FUNDAMENTAL STATISTICAL MEASURES

Before models and analysis of complex systems can be discussed it is necessary to review some fundamentals of statistics based on ideal random variables.

A **random variable** is a variable whose value is determined by a **random process**. If a process produces values that are not perfectly predictable from what is known, it is a random process. Any function of a random variable is also a random variable. A **realization** of the process produces one random value of the variable. A large collection of realizations produced under statistically identical conditions (the same deterministic parameters) is an **ensemble** of "identically prepared" observations. The act of flipping a coin is a random process. If we let $X = 1$ for "heads" and -1 for "tails", X is a random variable. A single flip produces a realization of X and an afternoon of flipping using the same technique would yield an ensemble of realizations.

1. Single variable statistics

Eventually we will be interested in fields of imperfectly predictable (random) variables which vary in time and space. And in most cases we will be interested in how these variables are related to others. But first, let us consider the statistical description of a single random variable, X.

Central to the idea of statistics is the **average** or **expected** value of a random variable. The average of X will here be denoted as X and is defined as

$$X \;=\; \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_n \tag{1}$$

where X_n is the value of X in the n^{th} realization. Throughout these notes the symbol $\langle\cdot\rangle$ will be reserved for the ideal average requiring an infinite number of realization. It is helpful to think of $\langle\cdot\rangle$ as a linear operator which can be applied to random variables. It is a linear operator because

$$\langle X+Y\rangle = \langle Y+X\rangle = \langle X\rangle + \langle Y\rangle, \quad \langle aX\rangle = \langle Xa\rangle = a\langle X\rangle, \quad \langle XY\rangle = \langle YX\rangle \neq \langle X\rangle\langle Y\rangle \tag{2}$$

when a is a constant.

The complete description of a single random variable is its **distribution function** D or, equivalently, its **probability density function** (pdf) F. These are defined by

$$D_X(r) \;=\; \text{Fraction of occurrences with } \; X < r \tag{3}$$

$$F_X(r) \;=\; \frac{d}{dr} D_X(r) \quad \text{so that} \quad F_X(r)\, dr \;=\; \text{Fraction of occurrences with } \; r < X < r + dr. \tag{4}$$

In the deterministic case that every realization produces the constant $X = A$, then $F_X(r) = \delta(r - A)$ and D would be a Heaviside unit step function with its edge at $r = A$. In these r is a dummy variable and X names the process producing the variable described by D and F. Note that $D(-\infty) = 0$, $D(\infty) = \int_{-\infty}^{\infty} dr\; F(r) \;=\; 1$. The pdf is the differential limit of a histogram describing the number of occurrences in each of several "bins" of X.

A handy representation of the probability density function is

$$F_X(r) = \langle \delta(r - X)\rangle. \tag{5}$$

To see why this is so, let N be an effectively infinite number of realizations of X and M be the number of realizations with $X < r$. Then, by definition,

$$D_X(r) \;=\; \frac{M}{N} \;=\; \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{r} dy\; \delta(y - X_n) \;=\; \int_{-\infty}^{r} dy \langle \delta(y - X)\rangle \tag{6}$$

from which (5) follows by differentiation with respect to r. The delta function is introduced into the third term to generate 1 when $X_n > r$ and 0 when it is not. The pdf is the complete description of X (taken

alone) because it immediately gives the average of any function, say G, of the random variable. This is easy to see using (5):

$$\langle G(X)\rangle = \langle \int dr\, G(r)\delta(r-X)\rangle = \int drG(r)\langle\delta(r-X)\rangle = \int dr\, G(r)\, F_X(r) \tag{7}$$

Given the pdf of a variable X it is a simple matter to find the pdf of any other variable which is a function of X, say $Y = Q(X)$. The number of realizations with $x < X < x+dx$ is the same as the number with Y between $Q(x)$ and $Q(x+dx) \;=\; Q(x) + \frac{dQ}{dx}dx$. Therefore

$$F_X(x)\;|dx| \;=\; F_Y[Q(x)]\;|dQ| \tag{8}$$

When the mapping of X to Y is one-to-one this leads to the simple relation

$$F_X(x)\;|dx| = \frac{dQ}{dx}\;F_Y[Q(x)]\cdot|dx| \tag{9}$$

The absolute value signs take care of cases where Q decreases as x increases. When Q vs. x is not one-to-one a form similar to (9) results but must include all contributions of dx that map into the same dQ and vice versa. The signs in (9) are easy to keep straight since F_X and F_Y must both be positive.

