DESIGN OF EXPERIMENTS

D. Drijard
CERN, Geneva, Switzerland

Introduction

This course is mostly devoted to simulation problems. The part which concerns detector optimization was essentially treated during the School in a seminar about the Split-Field Magnet (SRM) detector installed at the CERN Intersecting Storage Rings (ISR). This is not given in the written notes since very little of general use can be said about this subject, unless very trivial. We wanted rather to describe in a detailed way the tools which allow such studies to be made.

The notes start by a summary of statistical terms. The main emphasis is then put on Monte Carlo methods and generation of random variables. The last section treats the utilization of detector acceptance, which will be one of the most important parts to optimize when designing a detector.

1. Statistical terms

In this section we summarize the statistical material which we need in this course, so that no a priori detailed knowledge is assumed.

1.1 Continuous random variable (r.v.)

1.1.1 Range:

For simplicity of the expressions, we will use in the following a unidimensional r.v. and use explicitly its range \( x \in [m,M] \subseteq \mathbb{R} \).

1.1.2 Probability density function (p.d.f.):

\[
 f(x) = \lim_{\Delta \to 0} \frac{\text{Prob.} \left( x - \frac{1}{2} \Delta < X < x + \frac{1}{2} \Delta \right)}{\Delta} .
\]

From its definition, \( f(x) \) is a positive function of unit integral, i.e. \( \forall x \in [m,M] : f(x) \geq 0 \) and \( \int_{m}^{M} f(x) \, dx = 1 \). When it is not explicitly necessary we will not mention the range of integration; that is, we will have the implicit statement \( \forall x \in [m,M] : f(x) = 0 \) for the p.d.f. although the function \( f(x) \) which represents it in its range may take non-zero values outside this range. The abbreviations r.v. and p.d.f. will be used throughout.

1.1.3 Cumulative distribution:

\( F(x) = \int_{m}^{x} f(y) \, dy \) hence \( F(m) = 0 \) and \( F(M) = 1 \).

1.1.4 Conditional distribution:

For multidimensional r.v., we may consider subspaces. Let \( f(x,y) \) be the joint p.d.f. in the r.v. \( x \) and \( y \) (which may still be multidimensional). Then integrating over one r.v., we find the marginal p.d.f. in the other r.v. like \( g(y) = \int f(x,y) \, dx \). Finally, we define the conditional distribution \( h(x|y) = f(x,y)/g(y) \). It represents the p.d.f. of the r.v. \( x \), given a definite value of the variable \( y \). Notice that these functions are still normalized to unity, as p.d.f. should be \( \int f(x,y) \, dx \, dy = \int g(y) \, dy = \int h(x|y) \, dx = 1 \).

1.1.5 Expectation value:

For an arbitrary function \( g(x) \), we define the expectation value of \( g(x) \) over the p.d.f. \( f(x) \):

\[ E_f(g) = \int g(x) \, f(x) \, dx \ . \]

1.1.6 Variance:

\[ V_f(g) = E_f(g^2) - [E_f(g)]^2 \equiv E_f[(g - E_f(g))^2] \geq 0 \ . \]

The second expression exhibits the important positivity condition, which comes trivially from the integral of a positive function.

1.1.7 Change of variable:

We start from a r.v. \( x \) whose p.d.f. is \( f(x) \). We study a new variable \( y = g(x) \) which is a r.v., since it is a function of a r.v. Consequently it has a p.d.f. \( h(y) \) which is certainly defined completely by the knowledge of both \( f(x) \) and \( g(x) \) functions:

\[
 h(y) = \int f(x) \, \delta[y - g(x)] \, dx = \sum_{i} \frac{h(x_i)}{|g'(x_i)|} ,
\]

where the sum \( \sum \) runs over all roots of the equation \( y = g(x) \). For example, \( y = x^2 \) gives

\[
 h(y) = \frac{1}{2|x|} \left[ f(x) + f(-x) \right] = \frac{1}{2\sqrt{y}} \left[ f(\sqrt{y}) + f(-\sqrt{y}) \right] .
\]

1.1.8 Moments:

\( \mu_n = E[(x - E(x))^n] \) is the \( n \)-th central moment;

\[
\mu'_n = E(x^n) \quad \text{is the } n\text{-th algebraic moment.}
\]
The following notations are often used:
\[ \mu = \mu_r \] is the mean value;
\[ \sigma = \sqrt{\sigma^2} = \sqrt{\langle x \rangle} \] is the standard deviation.

