 |
| 9 | 6.97 | 12.50 | 3.45 | 16.77 |
| 10 | 7.98 | 12.50 | 3.96 | 17.82 |

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 central 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 0.05 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
it would be particularly misleading to draw conclusions about the parameter's value without further information from the likelihood.

References:
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. 236 of Ref. 3 gives the rule: all such points are included in the acceptance region (the domain \( D(x) \)). Some authors have assumed the contrary, and shown that one can then obtain null intervals.
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. Miguel 
(LBNL)

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 
assemblage 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 
L"uscher, 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 $10^{18}$. 
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. $f(x)$ for 
a given $u$ if we set

$$u = F(x),$$

(33.1) provided we can find an inverse of $F$, defined by

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

(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 
$\exp(x)$, $\sin(x)$, and $1/(1 + x^2)$ (Cauchy or Breit-Wigner), although it 
does not normally 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 $x_k$. Choose $u$ from a uniform 
distribution on $(0, 1)$ as before. Find $x_k$ such that

$$F(x_{k-1}) < u \leq F(x_{k}) \equiv \text{Prob}(x \leq x_{k}) = \sum_{j=1}^{k} f(x_j);$$

(33.3)

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

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.

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 $u C h(x) \leq f(x)$. If no, accept $x$; if not reject $x$ and try again. If 
we regard $x$ and $u C h(x)$ 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

For algorithms generating random numbers belonging to many different distributions are given by Press [1], 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, also using the fact that \( (1 - u) \) is uniform if \( u \) is uniform. To generate decays between times \( t_1 \) and \( t_2 \) according to \( f(t) = \exp(-t/\tau) \); let \( r_2 = \exp(-t_2/\tau) \) and \( r_1 = \exp(-t_1/\tau) \); generate \( u \) and let

\[
t = -\ln(r_2 + u(r_1 - r_2)).
\]

For \((t_1, t_2) = (0, \infty)\), we have simply \( t = -\ln u \). (See also Sec. 33.4.6.)

33.4.2. Isotropic direction in 3D

Isometry means the density is proportional to solid angle, the differential element of which is \( dD = d\cos\theta d\phi \). Hence \( \cos\theta \) is uniform (\( 2u - 1 \)) and \( \phi \) is uniform (\( 2\arctan u \)). For alternative generation of \( \sin\theta \) and \( \phi \), 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 \( \cos^2 \theta = v_1^2 + v_2^2 \). If \( \cos^2 \theta > 1 \), start over. Otherwise, the sine (\( \sin \theta \)) and cosine (\( \cos \)) of a random angle are given by

\[
\begin{align*}
S &= 2u_2 v_1 \cos^2 \theta \quad \text{and} \\
C &= (v_1^2 - v_2^2) \cos^2 \theta.
\end{align*}
\]

33.4.4. Gaussian distribution

If \( \mu \) and \( \sigma^2 \) are uniform on \((0, 1)\), then

\[
z_1 = \sin 2u_1 \sqrt{-2 \ln v_2} \quad \text{and} \quad z_2 = \cos 2u_1 \sqrt{-2 \ln v_2}
\]

are independent and Gaussian distributed with mean 0 and \( \sigma^2 \). There are many faster variants of this basic algorithm. For example, construct \( v_1 = 2u_1 - 1 \) and \( v_2 = 2 - 2u_1 \), which are uniform on \((-1, 1)\). Calculate \( \cos^2 \theta = v_1^2 + v_2^2 \), and if \( \cos^2 \theta > 1 \) start over. Otherwise, \( z_1 = v_1 \sqrt{-2 \ln v_2} \) and \( z_2 = v_2 \sqrt{-2 \ln v_2} \). (33.7)

33.4.5. \( \chi^2(n) \) distribution

For \( n \) even, generate \( n/2 \) uniform numbers \( u_i \); then

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

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

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

For \( n \geq 30 \), the much faster Gaussian approximation for the \( \chi^2 \) may be preferable; generate \( z \) as in Sec. 33.4.4 and use

\[
y = z^2 + \ln(2n-1)/2 \quad \text{if} \quad z < -\sqrt{2n-1} \quad \text{reject and start over.}
\]

33.4.6. Gamma distribution

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

- If \( k \geq 1 \) (the exponential distribution), accept \( x = -\ln u \). (See also Sec. 33.4.1.)