The pdf contains more information than can usually be determined for real processes. Consequently, practical analysis often involves simpler statistical measures. The simplest is, of course, the mean X. Others are concerned with variations about the mean and are most conveniently defined in terms of the **fluctuation**

$$X' = X - \langle X\rangle. \tag{10}$$

A prime on a random variable generally denotes a fluctuation. The **variance** μ and the **standard deviation** σ,

$$\mu_2 = \langle X'^2\rangle, \qquad \sigma = \mu_2^{1/2}, \tag{11}$$

describe, respectively, the "energy" of the fluctuations and a typical fluctuation size. The variance should not be confused with the mean square X^2. Beyond the mean and variance we might compute any number of higher **moments**

$$\mu_n = \langle X'^n\rangle. \tag{12}$$

As we will later see, as n increases it becomes progressively harder to obtain an accurate estimate of μ_n from a finite set of data. Thus a few lower moments are all that can typically be determined.

Just as there are random variables there can be random functions, say $X(t)$. This means that there be many realizations of $X(t)$ all running over the same range of t and between which $X(t)$ varies randomly (at the same t). In experimental settings, t might be time since a repeated experiment is started; the experiment is random if it does not evolve the same way in each realization. In an observational setting more creativity may be required to define the analogue of t and the realizations. Sometimes different places might produce different realizations over the same time range - anemometers placed all around an airport. Even when observations of X all come from a single long time series, it is still possible to have a random variable which is a function of time. For example, time series from each year might be regarded as a realization of the seasonal cycle random process.

In any case, if there is an ensemble of time series then averages are defined as for random variables. For example,

$$\langle X(t)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_n^N X_n(t). \tag{13}$$

Since $\langle\cdot\rangle$ is a deterministic linear operator, it passes through deterministic linear operators like integration, differentiation and summing:

$$\frac{d}{dt}\langle X(t)\rangle = \frac{d}{dt}\frac{1}{N}\sum_n^N X_n(t) = \frac{1}{N}\sum\frac{d}{dt}X_n(t) = \langle\frac{d}{dt}X(t)\rangle \tag{14}$$

It is important to understand how time can be used to generate multiple realizations and still be an independent variable. Further, one should have the same facility with the averaging operator as with integrating or differentiating. If there is conceptual difficulty, it is usually in the definition of the ensemble over which averages are to be taken.

Different types of random processes are distinguished by their pdfs. For example, computers easily generate realizations of a pseudo-random process producing a **uniformly distributed** variable, for example with the pdf

$$F(r) = \begin{bmatrix} 1 & \text{for} \quad 0 \le r \le 1 \\ 0 & \text{for} \quad r < 0 \ \text{ or } \ r > 1 \end{bmatrix} \tag{15}$$

Many physical processes which are approximately the sum of independent components are **normally distributed** or have a **Gaussian** pdf

$$F_X(r) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{(r - \langle X \rangle)^2}{2\sigma^2} \right]. \tag{16}$$

which is determined by the mean $\langle X \rangle$ and variance or mean square fluctuation $\sigma^2 = \langle X'^2 \rangle$. The Central Limit Theorem, presented below, shows that the sum of a very large number of variables will have a normal distribution so long as they are produced by the same random process. Since many physical variables are approximately the sum of many random processes, it is not surprising that many real-world observations are normally distributed. When the average of some variable or function of a variable (*e.g.* the square) is estimated from a large but finite sum its value is random and is well approximated as Gaussian. Other processes, characterized by explosive growth periods (like biological populations) have **lognormal distributions**, that is their logarithms are normally distributed. Normal and lognormal distributions are relatively simple because they depend on only two statistics, the mean and variance. More typically, more than two parameters are required to specify the statistics of a random variable.

The uniform, normal and lognormal pdfs corresponding to unit mean and standard deviation are plotted in Figure 2. One can say that the uniform distribution, and to a slightly lesser extent the normal distribution, are central compared with the lognormal distribution which admits a relatively large number of extreme position positive values. Note that only in the normal distribution is the most probable value (the maximum of the pdf) equal to the mean.

The method of generating the lognormal distribution in Figure 2 serves as a good example of transforming variables in a pdf. Let X be the lognormal variable whose logarithm is $Y = \ln X$ which has the normal pdf $F_Y(s)$ in (16) with mean $m = \ln X$ and standard deviation $\sigma = \langle [\ln X - \langle \ln X \rangle]^2 \rangle^{1/2}$. According to (9)

$$F_X(r)\,|dr| = F_Y(s = \ln r)\ [|ds| = |dr|\,/r] \tag{17}$$

$$F_X(r) = \frac{1}{\sqrt{2\pi}\sigma r} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2} \right]. \tag{18}$$

2. Statistics for multiple variables

All of dynamics, and most of what we want to know about the ocean involves how one variable is related to another. We now turn to ways that these relations may be defined and measured when neither the variables nor the relation is deterministic.

