# 4   Sums of Random Variables

Many of the variables dealt with in physics can be expressed as a sum of other variables; often the components of the sum are statistically independent. This section deals with determining the behavior of the sum from the properties of the individual components. First, simple averages are used to find the mean and variance of the sum of statistically independent elements. Next, functions of a random variable are used to examine the probability density of the sum of dependent as well as independent elements. Finally, the Central Limit Theorem is introduced and discussed.

Consider a sum $S_n$ of $n$ statistically independent random variables $x_i$. The probability densities for the $n$ individual variables need not be identical.

$$S_n \equiv \sum_{i=1}^{n} x_i$$

$$p_{x_i}(\zeta) \text{ does not necessarily } = p_{x_j}(\zeta) \text{ for } i \neq j$$

$$< x_i > \equiv E_i \qquad\qquad < (x_i - E_i)^2 > \equiv \sigma_i^2$$

The mean value of the sum is the sum of the individual means:

$$< S_n > \;=\; \int (x_1 + x_2 + \cdots + x_n) \underbrace{p(x_1, x_2, \ldots, x_n)}_{p_1(x_1)p_2(x_2)\cdots p_n(x_n)} \, dx_1 \, dx_2 \cdots dx_n$$

$$=\; \sum_{i=1}^{n} [\underbrace{\int x_i p_i(x_i) \, dx_i}_{E_i}][\prod_{i \neq j} \underbrace{\int p_j(x_j) \, dx_j}_{1}]$$

$$=\; \sum_{i=1}^{n} E_i$$

The variance of the sum is the sum of the individual variances:

$$Var(S_n) \;=\; < (S_n - < S_n >)^2 >$$

$$=\; < (x_1 - E_1 + x_2 - E_2 + \cdots + x_n - E_n)^2 >$$

The right hand side of the above expression is a sum of terms, each quadratic in $x$, having the following form:

$$< (x_i - E_i)(x_j - E_j) > \quad = \quad < x_i x_j > - < x_i > E_j - < x_j > E_i + E_i E_j$$

$$= \quad < x_i x_j > - E_i E_j$$

$$= \quad \begin{cases} < x_i^2 > - E_i^2 = \sigma_i^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Therefore,

$$Var(S_n) = \sum_{i=1}^{n} \sigma_i^2.$$

In the special case where all the individual $x_i$'s have the same probability density the above results reduce to

$$< S_n > = n E_i , \qquad\qquad Var(S_n) = n \sigma_i^2 .$$

Physically, the width of $p(S_n)$ grows as $\sqrt{n}$ while the mean of $S_n$ grows as $n$. The probability density becomes more concentrated about the mean as $n$ increases. If $n$ is very large, the distribution develops a sharp narrow peak at the location of the mean.

Now turn to the problem of finding the entire probability density, $p_S(\alpha)$, for the sum of two arbitrary random variables $x$ and $y$ represented by the joint density $p_{x,y}(\zeta, \eta)$. This is a straight forward application of functions of a random variable.
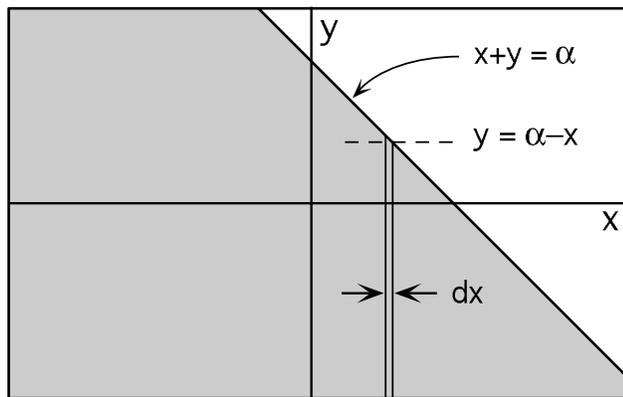
$$P_S(\alpha) = \int_{-\infty}^{\infty} d\zeta \int_{-\infty}^{\alpha-\zeta} d\eta \, p_{x,y}(\zeta, \eta)$$

$$
\begin{aligned}
p_s(\alpha) &= \frac{d}{d\alpha} P_S(\alpha) \\
&= \int_{-\infty}^{\infty} d\zeta \, p_{x,y}(\zeta, \alpha - \zeta)
\end{aligned}
$$

In this expression $p(S)$ is given as a single integral over the joint density $p(x, y)$. The result is valid even when $x$ and $y$ are statistically dependent. The result simplifies somewhat when $x$ and $y$ are statistically independent, allowing the joint density to be factored into the product of two individual densities.

$$p_S(\alpha) = \int_{-\infty}^{\infty} d\zeta \, p_x(\zeta) p_y(\alpha - \zeta) \qquad \text{if } x \text{ and } y \text{ are S.I.}$$

The integral operation involved in the last expression is known as *convolution*. The probability density for the sum of two S.I. random variables is the convolution of the densities of the two individual variables. Convolution appears in other disciplines as well. The transient output of a linear system (such as an electronic circuit) is the convolution of the impulse response of the system and the input pulse shape. The recorded output of a linear spectrometer (such as a grating spectrograph) is the convolution of the instrumental profile and the intrinsic spectrum being measured.