- If \( 0 < k < 1 \), initialize with \( v_1 = -(e+1)/c \) (with \( c = 2.71828 \ldots \) being the natural log base). Generate \( u_1, u_2, \) Define \( v_2 = v_1 u_2; \)

  - Case 1: \( v_2 \leq 1 \). Define \( x = -\ln v_2 \). Accept \( x \) and stop, else restart by generating new \( u_1, u_2, v_2. \)

  - Case 2: \( v_2 > 1 \). Define \( x = -\ln(v_1 + v_2/k) \). If \( u_2 \leq k \cdot v_1 \), accept \( x \) and stop, else restart by generating new \( u_1, u_2. \)

Note that, for \( k < 1 \), the probability density has a peak 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 \( c = 36 - 0.75 \cdot k \). Generate \( u_1 \) and compute \( v_1 = u_1(1 - u_2) \) and \( v_2 = (u_2 + 0.5) \sqrt{\ln v_1} \).

  - If \( x + k + v_1 - 1 \leq 0 \), go back and generate new \( u_1, u_2; \) otherwise generate \( u_1, u_2, \) and compute \( v_1 = 64 v_1 v_2^2 \).

  - If \( v_2 \leq 2 \cdot \ln(1/k + 1) \), accept \( x \) and stop, else go back and generate new \( u_1, u_2. \)

33.4.7. Binomial distribution

Begin with \( k = 0 \) and generate \( u \) uniform on \((0, 1)\). Compute \( P_k = (1 - p)^k \) and store \( P_k \) into \( B_k \). If \( u \leq P_k \) accept \( r_k = k \) and stop. Otherwise, increment \( k \) by one; compute the next \( P_k \) as \( P_k = \left( p/(1-p) \cdot \ln (1/k + 1) \right) + 1 \) to add this to \( B_k \). Again if \( u \leq P_k \) 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 \( f(r; n, p) \), 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 \( A = 0 \) to start. Generate \( u \). Replace \( A \) with \( u \); if now \( A < \exp(-\mu) \), where \( \mu \) is the Poisson parameter, accept \( r_k = k \) 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 \( \mu \geq 10 \) it may be satisfactory (and much faster) to approximate the Poisson distribution by a Gaussian distribution (see our Probability chapter, Sec. 31.4) and generate \( z \) from \( f(z; n, p) \); then accept \( x = \max(0, \mu + \sqrt{\mu + 0.75}) \) where \( |z| \) signify the greatest integer less than or equal to the expression. [12]

33.4.9. Student’s \( t \) distribution

For \( n > 0 \) degrees of freedom (\( n \) not necessarily integer), generate \( z \) from a Gaussian with mean 0 and \( \sigma^2 = 1 \) according to the method of 33.4.4. Next generate \( y \), an independent \( \chi^2 \) random variable with \( n-k/2 \) degrees of freedom. Then \( z \cdot \sqrt{n-1} / f \) 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 $\Gamma$, 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 FSC-SCRI-87-50 (1987). This generator is available as the CERNLIB routine V113, RANMAR, by F. Carminati and F. James.


11. J.L. Levy, 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. Köllig in the CERNLIB routine V136, RNPSSN.
34. Monte Carlo particle numbering scheme

Revised March 2004 by L. Barnes (Fermilab), I.G. Knowles (Edinburgh U.), S. Navas (U. Granada), T. Sjöstrand (Land U.), and T. Trippe (BNL).

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 NJOY/4 standard interface.

