8.044 Statistical Physics I
Spring 2008

Lecture Notes on Probability
for
8.044: Statistical Physics I

Notes prepared by: Thomas J. Greytak
Physics Department
Massachusetts Institute of Technology

Spring Term 2008
Lecturer: Young Lee

Preface

Probability is the language of statistical mechanics. It is also fundamental to the understanding of quantum mechanics. Fortunately, mastery of only a few basic concepts and techniques in probability is sufficient for most of the applications that are encountered in undergraduate physics. These notes will introduce the important elements of probability theory; the rest of the course will provide numerous examples of their application. A simultaneous course in quantum mechanics will supply examples of a slightly different nature.

Preface

Probability is the language of statistical mechanics. It is also fundamental to the understanding of quantum mechanics. Fortunately, mastery of only a few basic concepts and techniques in probability is sufficient for most of the applications that are encountered in undergraduate physics. These notes will introduce the important elements of probability theory; the rest of the course will provide numerous examples of their application. A simultaneous course in quantum mechanics will supply examples of a slightly different nature.

1 One Random Variable

A *random variable* is a quantity whose numerical value can not be predicted in advance from the information at hand. However, the processes governing the quantity are sufficiently well behaved that a certain set of rules, probability theory, can be used to obtain a limited knowledge about the possible values the quantity may take.

In general, there are two causes of the uncertainty associated with a random variable. The first is insufficient information about the system governing the random variable. The velocity of an atom selected at random from a classical gas in thermal equilibrium at a known temperature is a random variable. So is the number of atoms in a specific cubic millimeter of such a gas at known average density. If, on the other hand, the positions and velocities of all the atoms in that gas were known at some earlier instant (along with expressions for their differential scattering cross-sections) the variables in question would be *deterministic*. They could be calculated, in principal at least, from the information at hand. The amplitude of a given radio frequency pulse from a pulsar (rotating neutron star) at an antenna on earth is a random variable, unless the signal was monitored as it passed some intermediate location in space. The angle of a pendulum of known total energy would be a random variable, unless the angle at some previous time was known.

The second source of uncertainty giving rise to random variables is quantum mechanics. A system is *completely specified* when one knows all that it is physically possible to know about it. A primary feature of quantum mechanics is that even for systems that are completely specified, the results of measurements of some (but not all) quantities can not be predicted in advance. An electron may be known to be in its ground state in a hydrogen atom, but its radial distance from the proton is a random variable. The momentum of a particle specified to be in a particular energy eigenstate of a harmonic oscillator is a random variable.

Random variables can be divided into three classes according to the spectrum of values that they may take on. The velocity of an atom or the pulsar pulse amplitude mentioned above are examples of *continuous* random variables. The number of gas atoms in a specified region of space is a *discrete* random variable. It is also possible to have *mixed* random variables when the

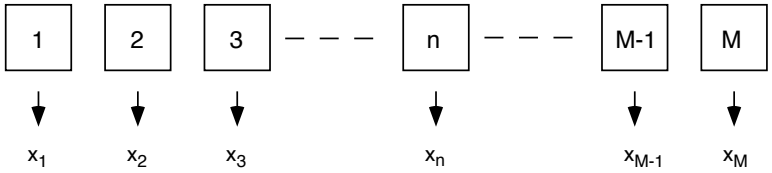
allowed values have both discrete and continuous contributions. For example, the energies available to the electron in a hydrogen atom are discrete for the bound states (energies less than zero) and continuous for the unbound states (energies greater than zero). In a steady state electrical discharge in hydrogen gas the electron energy can be treated as a random variable with a mixed spectrum of values.

Probability theory is based on the idealization of an *ensemble of similarly prepared systems*, each of which produces a single value for the random variable. The key word is "similarly". It means that all those variables or boundary conditions that are determined (fixed, known, specified, or whatever) for the system in question have been set to the same values for each member of the ensemble. For classical systems, similarly prepared does not mean identically prepared since there must be some degrees of freedom left to "chance". An example of a classical system where the members of the ensemble are not identical could be quartz crystals, each containing a single impurity with a classical magnetic moment. Similar preparation could involve specifying the temperature and pressure of the crystal, the applied magnetic field, and the specific lattice site at which the impurity resides. The unspecified degrees of freedom would be the instantaneous velocity and displacements of all the atoms in the crystal. The random variable could be the angle the magnetic moment makes with the applied field.

An example of a quantum ensemble where the members are truly identical is a collection of nuclear isotopes such as ${}_{92}\mathrm{U}^{237}$, each prepared in the same quantum state. No degree of freedom is left unspecified. The random variable could be the time until the isotope decays by beta emission.