The complete description of a pair of random variables X and Y is the **joint probability density** $F_{XY}(r, s)$. As in (5) and (7)

$$F_{XY}(r,s) = \langle \delta(r - X)\delta(s - Y) \rangle, \qquad \langle G(X,Y) \rangle = \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds\, G(r,s)\, F_{XY}(r,s) \tag{19}$$

Since $F_X(r)$ is the pdf of X, without regard for the value of Y

$$F_X(r) = \int_{-\infty}^{\infty} ds\, F_{XY}(r,s)\,. \tag{20}$$

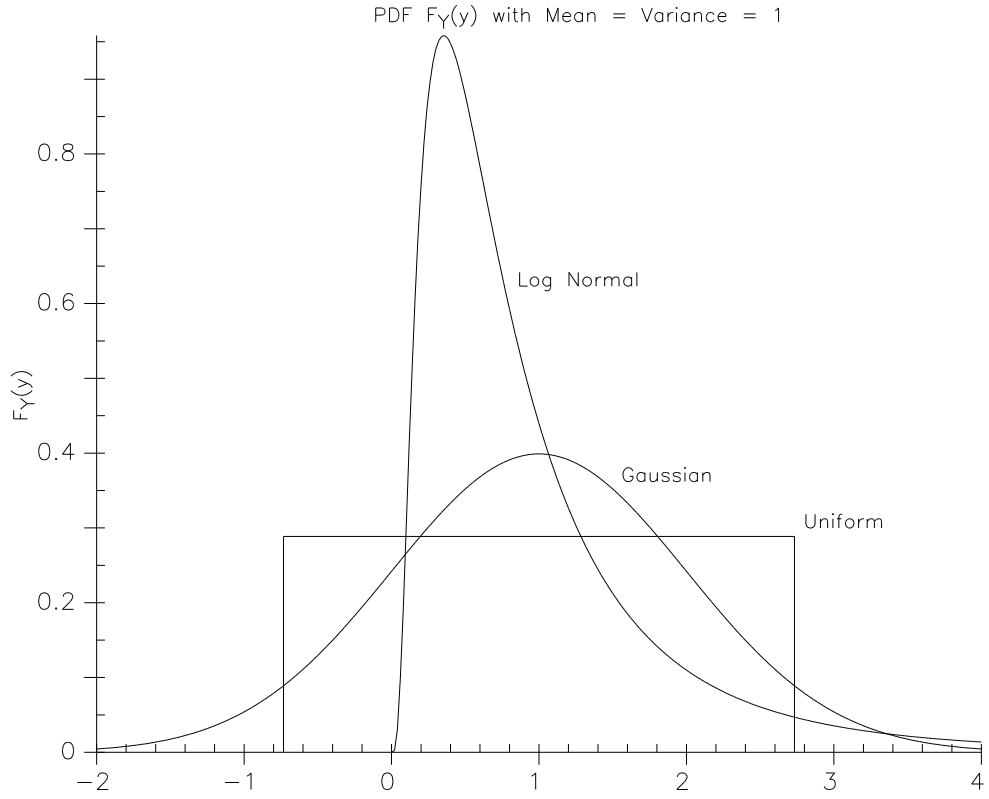


Figure 1: Examples of probability density functions with unit mean and variance.

It is straightforward to convert the jpdf of variables X, Y to the jpdf of functions of X, Y given their functional relation. To do this the procedure leading to (9) is applied making substitutions for both original variables. A special case of some importance is finding the pdf of $A = X + Y$ given the joint pdf of X and Y:

$$\langle G(A) \rangle = \int dx \int dy\, G(x+y) \cdot F_{XY}(x,y) = \int da\, G(a) \int dy\, F_{XY}(a-y,y) \tag{21}$$

so that $F_{X+Y}(a) = \int dy\, F_{XY}(a-y,y)$.

The pdf is simply the density of occurrences in parameter space. For example, if some tracer consists of molecules whose positions are X, Y then $F_{XY}(x,y)$ is the normalized (to unit integral) concentration of molecules at x, y. From the definition of a pdf, $F_{XY}(x,y)\, dx\, dy$ is the probability of finding any particular molecule in the small area $dx\, dy$ around x, y. If C is the mean tracer concentration, then $C(x,y)\, dx\, dy$ is the average amount of tracer found in $dx\, dy$ so that $C(x,y) = M \cdot F_{XY}(x,y)$ where M is the total mass of tracer.