The general form is a 1-digit number:

$$\pm \nu_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n \eta_n$$

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 $B^0$ 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) $\eta_{n-1}$ are quark numbers used to specify the quark content, while the rightmost digit $\eta_n = 2J + 1$ gives the spin of the system (except for the $S$ and $K^0_S$). The scheme does not cover particles of spin $J > 4$.
4. Diquarks have 4-digit numbers with $\eta_4 \geq \eta_3 \geq \eta_2 \geq \eta_1 = 0$.
5. The numbering of mesons is guided by the nonrelativistic $(L=\Sigma$ decoupled) quark model, as listed in Tables 14.2 and 14.3.
   a. The numbers specifying the meson’s quark content confirm to the convention $\eta_0 = 0$ and $\eta_4 \geq \eta_3 \geq \eta_2 \geq \eta_1 = 0$. The special case $K^0_S$ is the sole exception to this rule.
   b. The quark numbers of flavorless, light $(q,d,s)$ mesons are: 11 for the member of the isospin triplet $(\pi^0, \pi^+, \pi^-)$, 22 for the lighter isosinglet $(\eta, \omega, \phi)$, and 33 for the heavier isosinglet $(\eta', \phi')$. Since isosinglet mesons are often large mixtures of $u\bar{u}$ and $d\bar{d}$ states, 33 and 32 are assigned respectively.
   c. The special numbers 310 and 130 are given to the $K^0_S$ and $K^{*0}_S$ respectively.
   d. The fifth digit $\eta_5$ is reserved to distinguish mesons of the same total $(J)$ but different spin (S) and orbital (L) angular momentum quantum numbers. For $J > 6$ the numbers are $|L,S,F| = (J-1,1)$, $\eta_0 = 0$, $(0,0) \eta_1 = 1$, $(1,1) \eta_2 = 2$ and $(J+1,1) \eta_3 = 3$. For the exceptional case $J = 0$ the numbers are $(0,0) \eta_2 = 0$ and $(1,1) \eta_3 = 1$ (i.e. $\eta_3 = L$). See Table 34.1.

<table>
<thead>
<tr>
<th>$L=J-1$</th>
<th>$S=1$</th>
<th>$L=J,S=0$</th>
<th>$L=J,S=1$</th>
<th>$L=J+1,S=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>J code</td>
<td>$^{3P1}C$</td>
<td>L</td>
<td>J code</td>
<td>$^{3P1}C$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- 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 03133(1$^{2P1}$) $K_{1B}$ and the $K_1(1400)$ is numbered 20313(1$^{2P1}$) $K_{1A}$.

- The sixth digit $\eta_6$ is used to label mesons radically excited above the ground state.
- Numbers have been assigned for complete $\eta_n = 0$ S- and $P$-wave multiplets, even where states remain to be identified.
- In some instances assignments within the $\eta$ meson model are only tentative, here best guess assignments are made.
- Many states appearing in the Meson Listings are not yet assigned within the $\eta$ model. Here $\eta_0$ and $\eta_2$ are assigned according to the state’s likely flavors and spin; all such unassigned light isotensor states are given the flavor code 22. Within these groups $\eta_4 = 0, 1, 2, \ldots$ 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.
- The numbering of baryons is again guided by the nonrelativistic quark model, see Table 14.5.
   a. The numbers specifying a baryon’s quark content are such that in general $\eta_1 \geq \eta_2 \geq \eta_3$.
   b. Two states exist for $J = 1/2$ baryons containing 3 different types of quarks. In the lighter baryon $\{u, \bar{u}, d, \bar{d}, \ldots\}$ the light quarks are in an antisymmetric $(J = 0)$ state while for the heavier baryon $(S^0, \Sigma^+, \ldots)$ they are in a symmetric $(J = 1)$ state. In this situation $\eta_0$ and $\eta_4$ 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 90 $\beta$ numbers for excited baryons is recommended.
   d. For pentaquark states $n = 9$, $\eta_0 \eta_2 \eta_2 \eta_0 \eta_0$ gives the four quark numbers in order $n_2 \geq n_3 \geq n_4 \geq n_5 \geq n_6$ gives the antiquark number, and $J = 2J + 1$, with the assumption that $J = 1/2$ for the states currently reported.
- The gluon, when considered as a gauge boson, has official number 21. In codes for gluobos, however, 9 is used to allow a notation in close analogy with that of hadrons.
- The pomeron and odderon trajectories and a generic reggeon trajectory of states in QCD are assigned codes 900, 990, and 110 respectively, where the final digits indicate the 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 low-energy member.
- Two-digit numbers in the range 21–30 are reserved for generator-specific pseudoparticles and concepts.
- 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 boxon content of a two-Higgs-doublet scenario and of additional $SU(2)_L \times U(1)$ groups are found in the range 31–40.
   c. “One-of-a-kind” exotic particles are assigned numbers in the range 41–60.
   d. Fundamental supersymmetric particles are identified by adding a nonzero $n_6$ to the particle number.
   e. 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 skerons, the lighter physical state is given the smaller basis state number.
- Theorist states have $n = 3$, with technifermions treated like ordinary fermions. States which are ordinary color singlets have $n = 0$. Color octets have $n = 1$. If a state has non-trivial quantum numbers under the topcolor groups $SU(3)_c \times SU(3)_L$, the quantum numbers are specified by $t_{ij} j_{ij}$, where $i$ and $j$ are 1 or 2. $n_6$ is then 2$I + J$. The coloron, $V_5^b$, is a heavy gluon color octet and thus is 3$00021$.
7. L. Gerren, StedHep, Monte Carlo Standardization at FNAL, Fermilab PM-0001 and StedHep WWW site: http://omega.fnal.gov/pmg/lasthep/