The convolution of two functions $p(x)$ and $q(x)$ is designated by $\otimes$ and is defined as
$$p \otimes q \equiv \int_{-\infty}^{\infty} p(z) q(x - z) dz.$$

It is easy to show from this expression that convolution is commutative, that is, the result does not depend on which function is taken first:

$$a \otimes b = b \otimes a.$$

Convolution is also distributive,

$$a \otimes (b + c) = a \otimes b + a \otimes c,$$

and associative,
$$a \otimes (b \otimes c) = (a \otimes b) \otimes c.$$

Perhaps the best way to visualize the convolution integral is in terms of a series of successive steps.

1. FLIP $q(z)$ ABOUT THE ORIGIN OF ITS ARGUMENT TO FORM THE MIRROR IMAGE $q(-z)$.
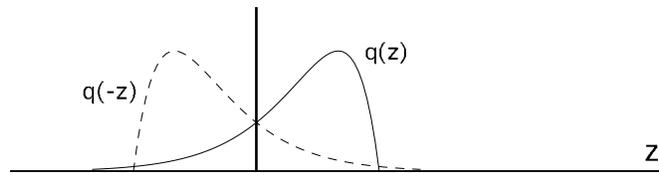


2. SHIFT $q(-z)$ TO THE RIGHT BY AN AMOUNT $x$ TO FORM $q(-z + x)$.
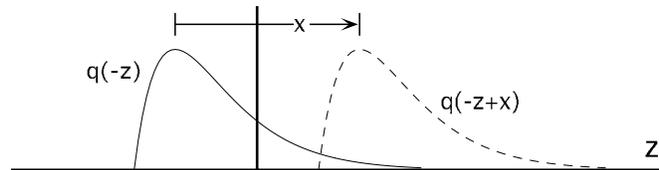


3. MULTIPLY $q(x - z)$ BY $p(z)$.
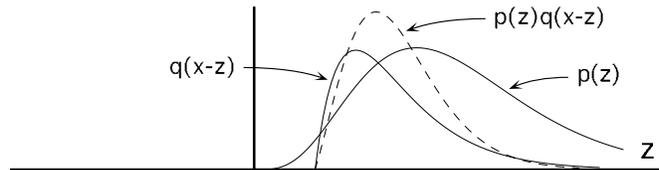


4. INTEGRATE THE PRODUCT AND PLOT THE RESULT.

5. REPEAT 2 THROUGH 4 FOR DIFFERENT VALUES OF $x$.



If $p(z)$ and $q(z)$ are each represented by a single analytic function for all $z$ one can work out the convolution integral mathematically with little thought given to the above graphic procedure. On the other hand, when the input functions are zero in some regions or are only piecewise continuous, then the graphic procedure is useful in setting the proper limits of integration.

**Example:** Sum of Two Uniformly Distributed Variables

Given $x$ and $y$ are two statistically independent random variables, uniformly distributed in the regions $|x| \leq a$ and $|y| \leq b$.
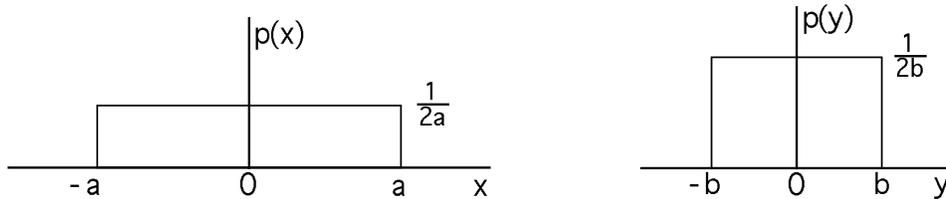


Problem Find the probability density $p(S)$ for the sum $S = x + y$.

Solution The form of the integral will depend on the value of $S$. Three separate regions need to be considered. Assume as indicated above that $b < a$.

$0 < S < a - b$
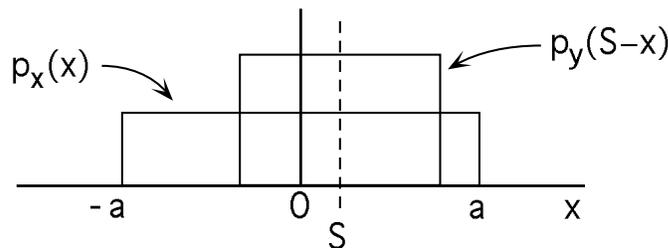
$$p(S) = \int_{S-b}^{S+b} (\frac{1}{2a})(\frac{1}{2b})\, dx = \frac{1}{2a}$$

$a - b < S < a + b$
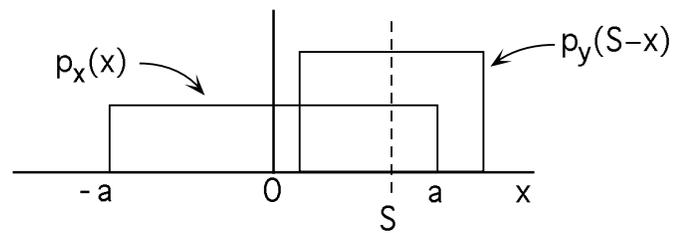


$$p(S) = \int_{S-b}^{a} (\frac{1}{2a})(\frac{1}{2b})\, dx = \frac{a+b-S}{(2a)(2b)}$$