Recall the figures in the previous section describing the variables X and Z of the Lorenz model. The plot of X vs. Z showed that the system usually travels around the edges of the "butterfly wings". Thus you would not be surprised to learn that the two-dimensional pdf is peaked along these rings and is small at the center of the "wings". Having this pdf gives some skill in predicting one variable given the other, even though there is no deterministic relation between the two.

An important conceptual point is the relation between the joint probability density function $F_{XY}(r,s)$ and the **conditional probability density** of X, denoted $F_X(r\,|Y=s)$. This conditional probability is the density of X occurrences near the value r given that $Y = s$. $F_X(r\,|Y=s)\, dr$ is the fraction of occurrences with $Y = s$ in the range $r < X < r + dr$. To see how the conditional pdf is related to other pdfs, suppose we have a nearly infinite number, N_0, of observed X, Y pairs and that the X, Y space is divided into infinitesimal bins of size $dr \times ds$. The number of occurrences with $r < X < r + dr$

and $s < Y < s + ds$ is $N_0 F_{XY}(r,s)\ dr\ ds$. The number of occurrences with $s < Y < s + ds$ is $N_0 ds \int_{-\infty}^{\infty} dr\ F_{XY}(r,s) = N_0 ds\ F_Y(s)$. $F_X(r\,|Y)\ dr$ is the fraction of those occurrences also having $r < X < r + dr$, *i.e.*

$$F_X(r\,|Y)\ dr = \frac{N_0 F_{XY}(r, s = Y)\ dr\ ds}{N_0 F_Y(s)\ ds}\ , \qquad \text{or} \qquad F_X(r\,|Y = s) = \frac{F_{XY}(r, s = Y)}{F_Y(s)} \tag{22}$$

The simplest joint distribution occurs when X and Y are **independent**, by which we mean that the statistics of X do not depend on the value of Y and *vice versa*. This is logically equivalent to saying that the conditional density $F_X(r\,|Y = s)$ is the same for all s and is, in fact, equal to the single-variable pdf $F_X(r)$. Combining this with (22) gives $F_X(r) = F_{XY}(r,s)/F_y(s)$ so that

$$F_{XY}(r,s) = F_X(r) \cdot F_Y(s) \tag{23}$$

The averages of the two delta functions in (19) separate because variations of X and Y are unrelated so that, for example, $\delta(r - X)$ has the same value regardless of the value of Y.

Another description of the pdf can be gained by examining how the average of a random function $G(X,Y)$ might be evaluated. The definition of G simply involves summing the value of G from a large number, N_0, of realizations and dividing the sum by N_0. An alternative would be to count the number of realizations in each of many small "bins" of size $\Delta \times \Delta$ in X, Y space. By small we mean that changing X and Y by $O(\Delta)$ does not change $G(X,Y)$ significantly. Now let $N(r,s)$ be the number of realizations from the total of N_0 which fall in $r\Delta < X < (r+1)\Delta$ and $s\Delta < Y < (s+1)\Delta$. Then

$$\langle G(X,Y)\rangle = \frac{1}{N_0}\sum_{n}^{N_0} G(X_n, Y_n)\ =\ \frac{1}{N_0}\sum_r\ \sum_s\ N(r,s)\cdot G(r\Delta, s\Delta). \tag{24}$$

$N(r,s)/N_0$ is the fraction of observations in bin r,s. According to (4), $N(r,s)/N_0 = \Delta^2\ F_{XY}(r\Delta, s\Delta)$. Taking the $N_0 \to \infty$ and $\Delta \to 0$ limits gives converts (24) to (19).

Just as many single variables in nature are normally distributed, many sets of variables are **joint normally distributed**. If the variables X_1, X_2 ... are independent and normally distributed with $X_n^2 = 1$, then any collection of variables Y_1, Y_2... which are linear combinations of the X_n, i.e.

$$Y_n\ =\ \sum_k\ A_{nk}\ X_k\ , \tag{25}$$

are joint normally distributed. Joint normal variables are completely determined by their means Y_n and their covariances

$$C_{nm} = \langle Y_n' Y_m'\rangle = \sum_k\ A_{nk} A_{mk} \langle {X'}_k^2\rangle \tag{26}$$

The joint pdf of M such variables Y_n is

$$F_{Y_1 Y_2 ....}(y_1, y_2 ...) = (2\pi)^{-M/2}\ |C|^{-1/2}\ \ \exp\left[\frac{-1}{2}\sum_{n,m}(y_n - \langle Y_n\rangle)\ C_{nm}^{-1}\ (y_m - \langle Y_m\rangle)\right] \tag{27}$$

where C^{-1} is the inverse of the covariance matrix and $|C|$ is the determinant of C.