1.1.9 Characteristic function:

The Fourier transform of a p.d.f. is called a characteristic function. It is most useful in the study of sums of r.v.:
\[ \phi_X(t) = E(e^{itX}) \equiv \int e^{itx} f(x) \, dx \]
and
\[ f(x) = \frac{1}{2\pi} \int \phi_X(t) e^{-itx} \, dt. \]
The main properties are the following:

i) \( \phi(0) = 1 \) and \( |\phi(t)| \leq 1. \)

ii) \( \phi_{X+Y}(t) = \phi_X(t) \phi_Y(t) \) if \( X \) and \( Y \) are independent r.v. (that is, the joint p.d.f. of \( X \) and \( Y \) is the product of the individual p.d.f. of \( X \) and \( Y \)).

iii) \( \phi_{aX+b}(t) = e^{ibt} \phi_X(at) \) if \( a \) and \( b \) are constant numbers.

iv) \( \mu' = \frac{1}{i} \frac{d}{dt} \phi(t) \bigg|_{t=0} \)

v) \( \mu_r = \frac{1}{i^r} \frac{d^r}{dt^r} e^{-i\mu t} \phi(t) \bigg|_{t=0} \)

1.1.10 Normal law:

Usually designated by \( N(\mu, \sigma^2) \)

p.d.f. \( f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \) with \( x \in [-\infty, \infty] \).

The mean value and standard deviation are explicit in the expression of \( f \):
\[ E(x) = \mu, \quad V(x) = \sigma^2. \]

The characteristic function is \( \phi(t) = \exp \left[ it\mu - \frac{t^2}{2} \sigma^2 \right] \).

It is useful to remember some numbers about the probability content outside ranges centred about the mean value \( \mu \):

\[
\begin{align*}
\text{Prob.} \ ((|x-\mu| > \sigma) & \sim 0.32 \\
\text{Prob.} \ ((|x-\mu| > 2\sigma) & \sim 0.046 \\
\text{Prob.} \ ((|x-\mu| > 3\sigma) & \sim 0.0027.
\end{align*}
\]

It is customary to present experimental results by quoting estimations \( \hat{\mu} \) and \( \hat{\sigma} \). If the p.d.f. of the r.v. analysed were a normal law, we could deduce information on the probability that the true \( \mu \) value lies in some range about \( \hat{\mu} \) expressed in units of \( \hat{\sigma} \).

It turns out that, in a low statistics experiment, the p.d.f. is far from being a normal law and, consequently, \( |\mu - \hat{\mu}| > 3\sigma \) may often be much more probable than the 3\% expected from a normal law. As far as we are concerned in this course, we shall see reasons why we expect asymptotically normal laws, but this must be checked in practical cases. We shall mention (4.8) a simple check. At this point mention is made of a semi-analytical check:

There is a one-to-one correspondence between p.d.f. and characteristic function. Thus we may check this last one, or equivalently the set of moments. It is usual to study quantities based on 3rd and 4th moments, adjusted to be zero for a normal law (and dimensionless):
\[ \gamma_1 = \mu_3/\mu_2^{3/2} \] is the coefficient of skewness,
\[ \gamma_2 = \mu_4/\mu_2^2 - 3 \] is the coefficient of kurtosis.

These coefficients are fairly easy to estimate (they are usually available in histogramming packages) and constitute something of a security for normality, although evidently not a certainty. As an example, a \( \chi^2 \) distribution with \( n \) degrees of freedom [sum of \( n \) r.v. with independent normal distributions \( N(0,1) \)] has \( \gamma_1 = 2\sqrt{n}/n \) and \( \gamma_2 = 12/n \). It is considered reasonably normal for \( n > 30 \), since \( \gamma_1 \) and \( \gamma_2 \) are less than 0.5.

1.1.11 Uniform law:

\[ \forall \ x \in [m,M] : f(x) = \frac{1}{M-m}; \text{ otherwise } f(x) = 0. \]
\[ E(x) = \frac{1}{2} (M-m) \text{ and } V(x) = \frac{1}{12} (M-m)^2. \]

We shall later refer to this law by the notation \( U(m,M) \).

1.2 Discrete random variables

1.2.1 Range

Using a one-to-one application, a discrete set can always be mapped on a set of consecutive integers. It will thus be assumed that the r.v. has a range \( r \in \{m, m+1, \ldots, M, M\} \).

1.2.2 Probability generating function:

\( p(r) \) is the probability that the r.v. takes the value \( r \).