$a + b < S$
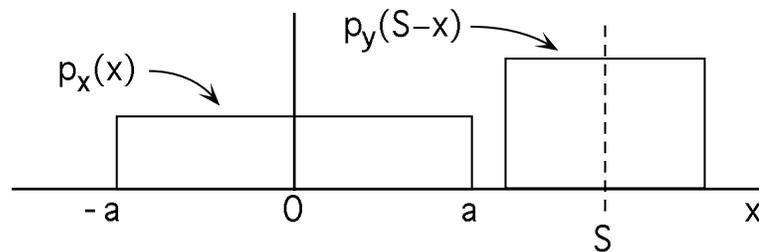


$$p(S) = 0 \qquad \text{since there is no overlap}$$

The symmetry of the problem demands that the result be even in $S$. The final result is plotted below.
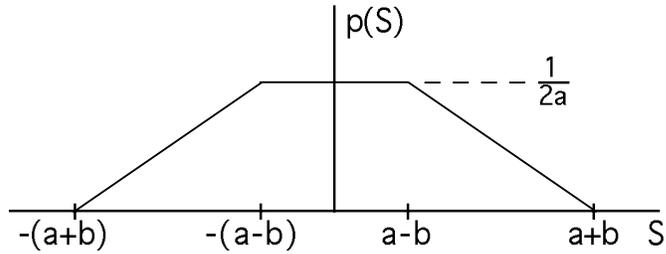
When the widths of the two probability densities are the same, the density for the sum becomes triangular.



. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
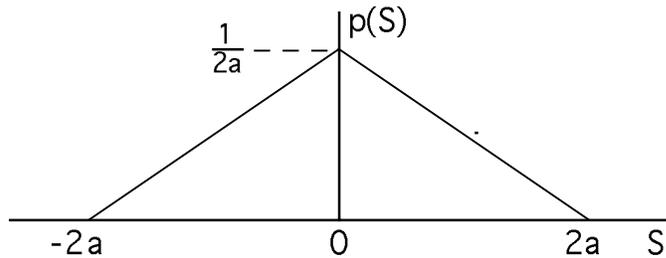
**Example:** Classical Intensity of Unpolarized Light

<u>Given</u> An unpolarized beam of thermal light can be viewed as two statistically independent beams of equal average intensity, polarized at right angles to each other. The total intensity $I_T$ is the sum of the intensities of each of the two polarized components, $I_1$ and $I_2$.

<u>Problem</u> Find the probability density for $I_T$ using the result from a previous example that

$$p(I_1) = \frac{1}{\sqrt{2\pi\alpha I_1}} \exp[-I_1/2\alpha] \qquad I_1 > 0$$

$$= 0 \qquad\qquad\qquad I_1 < 0$$

where $\alpha = < I_1 >$.

<u>Solution</u> Since the light is thermal and unpolarized, the intensity in each of the two polarization directions has the same density: $p_{I_1}(\zeta) = p_{I_2}(\zeta)$.
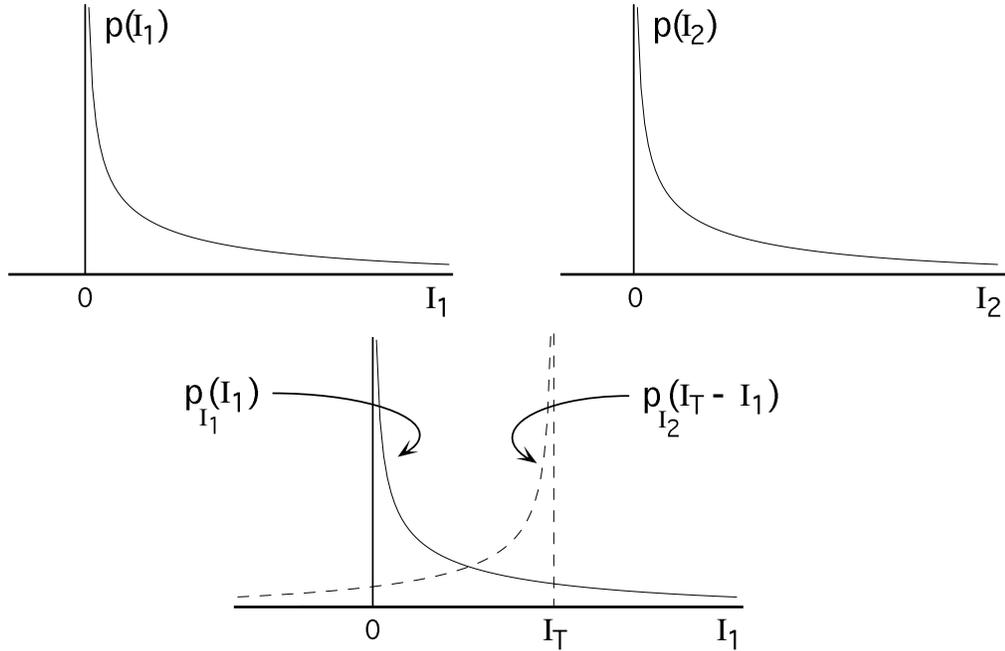
$$
\begin{aligned}
p(I_T) \;&=\; \int_{-\infty}^{\infty} p_{I_1}(I_1)\,p_{I_2}(I_T - I_1)\, dI_1 \\[2ex]
&=\; \int_{0}^{I_T} \left( \frac{\exp[-I_1/2\alpha]}{\sqrt{2\pi\alpha I_1}} \right) \left( \frac{\exp[-(I_T - I_1)/2\alpha]}{\sqrt{2\pi\alpha(I_T - I_1)}} \right) dI_1 \\[2ex]
&=\; \frac{1}{2\pi\alpha}\,\exp[-I_T/2\alpha] \int_{0}^{I_T} [I_1(I_T - I_1)]^{-1/2}\, dI_1 \\[2ex]
&=\; \frac{1}{2\pi\alpha}\,\exp[-I_T/2\alpha] \underbrace{\int_{-I_T/2}^{I_T/2} (\tfrac{1}{4}I_T^2 - x^2)^{-1/2}\, dx}_{\pi} \quad \text{using } x \equiv I_1 - \frac{1}{2}I_T \\[2ex]
&=\; \frac{1}{2\alpha}\,\exp[-I_T/2\alpha] \qquad I_T \geq 0 \\[2ex]
&=\; 0 \qquad\qquad\qquad\qquad I_T < 0
\end{aligned}
$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
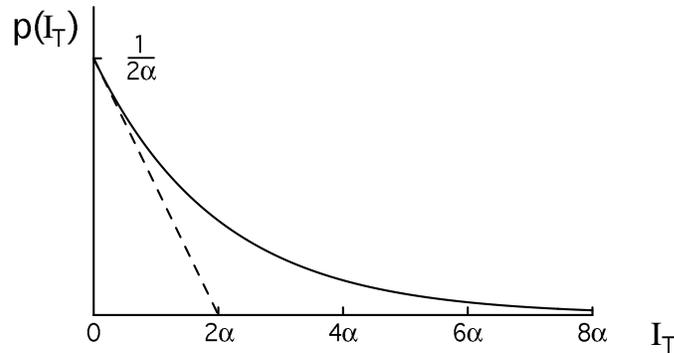
In the two examples just considered the variables being summed had probability densities of the same functional form, rectangles for instance. Certainly this will not always be the case; for example one might be interested in the sum of two variables, one having a uniform density and the other having a Gaussian density. Yet even when the input variables do have probability densities of identical form, the density of the sum will in general have a different functional form than that of the input variables. The two previous examples illustrate this point.

