Let's start with basics of random variables and then extend things to random fields. A random variable \( x \) has probability distribution function (PDF) \( p(x) \), so that

\[
\int p(x) \, dx = 1
\]

where the integral is over all possible values of \( x \).

Sometimes it is more convenient to characterize this function \( p(x) \) by a few numbers which tell us basic features of the PDF. One example are the moments of a distribution,

\[
\langle x \rangle \equiv \int p(x) \, x \, dx \quad \text{"mean"} \quad \text{tells about which value of } x \text{ the random variable is fluctuating around}
\]

\[
\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2 \equiv \int p(x) \, x^2 \, dx - \langle x \rangle^2 \quad \text{"variance"} \quad \text{tells how much variable } x \text{ fluctuates around } \langle x \rangle
\]

The typical example of a distribution (the only one) that can be characterized by its first 2 moments (more on this later) is the Gaussian PDF:

\[
p_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}
\]

\[
\Rightarrow \langle x \rangle = a
\]

\[
\Rightarrow \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2
\]

We can always make a random variable have zero mean by doing

\[
y = x - \langle x \rangle
\]

\[
\Rightarrow \langle y \rangle = 0
\]

and \( \sigma^2 = \langle y^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \)
Since this simplifies the appearance of many expressions, from now on we work with zero mean random variables.

The two next moments are also useful— for $n=3$ we have

$$\langle x^3 \rangle = \int p(x) x^3 \, dx$$

$$\Rightarrow s_3 \equiv \frac{\langle x^3 \rangle}{\langle x^2 \rangle^{3/2}} = \frac{\langle x^3 \rangle}{\sigma^3} \quad \text{"Skewness"}$$

characterizes how asymmetric between $x$ and $-x$ is the PDF.

For the fourth moment, we have

$$\langle x^4 \rangle = \int p(x) x^4 \, dx$$

In a Gaussian distribution, as we saw in homework #4, $\langle x^4 \rangle = 3 \langle x^2 \rangle^2$

It is useful to define a quantity known as "kurtosis", which tells us about how important are tails of PDF relative to a Gaussian:

$$s_4 \equiv \frac{\langle x^4 \rangle - 3 \langle x^2 \rangle^2}{\langle x^2 \rangle^2} = \frac{\langle x^4 \rangle}{\sigma^4} - 3 \quad \text{"Kurtosis"}$$

(some times called "mean kurtosis")

$s_4 > 0$ (tails enhanced)

$s_4 < 0$ (tails suppressed)
All the moments can be generated from the moment generating function \( M(t) \):

\[
M(t) = \langle e^{tX} \rangle = \int e^{tx} p(x) \, dx
\]

\[
\Rightarrow M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \langle x^n \rangle
\]

\[
\Rightarrow \langle x^n \rangle = \left( \frac{d^n M(t)}{dt^n} \right)_{t=0}
\]

A similar definition (which has nice convergence properties) is the characteristic function

\[
\phi(t) = M(it) = \int e^{itx} p(x) \, dx = \langle e^{itX} \rangle
\]

This is basically, up to a constant, the Fourier transform of the PDF.

Again we have:

\[
\left( \frac{d^n \phi(t)}{dt^n} \right)_{t=0} = i^n \langle x^n \rangle
\]

When integrals involved exist, one can invert this relationship and recover PDF from its characteristic function,

\[
p(x) = \int \phi(t) e^{itx} \frac{dt}{2\pi}
\]

It is in this sense that, knowing all moments and thus the characteristic function, one can recover the PDF. This actually fails sometimes (notably, for the lognormal distribution).