All definitions given for continuous r.v. are valid for discrete r.v. provided integrals are replaced by discrete sums. For example,
\[ \sum_{r=m}^{n} f(x) \, dx = 1 \rightarrow \sum_{r=m}^{n} p(r) = 1. \]

1.2.3 Binomial distribution

\[ B(r|p,n) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \]
with \( r \in \{0,1,\ldots,n\} \) and \( p \in [0,1] \),
\[ E(r) = np \text{ and } V(r) = np(1-p). \]
1.2.4 Poisson law

\[ P(r | u) = \frac{e^{-u} u^r}{r!} \quad \text{with} \quad r \in \{0, 1, \ldots, \infty\} \quad \text{and} \quad u > 0, \]

\[ E(r) = V(r) = u. \]

2. Central limit theorem

Let \( x \) be a continuous r.v. whose p.d.f. is \( f(x) \). Then the arithmetic mean \( \bar{x}_n = (1/n) \sum_{i=1}^{n} x_i \) of \( n \) such independent r.v. is again a r.v. of p.d.f. which we note by \( f_n(\bar{x}_n) \). We can in principle compute \( f_n \) from the knowledge of \( f \), but this is in general a non-trivial calculation \([\text{from 1.1.9}, \phi_n(t) = \left[\phi(t)\right]^n \] but the inverse Fourier transform is normally not simple\). However, we may state asymptotic properties (as \( n \to \infty \)) under fairly general conditions. If we restrict ourselves to the simple case which we study here, a sample of \( n \) r.v. (independent) issued from a common parent distribution, then: if \( V_f(x) < \infty \) (i.e. finite),

\[ \text{law of large numbers} \quad \bar{x}_n \xrightarrow[n \to \infty]{\text{d}} E_f(x), \]

\[ \text{central limit theorem} \quad f_n(\bar{x}_n) \xrightarrow[n \to \infty]{\text{d}} N\left[ E_f(x), \frac{1}{n} V_f(x) \right]. \]

As we will see later, these theorems are the basis of Monte Carlo integration. The first theorem is not very surprising, one expects \( \bar{x}_n \) to get close to \( E_f(x) \) when \( n \) increases; but the second one is less trivial since \( f_n(\bar{x}) \) converges towards a definite law, normal law, independently of the parent distribution \( f(x) \).

The only condition required \([V_f(x) \text{ finite}]\) seems harmless (a fortiori for physics applications where the range of r.v. observed is always finite), but it should be noticed that the theorem says nothing about the rate of convergence (how large \( n \) should be?). We can get a feeling by the study of a few examples.

2.1 Normal law

If \( f(x) \) is a normal law, then \( f_n(\overline{x}_n) \) is a normal law as well, whatever \( n \). Indeed, using the characteristic function \( (1.1.9) \) in the special case of normal law \( (1.1.10) \), then

\[ \phi_X(t) = \exp \left[ i t u - \frac{1}{2} t^2 \sigma^2 \right]. \]

If \( y = \frac{1}{n} \sum_{i=1}^{n} x_i \), then \( \phi_Y(t) = \exp \left[ i t u - \frac{1}{2} t^2 \sigma^2 \right]. \)

Finally \( \bar{x}_n = (1/n) \sum_{i=1}^{n} x_i \), then \( \phi_{\bar{x}_n}(t) = \phi_Y(t/n) = \exp \left[ i t u - \frac{1}{2} t^2 \sigma^2/n \right]. \) It then follows that \( f_n(\bar{x}_n) \equiv N(u, \sigma^2/n) \).

2.2 Cauchy law

\[ f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \quad \text{with} \quad x \in [-\infty, \infty]. \]

This is the usual Breit-Wigner formula after a change of variable such that the r.v. \( x \) is centred and rescaled. It is easily seen that the mean value \( E(x) \) is undefined and the variance \( V(x) \) is infinite. Let us restrict the range \( x \in [-A, B] \) with \( A, B > 0 \). Then

\[ f(x) = \frac{1}{\arctan A + \arctan B} \cdot \frac{1}{1 + x^2}, \]

\[ E(x) = \frac{1}{\arctan A + \arctan B} \cdot \frac{1}{2} \ln \frac{1 + B^2}{1 + A^2}, \]

\[ V(x) = \frac{A + B}{\arctan A + \arctan B} - 1. \]