There are three special cases, however, where the functional form of the density is preserved during the addition of statistically independent, similarly distributed variables. The sum of two Gaussian variables is Gaussian. This is shown in an example below. Simply knowing that the result is Gaussian, though, is enough to allow one to predict the parameters of the density. Recall that a Gaussian is completely specified by its mean and variance. The fact that the means and variances add when summing S.I. random variables means that the mean of the resultant Gaussian will be the sum of the input means and the variance of the sum will be the sum of the input variances.

The sum of two S.I. Poisson random variables is also Poisson. Here again, knowing that the result is Poisson allows one to determine the parameters in the sum density. Recall that a Poisson density is completely specified by one number, the mean, and the mean of the sum is the sum of the means.

A Lorentzian (or Cauchy) density depends on two parameters and is given by

$$p(x) = \frac{\Gamma}{\pi} \frac{1}{(x-m)^2 + \Gamma^2}$$

$$m = < x >$$

$\Gamma$ = half width at half height
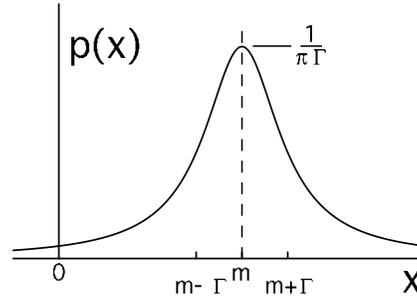


The sum of two S.I. Lorentzian random variables is Lorentzian with $m_S = m_1 + m_2$ and $\Gamma_S = \Gamma_1 + \Gamma_2$. Note that the widths add! Doesn't this contradict the rule about the variances adding, which implies that the widths should grow as the square root of the sum of the squared widths? This apparent contradiction can be resolved by trying to calculate the variance for the Lorentzian density.

**Example:** Sum of Two S.I. Gaussian Random Variables

   Given

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp[-(x - E_x)^2/2\sigma_x^2]$$

$$p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp[-(y - E_y)^2/2\sigma_y^2]$$

   Problem Find $p(S)$ where $S \equiv x + y$.