For a Gaussian distribution, we have (assuming zero mean):

\[
M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \langle x^n \rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \frac{t^2 \langle x^2 \rangle}{2} \right]^n = e^{t^2 \langle x^2 \rangle / 2}
\]
Then:
\[
\begin{align*}
M(t) &= e^{\frac{1}{2}t^2\langle x^2 \rangle} \\
\phi(t) &= e^{\frac{-1}{2}t^2\langle x^2 \rangle}
\end{align*}
\]

The fact that in a Gaussian everything is determined by its second moment is obvious from \(M(t)\) and \(\phi(t)\). This motivates the definition of a cumulant generating function \(\zeta(t)\):

\[
\zeta(t) = \ln[M(t)]
\]

In the case of the Gaussian, \(\zeta(t) = \frac{1}{2}t^2\langle x^2 \rangle\)

Derivatives of \(\zeta(t)\) generate the cumulants, same way we did for moments:

\[
\langle x^n \rangle_c = \left[ \frac{d^n \zeta(t)}{dt^n} \right]_{t=0}
\]

In general case the cumulants are related to moments as follows:

\[
\begin{align*}
\langle x \rangle_c &= \langle x \rangle \\
\langle x^2 \rangle_c &= \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \\
\langle x^3 \rangle_c &= \langle x^3 \rangle - 3 \langle x \rangle \langle x^2 \rangle + 2\langle x \rangle^3 = \langle x^3 \rangle - 3 \langle x \rangle \langle x^2 \rangle_c - \langle x^3 \rangle_c \\
\langle x^4 \rangle_c &= \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x \rangle^2 \langle x^2 \rangle - 6 \langle x^4 \rangle_c = \langle x^4 \rangle - 4 \langle x^3 \rangle_c \langle x \rangle_c - 3 \langle x^2 \rangle_c^2
\end{align*}
\]

Or for the zero mean case:

\[
\begin{align*}
\langle x^2 \rangle_c &= \sigma^2 \\
\langle x^3 \rangle_c &= \langle x^3 \rangle \\
\langle x^4 \rangle_c &= \langle x^4 \rangle - 3 \langle x^2 \rangle^2
\end{align*}
\]

So, kurtosis can be written as:

\[
\text{Kurtosis} = \frac{\langle x^4 \rangle_c}{\sigma^4}
\]

For a Gaussian random variable with zero mean, only cumulant non-zero is the second, \(\langle x^m \rangle_c = 0 \text{ for } n>2\)
The subscript "c" in \( \langle x \rangle_c \) stands for "cumulant", and also for connected in the language of statistical physics since when we consider random fields, these cumulant averages can be thought of in terms of diagrams that are always connecting all the points in an \( N \)-point correlation function.

An important property of cumulants is that, unlike moments, they are independent of each other, thus they are a much more compressed representation of the same information.

The PDF can be written in terms of the cumulant generating function,

\[
P(x) = \int \phi(t) e^{-itx} \frac{dt}{2\pi} = \int e^{-itx} \frac{dt}{2\pi}
\]

We saw before that \( S_3 = \frac{\langle x^3 \rangle_c}{\sigma^3} \), \( S_4 = \frac{\langle x^4 \rangle_c}{\sigma^4} \), and similarly we can define the \( S_n \) parameters by:

\[
S_n = \frac{\langle x^n \rangle_c}{\sigma^n}
\]

(\( \sigma^2 = \langle x^2 \rangle_c \))

Then we can see that:

\[
C_i(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{\langle x^n \rangle_c}{\sigma^n} = \sum_{n=2}^{\infty} \frac{(it)^n}{n!} S_n = -\frac{1}{2} + \sum_{n=3}^{\infty} \frac{(it)^n}{n!} S_n
\]

Cumulant G.F. starts from \( n=1 \), and we assume zero mean, \( \langle x \rangle = 0 \)

We see that if we define \( S(i) = C_i(t) \)

\[
\Rightarrow \quad S_n = \frac{d^n S(i)}{dt^n} \bigg|_{t=0}
\]

Important: Since cumulants obey \( \langle (b \, x)^n \rangle_c = b^n \langle x^n \rangle_c \)

\( \langle (b \, x) \rangle_c = \langle x \rangle_c \)

A linear transformation of a Gaussian variable is also a Gaussian variable.