The joint normal distribution (27) is formidable looking but easy to derive since the Y_n are simple functions of the X_k. To do this we will specify that $\langle X_k^2\rangle = 1$. As in arriving at (9), note that a given class of events has the same probability of occurring whether we define that class using X or Y coordinates. Thus

$$F_{Y_1, Y_2 .....}(y_1, y_{2,} ... y_N) dy_1 \cdot dy_2 \cdot .... \cdot dy_N\ \ =\ \ F_{X_1, X_2 .....}(x_1, x_{2,} ... x_N) dx_1 \cdot dx_2 \cdot .... \cdot dx_N \tag{28}$$

when the volume $dy_1 \cdot dy_2 \cdot ....dy_N$ corresponds to the volume $dx_1 \cdot dx_2 \cdot .... \cdot dx_N$. Because the X_k are independent and normal, their jpdf factors to

$$F_{X_1, X_2 .....}(x_1, x_{2,} ... x_N)\ \ =\ \ [2\pi]^{-N/2} \exp(-x_1^2/2) \cdot \exp(-x_2^2/2) \cdot ... \cdot \exp(-x_N^2/2) \tag{29}$$

Substituting $dx_1 \cdot dx_2 \cdot .... \cdot dx_N \; = \; \left|A^{-1}\right| \; dy_1 \cdot dy_2 \cdot .... \cdot dy_N$ and $x_k^2 \; = \; A_{ni}^{-1} A_{nj}^{-1} y_i y_j \; = C_{ij}^{-1} y_i y_j$ (sum over repeated indices) gives (27).

A joint normal distribution is often a reasonable model for physical variables that result from adding many random variables. The remarkable feature of this distribution is that the complete statistical relation between N variables is determined by the N mean values and the $N \cdot (N+1)/2$ associated variances and covariances. Thus all higher moments are detemined by first and second moments. For most other distributions a large or infinite number of moments and/or other statistics are required to define the jpdf.

A very handy result from the general joint-normal distribution is that if the variables $a, \; b, \; , c \; , ...$ are joint-normally distributed with zero mean then the mean product of 2N terms is the sum of all combinations of covariances. Thus

$$\langle abcd \rangle = \langle ab \rangle \langle cd \rangle + \langle ac \rangle \langle bd \rangle + \langle ad \rangle \langle bc \rangle \tag{30}$$

and

$$\begin{aligned} \langle abcdef \rangle \quad = \quad & \langle ab \rangle \cdot [\langle cdef \rangle = \langle cd \rangle \langle ef \rangle + \langle ce \rangle \langle df \rangle + \langle cf \rangle \langle de \rangle] + & (31) \\ & \langle ac \rangle \cdot \langle bdef \rangle + \langle ad \rangle \cdot \langle bcef \rangle + \langle ae \rangle \cdot \langle bcdf \rangle + \langle af \rangle \cdot \langle bcde \rangle & (32) \end{aligned}$$

Note that $a, \; , b \; , ..$ need not be distinct so that the above relations show that $\langle y^4 \rangle \; = \; 3 \langle y^2 \rangle^2$ and $\langle y^{2N} \rangle \; = \; (2N-1) \times (2N-3) \ldots 5 \times 3 \langle y^2 \rangle^N$

It is even more difficult to observe joint pdfs than single-variable pdfs. Consequently, we are generally limited to describing the typical relation between two variables using simpler measures like the **covariance** C_{XY} and the **correlation** ρ_{XY} defined by

$$C_{XY} \; = \langle X' Y' \rangle, \qquad \rho_{XY} = \frac{\langle X' Y' \rangle}{\sqrt{\langle X'^2 \rangle \langle Y'^2 \rangle}} \tag{33}$$

The correlation (limited to $-1 \; \leq \; \rho \; \leq \; 1$) is simply a normalized covariance.

The most important feature of the correlation is that its square ρ_{XY}^2 is the fraction of X or Y variance which can be determined by a linear relation or statistical model based on the other variable. The general linear model for Y' based on X' is

$$\hat{Y}' = \alpha X' \tag{34}$$

where α is a constant chosen to make $\hat{Y}'$ approximate Y'. The mean square of the error $\hat{Y}' - Y'$ is, by substitution of (34) for $\hat{Y}'$

$$E = \langle (\hat{Y}' - Y')^2 \rangle = \alpha^2 \langle X'^2 \rangle - 2\alpha \langle X' Y' \rangle + \langle Y'^2 \rangle. \tag{35}$$

E is a positive semi-definite quadratic form with a single minimum where $\partial E / \partial \alpha = 0$. The α producing this optimum (in the mean square error sense) model is easily found and the associated minimum mean square error calculated:

$$\alpha = \frac{\langle X' Y' \rangle}{\langle X'^2 \rangle} \; , \qquad \langle (\hat{Y}' - Y')^2 \rangle = \langle Y'^2 \rangle (1 - \rho_{XY}^2) \; . \tag{36}$$

It is from the second of (36) that we say ρ^2 is the fraction of variance which can be explained by a linear model.