   Solution

$$
\begin{aligned}
p(S) &= \int_{-\infty}^{\infty} p_x(x) p_y(S-x)\, dx \\
&= (2\pi\sigma_x\sigma_y)^{-1} \int_{-\infty}^{\infty} \exp[-(x - E_x)^2/2\sigma_x^2 - (S - x - E_y)^2/2\sigma_y^2]\, dx
\end{aligned}
$$

The argument of the exponent in the integral is quadratic in the integration variable $x$. It can be simplified by changing variables, $\zeta \equiv x - E_x$, and defining a temporary quantity $a \equiv S - (E_x + E_y)$, a constant as far as the

integration is concerned. The argument of the exponent can then be written and processed as follows.

$$-\frac{1}{2}\left[\frac{\zeta^2}{\sigma_x^2} + \frac{(a-\zeta)^2}{\sigma_y^2}\right] =$$

$$-\frac{1}{2}\left[\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^2}\zeta^2 - \frac{2a}{\sigma_y^2}\zeta + \frac{a^2}{\sigma_y^2}\right] =$$

$$-\frac{1}{2}\left[\left\{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^2}\zeta^2 - \frac{2a}{\sigma_y^2}\zeta + \frac{\sigma_x^2 a^2}{\sigma_y^2(\sigma_x^2 + \sigma_y^2)}\right\} + \left\{\frac{a^2}{\sigma_y^2} - \frac{\sigma_x^2 a^2}{\sigma_y^2(\sigma_x^2 + \sigma_y^2)}\right\}\right] =$$

$$-\frac{1}{2}\left[\left\{\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)^{-1}\left(\zeta - \frac{\sigma_x^2 a}{\sigma_x^2 + \sigma_y^2}\right)^2\right\} + \left\{\frac{a^2}{\sigma_x^2 + \sigma_y^2}\right\}\right]$$

The second of the two terms in { } brackets in the last line produces a constant which factors out of the integral. The integral over $\zeta$ coming about from the first term in { } brackets is just the normalization integral for a Gaussian and has the value

$$\left(2\pi\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)^{1/2}.$$

Bringing all of these terms together gives the probability density for the sum.

$$p(S) = (2\pi\sigma_x\sigma_y)^{-1}\exp[-\frac{a^2}{2(\sigma_x^2 + \sigma_y^2)}]\left(2\pi\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)^{1/2}$$

$$= \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_y^2)}}\exp[-\frac{(S - (E_x + E_y))^2}{2(\sigma_x^2 + \sigma_y^2)}]$$

The result is a Gaussian with mean $E_x + E_y$ and variance $\sigma_x^2 + \sigma_y^2$.

· · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·

The result above applies to the sum of two statistically independent Gaussian variables. It is a remarkable and important fact that the sum of two statistically dependent Gaussian variables is also a Gaussian if the two input variables have a bivariate joint probability density (the special joint density introduced in an earlier example). Those with a masochistic streak and a flair for integration can prove this by applying the expression for the density of the sum of two dependent variables to the bivariate Gaussian joint density.

**Mathematical Digression:** Convolution and Fourier Transforms

After the algebra of the last example it is reasonable to ask if there is some easier way to compute convolutions. There is, and it involves Fourier transforms. This digression is intended for interested students who are familiar with Fourier techniques.

Let the Fourier transform of $f(x)$ be $F(k)$ and use the notation $f \leftrightarrow F$ to indicate a transform pair.

$$F(k) \equiv \int_{-\infty}^{\infty} e^{ikx} f(x) \, dx$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} F(k) \, dk$$

(The Fourier transform of a probability density is called the *characteristic function* or *moment generating function* and is quite useful for more advanced topics in probability theory.)

One can show that the Fourier transform of the convolution of two functions is the product of the individual Fourier transforms.

$$\text{If } a \leftrightarrow A$$

$$\text{and } b \leftrightarrow B$$

$$\text{then } a \otimes b \leftrightarrow AB$$

To find the probability density for the sum of two statistically independent random variables one can multiply the Fourier transforms of the individual probability densities and take the inverse transform of the product. As a practical application and example one can show that the sums of Gaussians are Gaussian, sums of Poisson variables are Poisson, and sums of Lorentzians are Lorentzian.

Gaussian

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(x-E)^2}{2\sigma^2}] \leftrightarrow \exp[ikE - \frac{\sigma^2 k^2}{2}]$$

Poisson

$$\sum_{n=0}^{\infty}\frac{1}{n!}\lambda^n e^{-\lambda}\delta(x-n) \leftrightarrow \exp[\lambda(e^{ik}-1)]$$

Lorentzian

$$\frac{\Gamma}{\pi}\frac{1}{(x-m)^2+\Gamma^2} \leftrightarrow \exp[imk - |k\Lambda|]$$

Each of these three transforms $F(k)$ has the property that a product of similar functions preserves the functional form; only the parameters change. For example if $p_G(S)$ represents the sum of two Gaussians, then

$$
\begin{aligned}
p_G(S) &\leftrightarrow \exp[ikE_1 - \frac{\sigma_1^2 k^2}{2}]\exp[ikE_2 - \frac{\sigma_2^2 k^2}{2}] \\
&\leftrightarrow \exp[ik(E_1+E_2) - \frac{(\sigma_1^2+\sigma_2^2)k^2}{2}]
\end{aligned}
$$

The last expression is the Fourier transform of a Gaussian of mean $E_1 + E_2$ and variance $\sigma_1^2 + \sigma_2^2$. Check to see that the transforms for the Poisson and Lorentzian behave in a similar manner.

· · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·

Simple averaging revealed that for statistically independent random variables the mean of the sum is the sum of the means and the variance is the sum of the variances. The *Central Limit Theorem* gives information about the functional form of the resulting probability density.