### References:

<table>
<thead>
<tr>
<th>Particle, leptons are identified by setting $n = 4$.</th>
<th>setting $n = 99$.</th>
<th>Concerning the non-zero numbers, it may be noted that only quarks, excited quarks, antiquarks, and diquarks have $n_2 = 0$; only gluons, baryons (including psuedoquarks), and the odderon have $n_1 = n_3 = 0$; and only mesons, the reggeon, and the pomeron have $n_1 = n_2 = 0$. Concerning mesons (not antimesons), if $n_3$ is odd then it labels a quark and an antiquark if even.</th>
</tr>
</thead>
</table>
| This text and lists of particle numbers can be found on the WWW. The StedHep Monte Carlo standardization project maintains the list of PDG particle numbers, as well as numbering schemes from most event generators and software to convert to the different schemes. | | }

### QUARKS

<table>
<thead>
<tr>
<th>$d$</th>
<th>$u$</th>
<th>$s$</th>
<th>$c$</th>
<th>$b$</th>
<th>$t$</th>
<th>$b'$</th>
<th>$t'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

### DIQUARKS

<table>
<thead>
<tr>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
<th>$(u_d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

### SUSY PARTICLES

<table>
<thead>
<tr>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
<th>$\tilde{d}_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

### LIGHT I – 1 MESONS

<table>
<thead>
<tr>
<th>$\pi^+$</th>
<th>$\pi^0$</th>
<th>$\eta$</th>
<th>$\eta'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### LIGHT I – 0 MESONS

<table>
<thead>
<tr>
<th>$\rho(700)$</th>
<th>$\rho(700)^+$</th>
<th>$\rho(700)^-$</th>
<th>$\rho(700)^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

### LIGHT II – 1 MESONS

<table>
<thead>
<tr>
<th>$\rho(700)^0$</th>
<th>$\rho(700)^{+-}$</th>
<th>$\rho(1450)^0$</th>
<th>$\rho(1450)^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

### LIGHT II – 0 MESONS

<table>
<thead>
<tr>
<th>$\omega(782)$</th>
<th>$\omega(782)^0$</th>
<th>$\omega(782)^{+-}$</th>
<th>$\omega(782)^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

### TECHNOLOR PARTICLES

<table>
<thead>
<tr>
<th>$T_{\alpha}^{+}$</th>
<th>$T_{\alpha}^{-}$</th>
<th>$T_{\alpha}^{0}$</th>
<th>$T_{\alpha}^{0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

### SPECIAL PARTICLES

<table>
<thead>
<tr>
<th>$\varphi(1600)$</th>
<th>$\varphi(1600)^0$</th>
<th>$\varphi(1600)^{+-}$</th>
<th>$\varphi(1600)^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

### EXCITED PARTICLES

<table>
<thead>
<tr>
<th>$\pi^*(1400)$</th>
<th>$\pi^*(1400)^+$</th>
<th>$\pi^*(1400)^{+-}$</th>
<th>$\pi^*(1400)^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
</tbody>
</table>

### GAUGE AND HIGGS BOSONS

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$Z^0$</th>
<th>$W^+$</th>
<th>$H^+$</th>
<th>$Z^0/H^+$</th>
<th>$W^0/Z^0$</th>
<th>$W^+/H^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
</tr>
</tbody>
</table>