<table>
<thead>
<tr>
<th>Important: Since cumulants obey: ( \langle (b , x)^n \rangle_c = b^n \langle x^n \rangle_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \langle (b , x) \rangle_c = \langle x \rangle_c )</td>
</tr>
<tr>
<td>A linear transformation of a Gaussian variable is also a Gaussian variable.</td>
</tr>
</tbody>
</table>

That is, \( S(i) \) is the G.F. for the \( S_n \) parameters. We can rewrite PDF using this, therefore,
\[ P(z) = \int e^{\frac{-1}{2\sigma^2} (z - tx)^2} \, dx \]

Change variable to
\[ t' = \frac{t}{\sigma} \]

This will give just Gaussian distribution

Correction to Gaussian is fully characterized by \( s_g \) parameter

For this, we have to calculate the expected fluctuation in \( s_g \), mean estimated in a Gaussian field, that is, e.g. if we measure \( s_g \) by doing

\[ s_g = \frac{1}{N_r} \sum_{i=1}^{N_r} \frac{X_i^3}{\sigma^2} \]

Where a hat means that the RMS is an estimator of \( s_g \) given \( N_r \) realizations of the random variable \( X_i \) (with zero mean). This is how you calculated the skewness in homework #1. Note that \( s_g \) is a random variable, a linear combination of \( X_i \) of random variables

and thus it will fluctuate: if you generate another \( N_r = 1000 \) realizations \( s_g \) will have a different value. To see how much is expected to fluctuate we calculate its dispersion in a Gaussian field (to see whether variations you saw in HW #1 are consistent with \( s_g \) being zero):

\[ \left \langle \hat{s}_g^2 \right \rangle - \left \langle s_g^2 \right \rangle = \left \langle s_g^2 \right \rangle = \frac{1}{6} \frac{1}{N_r^2} \sum_{i \neq j} \left \langle X_i^3 X_j^3 \right \rangle \]

Note that, for simplicity, we consider \( \sigma^2 \) to be an exact parameter determined accurately enough that it has no fluctuation. Now, we decompose the sum as:

\[ \left \langle s_g^2 \right \rangle = \frac{1}{6} \frac{1}{N_r^2} \left \{ \sum_{i \neq j} \left \langle X_i^6 \right \rangle + \sum_{i \neq j} \left \langle X_i^3 X_j^3 \right \rangle \right \} \]
Now, \[ \frac{1}{N_r} \sum_i X_{i,3} = 15 \text{ for Gaussian} \]

\[ \frac{1}{N_r} \sum_i X_{i,5} \leq N_r \]

\[ \frac{1}{N_r} \sum_i X_{i,3} \leq \frac{1}{N_r} \sum_i X_{i,5} = 0 \]

\[ \Rightarrow \langle S_3^2 \rangle = \frac{15}{N_r} \]

\[ \Rightarrow \sqrt{\langle S_3^2 \rangle} = \sqrt{\frac{15}{N_r}} \approx 0.12 \]

This is the level of skewness that you should have seen in HMW#1, improving your LCGs should not change this level of skew you see.

Let's do the same for \( S_4 \):

\[ \hat{S}_4 = \frac{1}{0.4} \frac{1}{N_r} \sum_i X_{i,4}^4 - 3 \]

\[ \langle S_4^2 \rangle = \frac{1}{0.4} \frac{1}{N_r} \langle \sum_i \sum_j X_{i,4} X_{j,4} \rangle - \frac{6}{0.4} \frac{1}{N_r} \left( \langle \sum_i X_{i,4} \rangle \right)^2 + 9 \]

\[ \frac{1}{N_r^2} \langle \sum_i \sum_j X_{i,4} X_{j,4} \rangle = \frac{1}{N_r^2} \langle \sum_i X_{i,8} \rangle + \frac{1}{N_r^2} \langle \sum_i \sum_j X_{i,4} X_{j,4} \rangle \]

\[ \frac{1}{N_r} \sum_i X_{i,8} \]

\[ \text{for Gaussian:} \]

\[ \frac{1}{N_r^2} \sum_i X_{i,4} \sum_j \sum_k X_{j,4} \approx 9 \times 0.8 \]

\[ \Rightarrow \langle S_4^2 \rangle \approx \frac{105}{N_r} + 9 - 18 + 9 \]

\[ \Rightarrow \sqrt{\langle S_4^2 \rangle} = \sqrt{\frac{105}{N_r}} \approx 0.32 \]

Thus, definitely when you saw levels of \( S_4 \) of order unity for LCG, you were seeing flattening beyond expected for Gaussian, so the problem with LCG was evident.
Let's explore some other properties of cumulants before we discuss the Central Limit Theorem.