Central Limit Theorem (non-mathematical form)
Let $S_n$ be the sum of $n$ statistically independent, identically distributed random variables of mean $E_x$ and variance $\sigma_x^2$ (which must be finite). For large $n$, $p(S_n)$ can often be well represented by a Gaussian with mean $nE_x$ and variance $n\sigma_x^2$.

$$p(S_n) \approx \frac{1}{\sqrt{2\pi n\sigma_x^2}} \exp[-\frac{-(S_n - nE_x)^2}{2n\sigma_x^2}]$$

The vague wording of the theorem as presented here shows immediately why it is qualified by the adjective "non-mathematical". Yet this is perhaps the most useful form of the theorem for our purposes. Consider the specific wording used here.

- $\sigma_x$ (which must be finite)
  This excludes cases such as Lorentzian random variables where the density falls off so slowly with $x$ that the variance is infinite. Of course, for Lorentzians one does not need a Central Limit Theorem for it should be clear from earlier discussion in this section that the sum of $n$ statistically independent Lorentzian random variables will be Lorentzian for any value of $n$.

- For large $n$
  How large? Mathematicians would be more specific. It will be shown below that for practical purposes convergence with increasing $n$ is often quite rapid.

- can often be well represented by

  Certainly not the words of a mathematician! However, the vagueness is deliberate. The Gaussian result presented is a continuous function of the sum variable $S_n$. But a sum of $n$ discrete random variables will always be discrete, no matter how large $n$ might be. A continuous function is not going to converge in any mathematical sense to a sum of $\delta$ functions. What actually happens in such cases is that the <u>envelope</u> of the comb of $\delta$ functions approaches a Gaussian shape. A course grained average of $p(S_n)$ (averaging over a number of discrete values at a time) would approach the continuous density presented in the theorem.

So much for the limitations of the Central Limit Theorem; now consider some extensions.

- The Gaussian can be a good practical approximation for modest values of $n$. The examples which follow illustrate this point.

- The Central Limit Theorem may work even if the individual members of the sum are not identically distributed. A prerequisite is that no individual member should dominate the sum. The necessity for this restriction can be visualized easily. Consider adding 6 variables, each uniformly distributed between -1 and 1, and a 7th variable, uniform between -100 and 100. The density for the sum will look rectangular with rounded edges near $\pm 100$.

- The requirement that the variables be statistically independent may even be waived in some cases, particularly when $n$ is very large.

The Central Limit Theorem plays a very important role in science and engineering. Knowingly or unknowingly, it is often the basis for the assumption that some particular physical random variable has a Gaussian probability density.

**Example:** Sum of Four Uniformly Distributed Variables

Consider a random variable $x$ which is uniformly distributed between $-a$ and $a$.
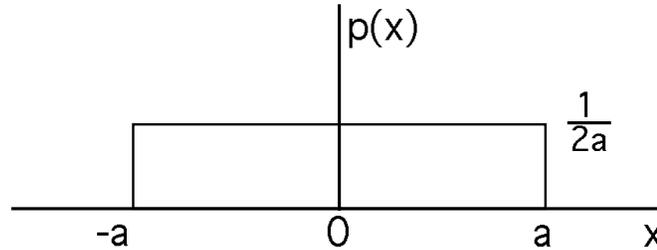


The sum of only four such variables has a density which is reasonably close to the CLT approximation. An earlier example found the density for a sum of two of these variables to be triangular. Since convolution is associative, the density of the sum of four can be found by convolving the triangular density with itself. The result is as follows.

$$
\begin{aligned}
p(S) &= \tfrac{1}{96a^4}(32a^3 - 12aS^2 + 3|S|^3) & 0 \le |S| \le 2a \\[2mm]
&= \tfrac{1}{96a^4}(4a - |S|)^3 & 2a \le |S| \le 4a \\[2mm]
&= 0 & 4a \le |S|
\end{aligned}
$$

The probability density for the sum is plotted below for $a = 1$. In this case the mean of the single random variable is zero and its variance is $1/3$. The CLT approximation has also been plotted, a Gaussian with zero mean and variance equal to $4/3$. Even though the density for a single variable is discontinuous, the density for the sum is quite smooth. The good match to the Gaussian in this example when the number of variables in the sum, $n$, is as small as 4 is due to the fact that the single density is symmetric and falls off rapidly with increasing argument. But note that the true density for the sum is zero beyond $|S| = 4$ while the Gaussian is finite. One could say that the percentage error is infinite beyond $|S| = 4$. Well, you can't have everything. This does illustrate, however, the problems encountered in formulating a mathematical statement of the convergence of the Central Limit Theorem.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
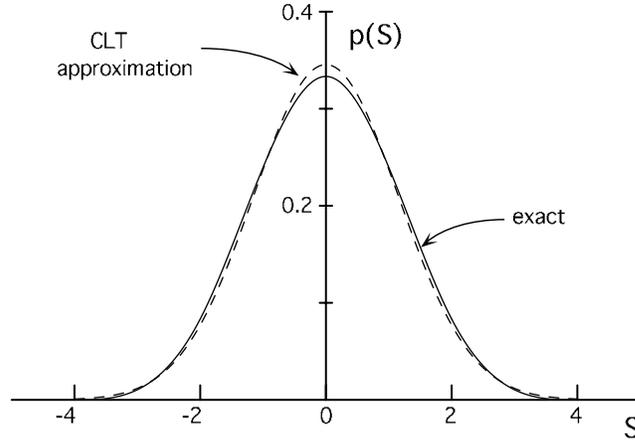
**Example:** Sums of Exponentially Distributed Variables