Of course, quantum mechanics and ignorance could both contribute to the randomness. For example, the quantum state of the isotopes discussed above, rather than being deterministic, could be distributed at random in a manner consistent with thermal equilibrium at some very high temperature. Each possible quantum state would then have its own decay mode and mean lifetime. Both quantum uncertainty and thermal weighting would contribute to the distribution of emission times.

Assume that an ensemble contains a very large number M of similarly prepared systems. Let the random variable in question be x. The n^{th} member of the ensemble produces the variable x_n.



Imagine compiling a histogram indicating the number of systems for which x falls within certain intervals. The *probability density function* $p_x(\zeta)$ is defined in terms of such a histogram:

$$p_x(\zeta) \equiv \lim_{\substack{M \to \infty \\ d\zeta \to 0}} \frac{\text{number of systems with } \zeta \leq x < \zeta + d\zeta}{M\, d\zeta}$$

As a consequence, if one were to examine a single system prepared in the same manner as the hypothetical ensemble, the probability that its output variable x would fall between ζ and $\zeta + d\zeta$ would be $p_x(\zeta)d\zeta$. [Note: Often $p_x(\zeta)$ is simply written as $p(x)$. When there is a chance of ambiguity or confusion, the longer form with the physical quantity as a subscript and a dummy variable as an argument will be used. In other cases, as in examples when the meaning is clear, the shorter form with the physical quantity as the argument may be used.]

Several properties of $p_x(\zeta)$ follow immediately from its definition as the limiting form of a histogram.

$$p_x(\zeta) \geq 0 \quad \text{for all } \zeta$$

$$\text{probabililty}(a < x \leq b) = \int_a^b p_x(\zeta)\, d\zeta$$

$$\int_{-\infty}^{\infty} p_x(\zeta)\, d\zeta = 1$$

A related quantity which will be found to be very useful is the *probability distribution function*, $P_x(\zeta)$; it is sometimes referred to by the more descriptive phrase *cumulative probability*.

$$P_x(\zeta) \equiv \text{probability}(x \leq \zeta)$$

$$= \int_{-\infty}^{\zeta} p_x(\zeta')\, d\zeta'$$

$$\Rightarrow p_x(\zeta) = \frac{d}{d\zeta} P_x(\zeta)$$

$P_x(\zeta)$ contains no new information; it can be found directly from $p_x(\zeta)$ by integration. Similarly $p_x(\zeta)$ can be found from $P_x(\zeta)$ by differentiation.

A random variable is completely specified by giving its probability density (or, what is equivalent, its distribution function). $p_x(\zeta)$ contains all that it is possible to know about the random variable x for a given system. Of course if the specification of the system is changed by changing the boundary conditions or by constraining some degrees of freedom that were previously left unconstrained, one is dealing with a different problem, one with a new $p_x(\zeta)$. The practical study of random variables reduces to two questions: "How can $p_x(\zeta)$ be found?" and "What information can be extracted from $p_x(\zeta)$ once it is known?". These notes focus on the latter question. The former is the subject of the remainder of this course and a companion course in quantum mechanics.

Example: Radioactive Decay

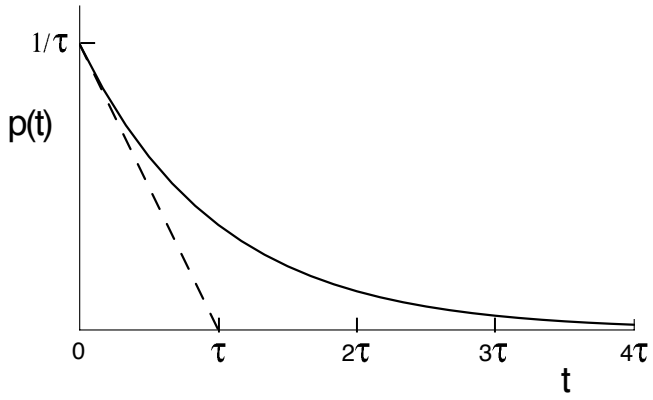
Given [Note: There will be many examples in this course. Such examples are much more satisfying when based on a physical phenomenon rather than a dimensionless mathematical expression. The danger is that the student may feel obligated to understand the physics of the specific system introduced, a pulsar for example. Although some background information may be presented, it is to be taken as "given". The student will be responsible only for the physics and techniques applied to the given situation.]

A radioactive source containing millions of excited nuclei is separated from a detector by a shutter. When the shutter is open, decays are detected at a mean rate of τ^{-1} events per second. If the shutter is opened at a time $t = 0$, the probability density for the waiting time t to the first detected event is

$$p(t) = \begin{cases} \tau^{-1} e^{-t/\tau} & t \geq 0 \\ 0 & t < 0. \end{cases}$$

Problem Examine and discuss $p(t)$ and $P(t)$.

Solution The probability density is presented in its abbreviated form, rather than the more complete form $p_t(\zeta)$. The random variable is continuous and is defined for all times greater than or equal to zero.