It is often quite useful to note that if X and Y are correlated variables then there are linear combinations of these two variables which are not. Suppose that X and Y were the velocities in the east and north directions from a position very near a coast with the orientation θ. One might prefer to think of these observations in terms of the alongshore and across-shore components

$$\hat{X} = X \cdot \cos(\theta) + Y \cdot \sin(\theta) \; , \qquad \hat{Y} = Y \cdot \cos(\theta) - X \cdot \sin(\theta) \; , \tag{37}$$

One would not be surprised if the $\hat{X}$ variability were much larger than $\hat{Y}$ variations and that they arose from quite different causes, so that $\langle \hat{X}' \hat{Y}' \rangle = 0$. Rather than guessing which vector components are uncorrelated,

one can calculate the **principal axes** in which the components are, by definition, uncorrelated. It is a simple matter to show that the principal axes are defined by the transformation (37) with

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{2\langle X'Y'\rangle}{\langle X'^2\rangle - \langle Y'^2\rangle}\right] \tag{38}$$

The process of finding principal axes for a collection of random variables essentially amounts to inverting the relation (25) from which correlated variables Y_n are generated by uncorrelated X_ks. In general the Y_n are not joint normally distributed so that the X may be neither normally distributed nor independent but they will always be uncorrelated.

3. The Central Limit Theorem

Because of the tremendous amount of information in a probability density functions, they usually can not be determined accurately from observations. In one special case, however, the pdf can be determined very directly from the Central Limit Theorem. This case is both very important in describing many physical processes and in determining how accurately statistical constructs made from finite amounts of data approximate statistics. The proof of the Central Limit Theorem is marvelously powerful and its outline serves to define very clearly the conditions under which the Theorem applies.

Consider the normalized sum

$$X \;=\; N^{-1/2}\sum^{N} V_n \tag{39}$$

where the V_n are **identically distributed**, independent random variables with

$$\langle V\rangle = 0, \qquad \langle V^2\rangle = \sigma^2, \qquad \lim_{N\to\infty}\langle (V/\sigma)^n\rangle N^{-1/2} = 0 \qquad \text{for all } n \tag{40}$$

The third of these restrictions is crucial and yet quite reasonable. The only restriction on the probability density of V is that the high moments of V/σ can all be bounded by $N^{1/2}$, that is to say, the pdf of V is well behaved at large values of V. The first two conditions insure that

$$\langle X\rangle = 0, \quad \langle X^2\rangle = \frac{1}{N}\sum^{N}\sum^{N}\langle V_n V_m\rangle = \frac{1}{N}\sum^{N}\langle V^2\rangle = \sigma^2 \tag{41}$$

The ingenious feature of the proof involves using the **characteristic function** of X which is

$$\begin{aligned}\langle\exp(isX)\rangle &= \langle\exp(isV_1N^{-1/2})\cdot\exp(isV_2N^{-1/2})\ .....\rangle \\ &= \langle\exp(isV_1N^{-1/2})\rangle\langle\exp(isV_2N^{-1/2})\rangle....\langle\exp(isV_NN^{-1/2})\rangle \\ &= \langle\exp(isVN^{-1/2})\rangle^N\end{aligned} \tag{42}$$

The second equality follows from the fact that the V_n are independent so that the average product is the product of the averages. The third follows from making all V_ns identically distributed. The Central Limit Theorem applies to the case where $N\to\infty$ in which case we can make use of a Taylor series expansion of the exponential function

$$\begin{aligned}\langle\exp(isVN^{-1/2})\rangle &= \langle 1 + isVN^{-1/2} - \frac{1}{2}s^2V^2N^{-1} + \ O(s^3V^3N^{-3/2})\rangle \\ &= 1 - \frac{1}{2}s^2\sigma^2\frac{1}{N} + O(N^{-3/2})\end{aligned} \tag{43}$$