A single sided exponential is an example of a density that is both discontinuous and asymmetric. It falls off more gradually with increasing argument than does a rectangular or Gaussian density. In this case convergence to the CLT approximation requires a larger number of elements, $n$, in the sum.



$$p(x) = \frac{1}{a} e^{-x/a} \qquad x > 0$$

$$= 0 \qquad x < 0$$
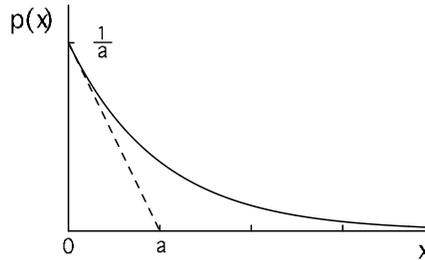
Here the mean is $a$ and the variance is $a^2$. This density is sufficiently simple that the repeated convolution necessary to find the density for the sum of $n$ identical statistically independent variables can be carried out analytically.

$$p(S) = \left[\frac{S^{n-1}}{(n-1)!a^n}\right] e^{-S/a} \qquad S \geq 0$$

$$= 0 \qquad S < 0$$

This result is shown below for $a = 1$ and several values of $n$. The Gaussian approximation is also shown for comparison.
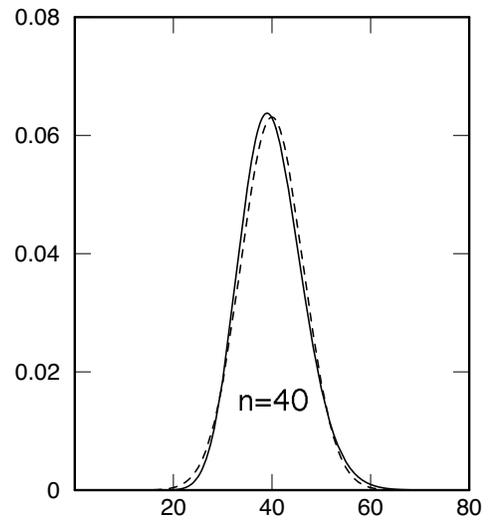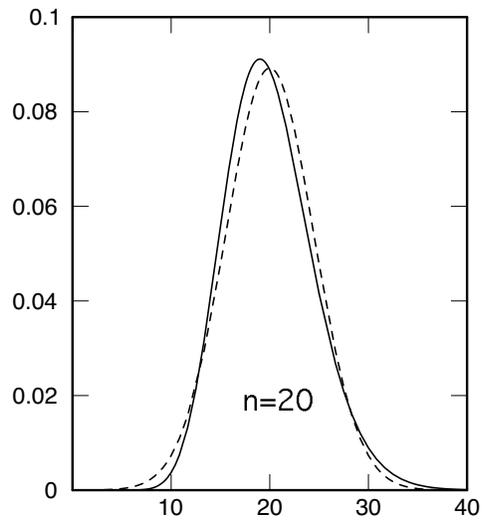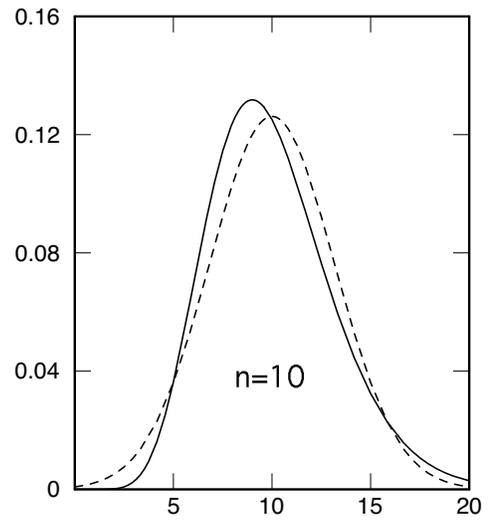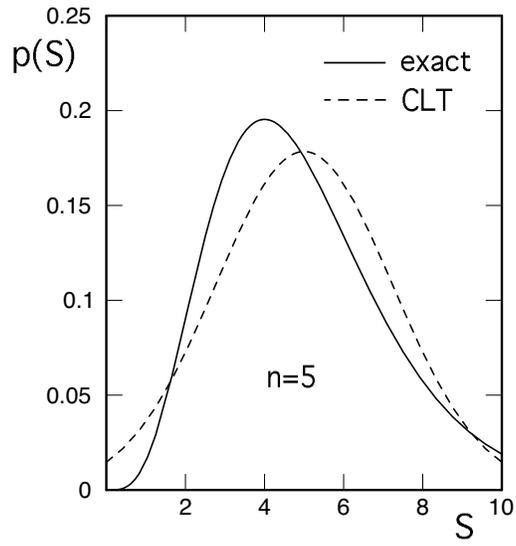
**Example:** Poisson Density for Large $< n >$

The Central Limit Theorem applied to sums of discrete random variables states that the envelope of the train of $\delta$ functions representing $p(S)$ should approach a Gaussian. Here the Poisson density is used to illustrate this behavior. The graphs on the following page show the Poisson density for several values of the mean number of events $< n >$. (Recall that for a Poisson density, the variance is equal to the mean.) The dots indicate the coefficients of the delta functions that would be located at each positive integer value of $S$. The dashed lines show the CLT approximation

$$p(s) = \frac{1}{\sqrt{2\pi < n >}} \exp[-\frac{(S- < n >)^2}{2 < n >}].$$

Since the sum of Poisson random variables also has a Poisson density, the case for which $< n > = 20$ could be the result of adding 10 variables of mean 2 or of adding 5 variables of mean 4. For the Poisson, one can see that it is the mean of the final density, not the number of statistically independent elements in the sum, which determines the convergence to the CLT result. The envelope of the density of a <u>single</u> Poisson random variable will approach a Gaussian if the mean is large enough!