If now we let \( A \) and \( B \) tend towards \( \infty \), then \( E(x) \) will depend on the ratio \( B/A \), and is thus undefined. We may cure this difficulty if we impose, say, \( A = B \), then \( E(x) = 0 \), but inevitably \( V(x) \) is infinite. So we are in the case where our asymptotic theorems should fail. Indeed, using again characteristic functions, we can show that

\[ \phi_X(t) = e^{-|t|} \quad (\text{no derivative at} \ t = 0, \ \text{hence no moments}). \]

Then if

\[ y = \sum_{i=1}^{n} x_i, \quad \phi_Y(t) = e^{-n|t|}. \]

Finally for

\[ \bar{x}_n = \frac{1}{n} y, \quad \phi_{\bar{x}_n}(t) = e^{-|t|}. \]

Thus \( f_n(\bar{x}_n) \) is identical to \( f(x) \). As in the preceding case, the type of law is not modified by the arithmetic mean process but, now, the width does not decrease with \( n \). This may seem an academic problem since the range is always finite in real life, but the rate of convergence towards a normal law will be slow. To get an order of magnitude, we may study the kurtosis \( \gamma_2 \) as a function of the size \( n \) of the sample and of the parameter \( A \). Using the moments expansion of the characteristic function given in 1.1.9, we find, without explicit reference to the Cauchy law in fact, \( (\gamma_2)_n = 1/n \ (\gamma_2)_\infty \). Then for \( A > 10 \) we may approximate \( \arctan A \) by \( \pi/2 \), and we get \( (\gamma_2)_\infty = (\pi A/6) - 3 \). Consequently, the larger \( A \), the larger \( n \) will be necessary to have, say, \( (\gamma_2)_n < 0.5 \). Notice that \( (\gamma_2)_\infty = 0 \) for \( A = 6 \), which shows that \( \gamma_2 = 0 \) is not so good a test of normality.
2.3 Uniform law

If \( x \) follows a uniform law \( U(-0.5, 0.5) \), then the r.v. \( u_n = x / \sqrt{n} \) follows asymptotically a normal law \( N(0,1) \). This property is often used to generate r.v. according to a normal law, and \( n = 12 \) is frequently employed, which saves the calculation of a square root. One can compute directly
\[
(y_2)_n = -1.2, \quad \text{and from the preceding paragraph}
\]
\[
(y_2)_n = (1/n) (y_2) = -0.1.
\]
The range of \( u_n \) is \([-6, 6]\), which is large enough for most applications.

3. Monte Carlo integration

If we want to estimate an integral \( I = \int g(x) \, dx \), we may use standard techniques, such as the Simpson rule. If the integral is to be performed in multidimensions, then this method must be abandoned. Even more elaborate methods, such as Gauss methods, cannot be used for, say, more than five dimensions, and indeed in particle physics integrals of dimensions larger than 10 are often encountered.

Such methods require sampling the function on a regular grid; so, for example, \( n = 10^d \) points if we require \( d \) dimensions 10 points per dimension. This evidently becomes rapidly prohibitive. In addition, the error made on the integral varies asymptotically (large \( n \)) like \( n^{-2/d} \) for the Simpson rule. The Monte Carlo method, which we study now, does not suffer from such difficulties; in particular, we will see that the error on the integral varies like \( n^{-1/2} \), whatever the dimensions.

3.1 Conventional method

Let \( x_1, x_2, \ldots, x_n \) be a set of \( n \) independent r.v. which follow a common uniform law \( U(0, M) \), then a new r.v. \( \bar{I} = \sum_{i=1}^{M} g(x_i) / n \) is constructed. By construction \( E(\bar{I}) = I \) and \( V(\bar{I}) = 1/n \, V(g) \).

From the central limit theorem (apart from rare cases where \( V(g) \) is infinite), the p.d.f. of \( \bar{I} \) tends for large \( n \) towards a normal law \( N[I, (1/n) \, V(g)] \). The (asymptotic) normality allows us to make probabilistic statements about the closeness of \( \bar{I} \) and \( I \).

In particular, there is about 95% probability that
\[
|I - \bar{I}| < 2 \sigma (1.1.10), \quad \text{where} \quad \sigma^2 = (1/n) \, V(g).
\]
As mentioned earlier, \( \sigma \propto n^{-1} \). The same conclusions are obtained in multidimensions, simply \( x_i \) is now a set of \( d \) uniform r.v. (instead of one) and, in the expression of \( \bar{I} \), one must replace \( (M-m) \) by the volume \( \Omega \) spanned.