It is in saying that the remainder in this expression is of $O(N^{-3/2})$ that we have made use of the third restriction in (40). Now using the binomial expansion

$$\langle\exp(isX)\rangle \;=\; \left[1 \;-\; \frac{1}{2}s^2\sigma^2\frac{1}{N} + O(N^{-3/2})\right]^N$$

$$
\begin{aligned}
&= \quad 1 \; - \; N(\frac{1}{2}s^2\sigma^2\frac{1}{N}) \; + \; \frac{N(N-1)}{2!}(\frac{1}{2}s^2\sigma^2\frac{1}{N})^2 \; + \; ..... \\
&= \quad 1 - \left(\frac{1}{2}s^2\sigma^2\right) + \frac{1}{2!}\left(\frac{1}{2}s^2\sigma^2\right)^2 \; - \; \frac{1}{3!}\left(\frac{1}{2}s^2\sigma^2\right)^3 + ... + O(N^{-1/2}) \\
&= \quad \exp\left(-\frac{1}{2}s^2\sigma^2\right) + \; O(N^{-1/2}) \qquad (44)
\end{aligned}
$$

In the limit $N \to \infty$ we may neglect the remainder. Then making note of (7) and carrying out an inverse Fourier transform

$$\langle \exp(isX) \rangle = \int dx \; F_X(x) \; \exp(isx) \; = \; \exp\left(\frac{-1}{2}s^2\sigma^2\right) \qquad (45)$$

$$F_X(x) \; = \; \frac{1}{2\pi}\int ds \; \exp(-isx) \; \exp(\frac{-1}{2}s^2\sigma^2) \; = \; (2\pi)^{-1/2} \; \frac{1}{\sigma} \; \exp(-x^2/2\sigma^2) \qquad (46)$$

Thus by saying only that X is the sum of an infinite number of almost arbitrarily distributed random variables it is possible to specify the pdf of X. In general $\langle X \rangle \neq 0$ and (16) applies.

How many terms are required to approximate the "infinite number" of the Central Limit Theorem? The answer depends on the distribution of the things to be added. If they have very dispersed distributions then it will take many, if they already look somewhat like normal variables it will take few. The depends even more strongly on how you measure similarity to the normal distribution. It is easy to make the occurrence of likely outcomes follow the normal curve but it is hard to get sums of few variables to have the right occurrence of rare events. An example is shown in Figure **??** which shows the probability density of variables X_N defined as

$$X_N \; = \; \left[\frac{12}{N}\right]^{1/2} \sum_{n=1}^{N} V_n \qquad (47)$$

where the distribution of V_n is uniform over $-1/2 < V < 1/2$. The X_N have zero mean and unit variance and X_∞ is, of course, normally distributed. The distributions of X_3 and X_{10} are shown. Table 1 also lists the moments $\langle X^4 \rangle$ and $\langle X^6 \rangle$ of these variables. Both sums do a fair job of having the right fourth moment but even X_{10} misses the sixth moment badly. The problem is, of course, that a normal variable can take on any value (albeit some are infrequent) whereas the finite sums have very distinct maximum and minimum values.

4. Numerical generation of random numbers

While the distribution of functions of random variables can theoretically be computed analytically from pdfs, it is often desirable to simulate a random process starting from numerically generated random numbers with a known distribution. Like all "random" processes, such numbers are really deterministic if the algorithm by which they are generated is known, but we can call them random, or at pseudo-random.

Most computer systems support a random number generator based on the "Congruential" method of producing numbers X with a uniform distribution over $0 < X < 1$. The basic idea is that the insignificant digits in a sequence of numbers will (a) probably not have a pattern regardless of how the numbers are generated and (b) will be uniformly distributed from 0 to the largest number the selected number of digits can generate. If MOD(I, m) means the remainder when I is divided by m, then the basic Congruential method generates the next random integer I from the last according to

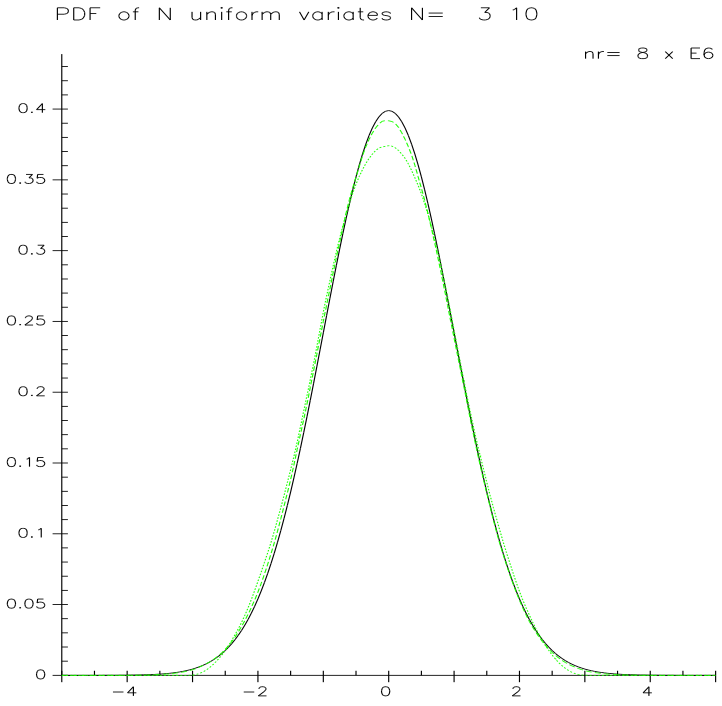
$$I_{n+1} \; = \; \mathrm{MOD}(aI_n + b, \; m) \, , \qquad\qquad X_n \; = \; \frac{1}{m} I_n \qquad (48)$$