The fact that a single Poisson density approaches a Gaussian envelope for large mean is not typical of discrete densities. Consider the Bose-Einstein (or geometric) density that was treated in a homework problem. For that density the probability of obtaining a given integer decreases monotonically with increasing value, no matter how large the mean of the density. In contrast, the density for the sum of 10 identical Bose-Einstein random variables, each with mean 2, would not be of the Bose-Einstein form and, following the CLT, would exhibit a peak in its envelope near $S = 20$.
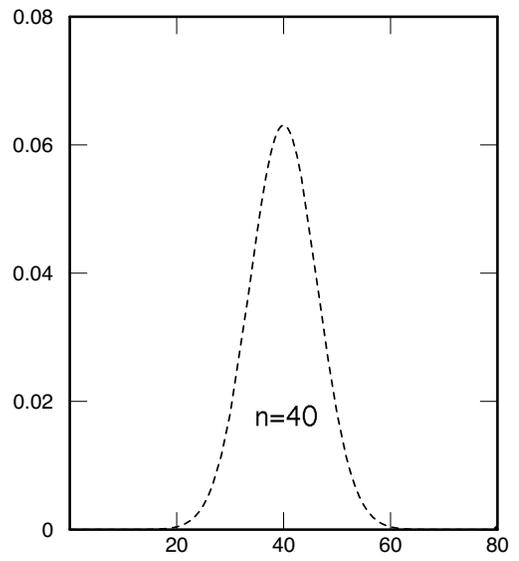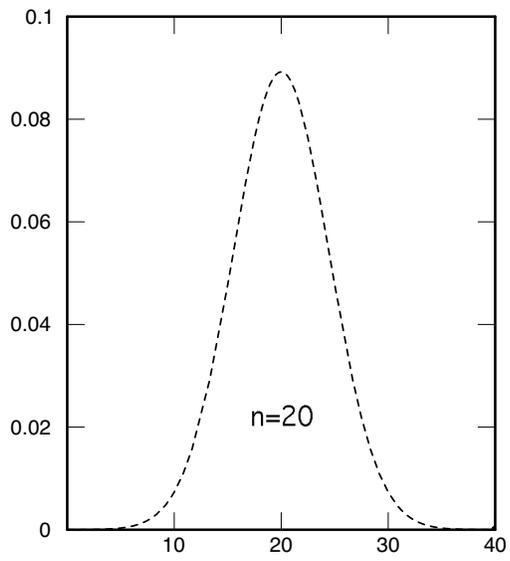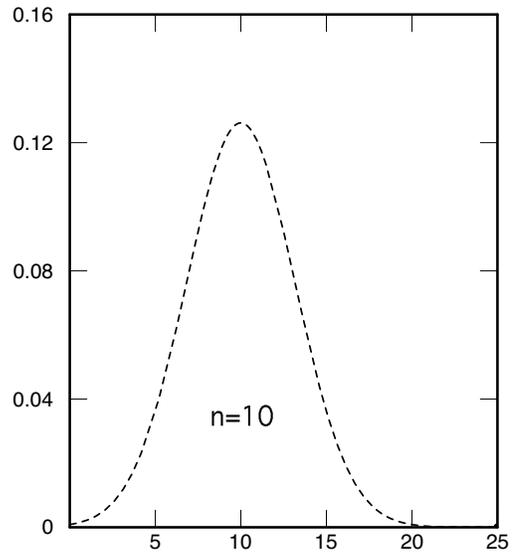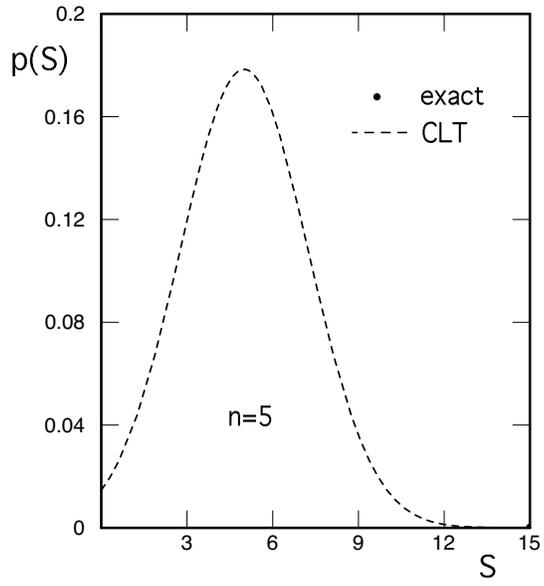
8.044 Statistical Physics I

Spring 2013