3.2 Importance sampling

We covered the region of integration with a uniform density of points (since uniform r.v. were used), but better results can be expected if the function \( g(x) \) is sampled with higher density in the regions where it is important. Let us recast the problem using an arbitrary p.d.f. \( f(x) \) instead of uniform law.

Let us define the "weight" \( w(x) = g(x) / f(x) \). Then the original integral is \( I = \int g(x) \, dx = \int w(x) \cdot f(x) \, dx \). Our estimate will be \( \bar{I} = 1/n \sum_{i=1}^{n} w(x_i) \) and, from central limit theorem, the p.d.f. of \( \bar{I} \) will tend towards a normal law \( N[I, (1/n) \, V(w)] \). The aim is now to adjust \( f(x) \) in order to minimize \( V(w) \) which governs the precision of the estimation. We first study this optimization problem; then in the next section we shall describe methods of generating r.v. according to some p.d.f. \( f(x) \).

The ideal case would be \( V(w) = 0 \), which implies \( g(x) = I \cdot f(x) \). Evidently, it would not even be necessary to generate a single r.v. since we can pretend to know explicitly the integral of \( g(x) \) \( [f(x) \) is of unit integral]. We address ourselves to the cases where such calculations are not possible, or at least not known.

At the other extreme, we may find \( V(w) = \infty \), in which case our theorems would not apply. This is apparently another academic problem but it happens in some careless applications, and stays unnoticed since no infinite number appears in the calculations. Assume that in some region of the space of integration the function \( f(x) \) is equal to 0 [but not the function \( g(x) \) to integrate]. Then in this region \( w(x) \) is infinite and consequently \( V(w) \) is infinite. However, this will not show up in the calculations since there will be no occurrence of the r.v. \( x \) in this region. Simply \( g(x) \) will be integrated in a restricted range, hence wrongly estimated; \( V(\bar{I}) \) infinite simply reflects the absence of knowledge of a part of the integral.

Practically, one will search for a function \( f(x) \) reasonably close to \( g(x) \), in order to minimize \( V(w) \), but at the same time one must have a way of generating r.v. according to \( f(x) \). This will be the subject of the next section.

4. Random variable generators
4.1 Uniform law

This is the basis of all r.v. generators out of which they are constructed, as we will see in the following paragraphs. Most computers have a uniform r.v. generator, so we will not study this problem.
Simply, it should be remembered that the quality of such a generator (true uniformity, absence of correlations) will reflect itself on subsequent generators; hence one should not adopt a new generator without elaborate tests.

4.2 Inverse cumulative distributions

This is something of a brute force method — that is, it does not require much finesse. Unfortunately it works only for simple cases. Let \( x \) be a r.v. of p.d.f. \( f(x) \) and cumulative distribution \( F(x) \). Then one can show that the r.v.

\[ y = F(x) \]

follows a uniform law \( U(0,1) \). Reversing the statement, if \( y \) follows \( U(0,1) \) then \( x = F^{-1}(y) \) follows the original \( f(x) \). Thus the method requires a knowledge of the inverse cumulative distribution. This is rarely possible in an analytical way, but, at least in the one-dimensional case, we often construct a table of \( F(x) \) by numerical integration and later solve the inverse equation by interpolation.

As an example, let us take an exponential law

\[ f(x) = e^{-x} \text{ with } x \in [0,\infty] \]

then

\[ y = F(x) = \int_0^x f(u) \, du = 1 - e^{-x} \]

and finally

\[ x = -\ln(1-y) \].

So, using for \( y \) a uniform r.v., we get for \( x \) an exponential p.d.f.

4.3 Change of variable

We have seen (1.1.7) how to compute the p.d.f. of a function of r.v. If we start from a uniform law, in one dimension, then the method is identical to the preceding one [i.e. if we find a change of variable \( x = h(y) \) such that, when \( y \) follows a normal law, \( x \) follows the p.d.f. \( f(x) \) in which we are interested, then this change of variable is certainly \( x = F^{-1}(y) \)]. However, in multidimensions this may not be so trivial. The following example shows how to generate r.v. according to a normal law (it is standard on most computers).

Let \( \theta \) be an angle which follows \( U(0,2\pi) \), and \( z \) follow \( U(0,1) \). Then

\[ x = \sqrt{-2 \ln r} \cos \theta \]

and

\[ y = \sqrt{-2 \ln r} \sin \theta \]

both follow normal laws \( N(0,1) \), and they are uncorrelated. Compared to the method mentioned earlier, the average of \( n \) uniform r.v., the present method is exact but a priori more time consuming.