where $a, \; b$ and m are large integers. The resulting X is uniform over $0 < X < 1$.

It is also useful to generate normally distributed random variables. This is easy if you can generate uniformly distributed numbers. The brute force approach would be sum a large number of the uniform variates. By the Central Limit Theorem the sum will be normally distributed; it is surprising how few numbers need be summed to get fairly accurate results but the extreme values will not be right until a large number is used.

N	X^2	X^4	X^6
3	0.9988	2.591	9.72
10	0.9998	2.870	13.16
∞	1.0	3.0	15.0

Table 1: Table of PDF moments corresponding to Figure 2.



t

Figure 2: Probability Density Function of the sum of 3 and 10 unifrom variates (gray, scaled so the mean of the sums is zero and they have unit standard deviations) compared with Gaussian distributions. Table 1 shows moments.

A more efficient, but less obvious, method is to generate two variates U and V to be uniform on 0,1 and 0,π, respectively. Then

$$X = (-2\ln U\ )^{1/2}\cos(V) \tag{49}$$

is normally distributed with $\langle X\rangle = 0$ and $\langle X^2\rangle = 1$. This can be proven by noting that U and V are independent so that

$$\langle X^n\rangle = \langle(-2\ln U)^{n/2}\rangle\langle\cos^n(V)\rangle \tag{50}$$

Clearly the odd moments vanish. The even moments are found as

$$\langle(-2\ ln\ U)^n\rangle = \int_0^1 du\ (-2\ ln\ u)^n\ = \int_0^\infty dw\ e^{-w}(2w)^n\ =\ 2\times 4\times 6...2n \tag{51}$$

$$\langle\cos^{2n}(V)\rangle = \frac{1}{\pi}\int_0^\pi\ dv\cos^{2n}(v) = \frac{1\times 3\times 5....(2n-1)}{2\times 4\times 6....2n} \tag{52}$$

Thus the n^{th} moment of X is $1\times 3\times 5...(2n-1)$. The even moments of a normally distributed variable with zero mean and unit variance are

$$\begin{aligned}\langle Y^{2n}\rangle &= (2\pi)^{-1/2}\left[\int_{-\infty}^{\infty} y^{2n-1}\ y\ \exp\left(-\frac{1}{2}y^2\right)dy\right]\\ &= (2n-1)\int_{-\infty}^{\infty} y^{2(n-1)}\exp\left(-\frac{1}{2}y^2\right)dy\ =\ (2n-1)\langle Y^{2n-2}\rangle\end{aligned} \tag{53}$$

This is the same as the moment of X. Since the moments are uniquely determined by the pdf, the moments determine the pdf and, therefore, Y must also be normally distributed.

Problems

(1) Define the lognormal variable X such that $Y = \ln(X)$ is normally distributed. Write down $F_Y(y)$ in terms of $\langle Y\rangle$ and σ_Y. Then find the mean and variance of X.

(2) If X is normally distributed with zero mean and $Y = X^2$ find $F_Y(y)$.

(3) Write down a general expression for the pdf of $Z = X\cdot Y$ in terms of $F_{XY}(x,y)$.

(4) Given independent normally distributed variables X and Y with zero mean and unit variance, show the simplest way to generate normally distributed variables A and B with arbitrary means and specified correlation ρ_{AB}.

(5) From the general multi-variate joint normal distribution (27), find the pdf of the two joint-normal variables X and Y in terms of their standard deviations and their correlation.

(6) The sum of two normally distributed variables is also normally distributed. Thus the low moments of the sum should be related as above (30). Let $Z\ =\ X\ +\ Y$ where X and Y are independent normally distributed variables. What is the third moment of Z in terms of statistics of X and $\langle Y\rangle$? Show that the fourth moment of $Z = X + Y$ obeys (30) when $\langle X\rangle = \langle Y\rangle = 0$ but both have different variances.