Suppose we have a sum of random variables:

\[ y = x_1 + \ldots + x_N \]

each of them independent of the others, but not necessarily with the same PDF. Then, the PDF for \( y \) is given by:

\[
P(y) = \int P(x_1, \ldots, x_N) \, \delta \left[ y - \sum_{i=1}^{N} x_i \right] \, dx_1 \ldots dx_N
\]

\[ \uparrow \]

\[ x_i \, \text{indep} \]

We have for the characteristic function of \( y \):

\[
\phi_y(t) = \int e^{i t y} P(y) \, dy = \int e^{i t \sum x_i} P_1(x_1) \ldots P_N(x_N) \, dx_1 \ldots dx_N
\]

\[ \Rightarrow \phi(y) = \phi_1(t) \ldots \phi_N(t) \]

That is, the characteristic function of the sum is the product of the characteristic functions. If the variables have all the same PDF, we have:

\[
\phi(t) = \left[ \phi_x(t) \right]^N
\]

Now, since the cumulant GF is \( C(t) = \ln M(t) \) and \( \phi(t) = M(it) \), we have:

\[
C_y(t) = \sum_{i=1}^{N} C_i(t)
\]

In other words, the cumulant GF of a sum of independent variables is the sum of the cumulant GFs. For equally distributed variables, we have

\[
N C(t)
\]

Now, we are ready to take a look at the Central Limit Theorem...
The central limit theorem says that the sum of uncorrelated variables becomes Gaussian in the limit $N \to \infty$.

We will assume all moments exist to simplify our proof. All we have to prove is that the $S_n$ parameter goes to zero as $N \to \infty$.

Let $y = x_1 + \cdots + x_N$ 

\[ E[y^2] = \sum_{i=1}^{N} E[x_i^2] = \sum_{i=1}^{N} \sigma_i^2 \]

\[ E[y^3] = \sum_{i=1}^{N} E[x_i^3] = \sum_{i=1}^{N} S_3 \sigma_i^3 \frac{1}{\sqrt{3}} \]

\[ |S_2| = \left| \frac{E[y^2]}{E[y^2]^{3/2}} \right| = \left| \frac{\sum_{i=1}^{N} \sigma_i^2}{\left( \sum_{i=1}^{N} \sigma_i^2 \right)^{3/2}} \right| \leq \frac{N S_2^{\text{max}}}{\sigma_{\text{max}}^{3/2}} \frac{1}{\sqrt{N}} \quad N \to \infty
\]

where $S_2^{\text{max}} = \max \{ S_2, \sigma \}$, $\sigma_{\text{max}} = \max \{ \sigma_i \}$, $\sigma_{\text{min}} = \min \{ \sigma_i \}$

Similarly, for $S_4$:

\[ |S_4| = \left| \frac{\sum_{i=1}^{N} \sigma_i^4 S_4}{E[y]^{4/2}} \right| \leq \frac{1}{N} S_4^{\text{max}} \left( \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \right)^4 \to 0 \quad N \to \infty
\]

Thus, in general $S_n \sim \frac{1}{N^{n/2-1}} \to 0 \quad N \to \infty$ ($n \geq 3$)

So, PDF of $y$ becomes Gaussian.

Note: A lognormal distribution is such that $\ln x$ is Gaussian.

The variance$S_2$ will formally diverge as $N \to \infty$.

You can make it finite by replacing $y \to y/N^p$.

With $p > 1/2$, $S_n$'s do not depend on $p$, so arguments

Thus, product of uncorrelated variables becomes lognormal distributed.