4.4 Special tricks

If \( r_1 \) and \( r_2 \) both follow a uniform law \( U(0,1) \), then

\[ x = -\ln(r_1 \cdot r_2) \]

follows \( f(x) = xe^{-x} \). It can be seen by \( f(x) = \int_0^x f(d_1) dr_2 \delta[x + \ln(r_1 \cdot r_2)] \). This method is used, in particular, to generate transverse momentum distributions of the type \( d\sigma/dp_T^2 \sim e^{-\delta p_T} \).

4.5 Keep or reject method

Assume we want to generate \( x \) according to \( f(x) \) with \( x \in [m,M] \). Let \( \alpha = \max f(x) \). Then we may use the following algorithm:

1) generate \( u \) according to \( U(m,M) \);
2) generate \( r \) according to \( U(0,\alpha) \):

- if \( r > f(u) \), restart the procedure from point 1;
- if \( r < f(u) \), keep the r.v. \( u \).

The set of r.v. \( u \) generated by such a method have \( f(u) \) as p.d.f. on the average, i.e. at the limit of a large number of trials, the efficiency of the method is

\[ e = 1/\alpha \cdot (M-m) \]

which represents the ratio of the number of accepted generations to the number of times the algorithm was executed [i.e. the ratio of the integral of the curve to the area of the rectangle where points \( (r,n) \) were uniformly generated]. Notice that it is mandatory to redo both steps 1 and 2 when the test fails. This method may be easily generalized in multidimensions, but its efficiency may then become too small. Even in one dimension the method may be useless when \( f(x) \) exhibits sharp peaks, i.e. of small width with respect to the range.

4.6 Weighting method

Assume that we found a random generator \( g(x) \), close enough to the p.d.f. \( f(x) \) which we want, but not identical. We have now the choice between two methods. Either we keep the r.v. \( x \) thus generated but every subsequent calculation must be done using the weight \( w(x) = f(x)/g(x) \), or we use keep or reject method. We shall study how to choose between these two methods, given the precision required, by comparing the time spent (computer time nowadays).

Let \( \tau_\theta \) be the time necessary to generate the r.v. \( x \), and \( \tau_\delta \) the time necessary to utilize it. This last time may vary in very large proportions. As extreme cases, the shortest time may be to simply record a dichotomic variable, such as a forward backward asymmetry study; the longest may be the complete analysis of a high multiplicity event. Practically, \( \tau_\delta \) may vary from 10 ms to 5 s (in the CDC 7600 computer).

4.6.1 With weights

From the foregoing, it can be deduced that, loosely speaking, a weighted event is worth

\[ E^2(w)/E(w^2) \]

and requires a time \( (\tau_\theta + \tau_\delta) \). Thus the
time required for an "equivalent event" is 
\[ \tau_1 = (\tau_g + \tau_u) \cdot (E(w)/E^2(w)). \]

4.6.2 Without weights

The keep or reject method is applied to the function \( w(x) \) (it was described in Section 4.5 for a p.d.f., but it need not be the case). So we must find an estimation of the maximum weight \( w_{\text{max}} \). In practical cases, it is rarely obtained by analytical methods but rather on a large enough number of trials, by keeping track of the largest weight observed. Then, in subsequent applications, for each r.v. \( x \) generated with a weight \( w \), we generate \( r \) according to \( U(0, w_{\text{max}}) \) and keep the generated value \( x \) if \( w > r \) (otherwise we restart \( x \) and \( r \) generation). The r.v. thus accepted are then given a unit weight (or any arbitrary but fixed value). Reformulating the efficiency given in Section 4.5 for a function \( w(x) \) whose integral is not unity, one finds 
\[ e = E(w)/w_{\text{max}}. \]
Finally, the time needed for an "equivalent event" is 
\[ \tau_2 = \tau_g \cdot \left[ w_{\text{max}}/E(w) \right] + \tau_u. \]

4.6.3 Example

We have just obtained the times \( \tau_1 \) and \( \tau_2 \) for the two methods of generating and utilizing one equivalent event; that is, corresponding to a common statistical precision (asymptotically). We can see how different these times may be in the case of high-energy, high-multiplicity events 
\( \sqrt{s} \approx 50 \text{ GeV}, E(m) \approx 20 \), where \( m \) is the total multiplicity.