### References:

7. L. Gerren, StedHep, Monte Carlo Standardization at FNAL, Fermilab PM-0001 and StedHep WWW site: http://omega.fnal.gov/pmg/lasthep/
Footnotes to the Tables:

a) Numbers or names in bold face are new or have changed since the 2002 Review, §2.

b) The physical states are admixtures of the pure $\bar{s}s, \bar{u}u, \bar{d}d, \bar{c}c$, and $\bar{b}b$ states.

c) In this draft we have only provided one generic lepton number 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) $\Sigma^{+}$ and $\Sigma^{-}$ are alternate names for $\Sigma(1385)$ and $\Sigma(1530)$.

---
## 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 $\sqrt{-8/15}$.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\nu'$</th>
<th>$d^2\beta = \cos \theta$</th>
<th>$d_2\beta = \cos \theta$</th>
<th>$d_2\beta = \sin \theta$</th>
<th>$d_2\beta = -\sin \theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}$ cos $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}$ cos $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}$ cos $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
<td>$\frac{1}{2}$ sin $\theta$</td>
</tr>
</tbody>
</table>

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 single, 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 coefficients 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 $\sqrt{2}$ 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 $\Xi \rightarrow \mathcal{D}K$ element of the $10 \otimes 10 \otimes 8$ matrix is $-\sqrt{6}/\sqrt{27} = -1/2$.

The matrices, $\mathcal{D}$ for example, in decuplet ratio of $\Sigma(\text{symmetric})$ and $\Xi(\text{antisymmetric})$ relationships: alternative isospin Clebsch-Gordan coefficients, we obtain, e.g.,

$$\Gamma (\mathcal{D} \rightarrow \Xi \mathcal{K}) \sim \left( -\sqrt{3}/2 \right) \phi_1 + \sqrt{7}/2 \phi_2 .$$

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

$$\Gamma (\mathcal{D} \rightarrow \phi) = \frac{1}{2 \sqrt{3}} \phi_1 + \phi_2 .$$

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

$$\Gamma (\Xi \rightarrow \mathcal{D} \mathcal{K}) = \left( -\sqrt{3}/2 \right) \phi_1 + \sqrt{7}/2 \phi_2 .$$

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,

$$\mathcal{L} = -\sqrt{2} \bar{D} \mathcal{T} (\mathcal{D}, \mathcal{D}) \mathcal{F} + \sqrt{2} \bar{F} \mathcal{T} (\mathcal{D}, \mathcal{D}) \mathcal{F},$$

where $[\mathcal{D}, \mathcal{D}] = \mathcal{D} \mathcal{D} - \mathcal{D} \mathcal{D}$ and $[\mathcal{D}, \mathcal{D}] = \mathcal{D} \mathcal{D} + \mathcal{D} \mathcal{D}$, are

$$D = \frac{\sqrt{30}}{4} g_1 , \quad F = \sqrt{24} g_2 .$$

Thus, for example,

$$\Gamma (\Xi \rightarrow \mathcal{D} \mathcal{K}) \sim (F - D) \sim (1 - 2\alpha)^2 ,$$

where $\alpha = \mathcal{D} \mathcal{D} + \mathcal{F} \mathcal{F}$. (This definition of $\alpha$ is de Swart's. The alternative $D(D + \mathcal{F})$, due to Gell-Mann, is also used.) The generators of SU(3) transformations, $\lambda_\alpha$ ($\alpha = 1, 8$), are $3 \times 3$ matrices that obey the following commutation and anti-commutation relationships:

$$[\lambda_\alpha, \lambda_\beta] = i \epsilon_{\alpha \beta \gamma} \lambda_\gamma , \quad [\lambda_\alpha, \lambda_\beta] = \frac{1}{2} f_{\alpha \beta \gamma} \lambda_\gamma ,$$

$$[\lambda_\alpha, \lambda_\beta] = i \epsilon_{\alpha \beta \gamma} \lambda_\gamma + \frac{1}{2} f_{\alpha \beta \gamma} \lambda_\gamma ,$$

where $I$ is the $3 \times 3$ identity matrix, and $\frac{1}{2} f_{\alpha \beta \gamma}$ is the Kronecker delta symbol. The $f_{\alpha \beta \gamma}$ are odd under the permutation of any pair of indices, while the $d_{\alpha \beta \gamma}$ are even. The nonzero values are

$$1 \rightarrow 8 \otimes 8$$

$$(\Delta) \rightarrow (N \mathcal{K}) \pi \mathcal{K} = \frac{1}{\sqrt{3}} \left( \begin{array}{ccc} 2 & 3 & -1 \\ -2 & 4 & -2 \\ 2 & -3 & -2 \end{array} \right) .$$

$$8_1 \rightarrow 8 \otimes 8$$

$$\bar{N} \rightarrow \bar{N} \mathcal{K} \pi \mathcal{K} = \frac{1}{\sqrt{3}} \left( \begin{array}{ccc} 9 & -1 & -1 \\ -2 & -3 & -2 \\ 2 & -3 & -2 \end{array} \right) .$$

$$10 \rightarrow 8 \otimes 8$$

$$N \rightarrow N \mathcal{K} \pi \mathcal{K} = \frac{1}{\sqrt{3}} \left( \begin{array}{ccc} 3 & 3 & -3 \\ -2 & 3 & 3 \\ 2 & -3 & 3 \end{array} \right) .$$

$$10 \rightarrow 10 \otimes 8$$

$$\Delta \rightarrow \Delta \mathcal{K} \mathcal{K} = \frac{1}{\sqrt{15}} \left( \begin{array}{ccc} 15 & -3 & -3 \\ 8 & 0 & -8 \\ 0 & 8 & -8 \end{array} \right) .$$

The $\lambda_\alpha$'s are

$$\lambda_\alpha = \left( \begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right) , \quad \lambda_\beta = \left( \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right) .$$

Equation (36.7) defines the Lie algebra of SU(3). A general 8-dimensional representation is given by a set of $d \times d$ matrices satisfying Eq. (36.7) with the $f_{\alpha \beta \gamma}$ 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 (LINL).

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 $\Delta$-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 set of integers specifies a multiplet. For an SU(2) multiplet such as an isospin multiplet, the single integer $\alpha$ 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 multiplets). In SU(3), the two integers $\alpha$ and $\beta$ 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

![SU(3) Octet and Decuplet Diagrams]

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

37.2. Number of particles

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

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

$$N = \frac{\alpha + 1}{1}.$$  

(37.1)

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

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

(37.2)

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

$$N = \frac{(\alpha + 1)(\beta + 1)(\gamma + 1)}{1}.$$  

(37.3)

Note that in Eq. (37.3) there is no factor with $(\alpha + \gamma + 2)$: only a nonnegative sequence of the label integers appears in any factor. One more example is the SU(5) multiplet

$$N = N(\alpha, \beta, \gamma, \delta) = \frac{(\alpha + 1)(\beta + 1)(\gamma + 1)(\delta + 1)}{1}.$$  

(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 (of some other symbol) arranged in one or more left-justified rows, with each row being at least as long as the row beneath it. The correspondence between a diagram and a multiplet label is: The top row juts out $\alpha$ boxes to the right past the second row, the second row juts out $\beta$ 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

$$\begin{array}{ccc}
\text{a} & \text{b} & \text{c} \\
\end{array}$$

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 $aba$ and $aabc$ are admissible sequences and $abb$ and $aaba$ 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 $n$-meson octet and the baryon octet), we start with

$$\begin{array}{ccc}
\text{a} & \text{a} & \text{a} \\
\end{array}$$

and

$$\begin{array}{ccc}
\text{a} & \text{b} & \text{b} \\
\end{array}$$

(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:

$$\begin{array}{ccc}
\text{a} & \text{a} & \text{a} \\
\end{array}$$

$$\begin{array}{ccc}
\text{a} & \text{b} & \text{b} \\
\end{array}$$

(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:

$$\begin{array}{ccc}
\text{b} & \text{b} & \text{b} \\
\end{array}$$

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:

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

In terms of numbers of particles, it may be written:

$$8 \oplus 8 - 27 \oplus 10 = 24 \oplus 8 \oplus 8 \oplus 1.$$