Known algorithms which allow complete events to be simulated exist with the following characteristics:

\[ \tau_g \approx 0.1 \text{ ms/particle}; \]
\[ E^2(w)/E(w^2) \approx 0.1; \]
\[ E(w)/w_{\text{max}} \approx 0.01. \]

We will assume that the events generated are saved and later used many times for different studies (setting up of the analysis chain, for example). An order of magnitude may be 100 utilizations at the cost of 25 ms per particle. Then we get
\[ \tau_1 = (10^{-4} + 2.5) \times 10 = 25 \text{ s/particle}, \]
\[ \tau_2 = (10^{-4} \times 10^2 + 2.5) = 2.5 \text{ s/particle}. \]

A sample, rather meager, of 1000 events of multiplicity \( \approx 20 \) was used in such conditions but, although these events have been used over three or four years, we were certainly legitimatized in adopting the "no weight" method in view of the factor of 10 in time spent. Evidently this conclusion depends on specific cases, and we may be led to use the first method when the utilization time is low enough.

4.7 Johnson distributions

The Johnson family of distributions provides a wide range of shapes which we may use to represent empirical distributions of which we do not a priori have the p.d.f., or as a first step followed by a "weight or no weight" method. These distributions result from a change of variable applied to a r.v. which follows a normal law \( \mathcal{N}(0,1) \). The interested reader should consult: G.J. Hahn and S.S. Shapiro, Statistical models in engineering (John Wiley and Sons, New York, 1967).

4.8 Remarks

The preceding statements, based on mean values, imply asymptotic limits, i.e. large samples. It is most useful to study the finite sample behaviour. The simplest way seems to be the following.

Every study that can be made is eventually summarized by a few numbers, such as, for example, the width of a mass distribution when resolution is analysed. If it is decided to generate \( n \) events for this study, then, say, 10 independent studies could be done on samples of size \( n/10 \). Then the distribution of the results could be compared with their otherwise estimated variance.

4.9 Conclusions

A serious pre-study of the problem in hand is extremely useful, especially when large computer time is needed:

- invest brains in r.v. generation, factors of 10 in time spent are not rare;
- choose "weight or no weight" method depending on subsequent use;
- not only may computer time be wasted by careless study but, in addition, stupid results may be obtained (that is, very imprecise).

5. Event simulation

In most cases, we want to generate exclusive events, according to some physics hypothesis. Sometimes the study may be done on the basis of the four-momenta of the particles. However, on many occasions, we may want a more refined study which takes into account the characteristics of the detector. In this case we must simulate completely the response of the detector both in terms of its
capacity to record information and in terms of background generated.

We must in general follow the particles in a magnetic field using Runge-Kutta methods or pieces of helix. Then global effects must be simulated, multiple scattering and energy loss, and discrete effects such as decays, secondary interactions, 9-rays, etc. Eventually the same procedure must also be done for secondary particles. The formulae for such processes exist in classical text books; only for special studies do we need to search in the recent literature. Eventually the detector response must be simulated in terms of multiwire proportional chambers, drift chambers, calorimeters, etc.

In some cases, we must artificially enhance an effect. For example, in an experiment which looks for prompt electrons, we must study sources of background, such as 9-rays. A priori, the experiment was designed in such a way that the particles cross a small thickness of matter, so that few 9-rays are produced. A straightforward analysis would give rise to a small number of such background electrons (hence poor statistical precision from Poisson law or binomial law) at the price of a lengthy analysis of many events. So the idea is to enhance 9-ray production. For example, at the level of individual particles, we may compute the probability that a 9-ray was produced in the thickness traversed, attach it as a weight to the track, and generate a 9-ray in this thickness according to the relevant exponential law but with certainty (i.e. probability 1). Thus we save the generation of the track and the calculation of its trajectory in the magnetic field for the vast majority of cases where a 9-ray would not be produced. When two or more electrons should be studied in the same event, then the problem becomes more complicated (since one wants to preserve the right multiplicity of 9-rays).

6. Acceptance

Given an event, defined by the four-momenta of the particles produced, we observe in general only a subset of these particles, with four-momenta slightly different. The imperfection of the detector has two components (although one goes continuously from one to the other): blindness in some regions of phase space, and imprecision in the observations. The concept of acceptance is used in a loose sense to measure this imperfection. To be definite, the p.d.f. of the observations f(o) results from a convolution integral between the p.d.f. φ(p) of the four-momenta p defined by physics, and the acceptance a(o/p) of the detector

\[ f(o) = \int a(o/p) \phi(p) \, dp. \]

The object of experimental physics is to estimate φ(p) from observation of f(o) and calculations of a(o/p). In full generality, for non-trivial cases, the problem cannot be solved since it is ill-conditioned. As is often the case, we have two methods: the first one is the good one but it is difficult to apply; the second one is not very good but easier to apply, hence widely used.

6.1 Distorting physics

We start with an assumption about the underlying physics φ(p). Then we estimate the expected results through the convolution integral mentioned earlier, and compare it with the actual observations, with possible estimates of some free parameters in φ(p) (using classical statistical methods). In the case of clear disagreement, we must modify the assumption φ(p), but in the case of agreement we may infer only compatibility, not rightness of hypothesis (the problem is indeed ill-conditioned). The main difficulty of this method is that on the one hand we must have a complete description of physics through φ(p), and on the other hand the calculations must be done with a statistical precision at least as good as that of the experiment proper.

6.2 Undistorting observations

Let us pretend that o ≥ p, at least on the average, for this phase of calculations. Then we may say that an observed event is worth a⁻¹ produced events (a ≤ 1 by definition). This method is often the only one feasible but, unfortunately, it has many drawbacks. This weighting method is not statistically sound as we will see later, but at least it deserves some care. In these calculations we have lost track of the difference between the variables o and p. So, useful information such as the resolution must be estimated independently. In phase-space regions where the detector is blind, infinite weights should be applied to absent observations. This difficulty may be partly solved by symmetry arguments. For this, we need factorization φ(p) = f(x) · g(y), where, say, x is one variable and y represents all the other variables. In addition, we need to know the function f(x). Finally, for all regions spanned by y, there must be a region spanned by x where the detector is not blind. Then, for each y value, we will compute an over-all factor
\[ n(y) = \frac{f(x)}{\omega(y)} \int f(x) \, dx \text{, where the first integral is restricted to the } x \text{ region } \omega(y) \text{ where the acceptance } a(x,y) \text{ is non-zero. Finally, we use a weight } [n(y) \cdot a(x,y)]^{-1} \text{ for each } (x,y) \text{ observation.} \\
\text{The } x \text{ distribution will have the right shape in the region } x \in \omega(y), \text{ scaled up by } n^{-1} \text{ to take into account the blind region. Typically, } x \text{ may be the azimuthal angle around unpolarized beams, in which case } f(x) = 1/2\pi. \text{ If these conditions are not satisfied, then the experiment must be restricted to the observable } y \text{ region.} \\
\]

6.3 Exclusion of large weights

When events of different weights are considered, we can estimate the number of equivalent events by
\[ n_{\text{eq}} \sim \frac{(Dw)^2}{DW^2}. \]
If \( n \) is the number of events, then we can easily verify that \( n_{\text{eq}} \leq n \),
\[ \frac{n}{n_{\text{eq}}} = 1 + \frac{1}{n} \sum_{i=1}^{n} \left( \frac{w_i - \frac{1}{n} \sum_{j=1}^{n} w_j}{n_{\text{eq}}} \right)^2 \geq 1. \]

Coming back to the first expression, if one event has a weight much larger than the weights of the other events, then \( n_{\text{eq}} \ll n \). For example, one event of weight \( 10^3 \) and 999 events of weights 1 give \( n_{\text{eq}} \sim 4 \) for \( n = 10^3 \). Since the statistical precision is governed by \( n_{\text{eq}} \) (uncertainty varies like \( n_{\text{eq}}^{-1} \)), we may question the use of that one event which ruins the sample.

Assume that we want to estimate a parameter; that is, the function \( f(p) \) which describes physics has one free parameter. Then we may compare the two methods in terms of the variance \( \sigma^2 \) of the parameter estimated. If we use the first method ("distorting physics"), then we can show that every event contributes individually to decrease \( \sigma^2 \) (the information of the sample about this parameter, \( 1 \sim 1/\sigma^2 \), is the sum of the information of each event, each piece of information being a positive number). If now we consider the second method, which uses weights, the situation may be very different. In statistical terms, this method is not optimal and \( \sigma^2 \) may increase when a large-weight event is added to the sample contrary to the intuitive feeling.

This is of fundamental importance. One is naturally not inclined to discard events, but such a conservative attitude is sometimes opposite to one's interest! For a more detailed discussion see: W.T. Eadie et al., Statistical methods in experimental physics (North-Holland Publ. Co., Amsterdam, 1971).