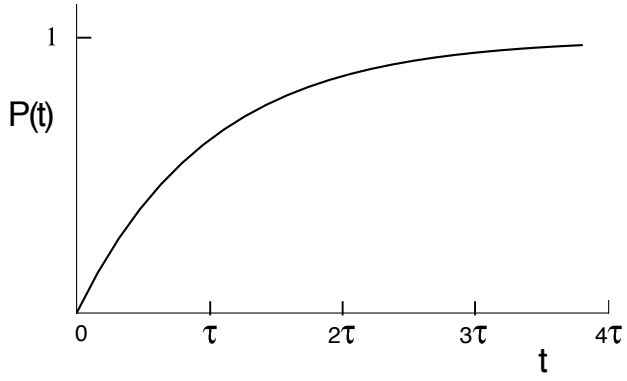


The density is normalized.

$$\begin{aligned}\int_{-\infty}^{\infty} p(t)\,dt &= \int_0^{\infty} e^{-t/\tau}\,\frac{dt}{\tau} = \int_0^{\infty} e^{-y}\,dy \\ &= 1\end{aligned}$$

The cumulative function is found by integration.

$$\begin{aligned}P(t) &= \int_{-\infty}^{t} p_t(\zeta)\,d\zeta = \int_0^{t} e^{-\zeta/\tau}\,\frac{d\zeta}{\tau} = \int_0^{t/\tau} e^{-y}\,dy \\ &= 1 - e^{-t/\tau} \qquad t \geq 0\end{aligned}$$



The density can be recovered from the cumulative by differentiation.

$$\begin{aligned} p(t) &= \frac{d}{dt} P(t) = -(\frac{-1}{\tau})\, e^{-t/\tau} \\ &= \tau^{-1} e^{-t/\tau} \qquad t \geq 0 \end{aligned}$$

. . . . . .

The next example concerns a discrete random variable. It could be dealt with by presenting a table of probabilities, one for each allowed value of the variable. However, it is easier to visualize pictures than tables. Also, it is convenient to deal with continuous, discrete, and mixed variables with a single formalism. Therefore, the concept of a δ function is employed.

Mathematics: The δ Function

The δ function has the following properties.

$$\delta(x) = 0 \qquad \text{if } x \neq 0$$

$$\delta(x) \to \infty \text{ (it is undefined)} \qquad \text{if } x = 0$$

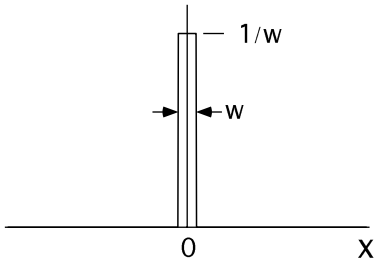
$$\int_{-\epsilon}^{\epsilon} \delta(x)\, dx = 1 \qquad \text{for all } \epsilon > 0$$

Its graphical representation is a vertical arrow.

The δ function can be thought of as the limit of a narrow, normalized function as the width goes to zero.

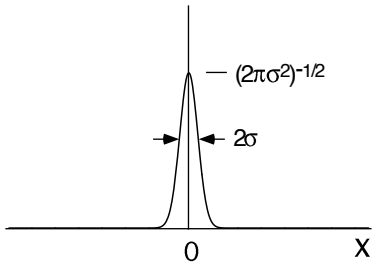
Rectangular Model

$$\delta(x) = lim_{w\to 0} \begin{cases} \frac{1}{w} & |x| \le \frac{w}{2} \\ 0 & |x| > \frac{w}{2} \end{cases}$$

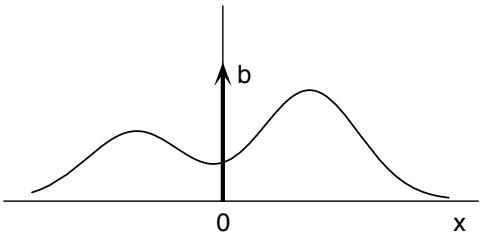


Gaussian Model

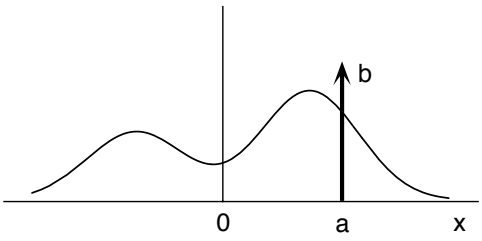
$$\delta(x) = lim_{\sigma\to 0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{x^2}{2\sigma^2}]$$



If a well behaved function $f(x)$ is multiplied by a δ function and that product is integrated over all space, the result is the value of $f(x)$ at the location of the δ function.



$$\int_{-\infty}^{\infty} f(x)(b\,\delta(x))\,dx = bf(0)$$

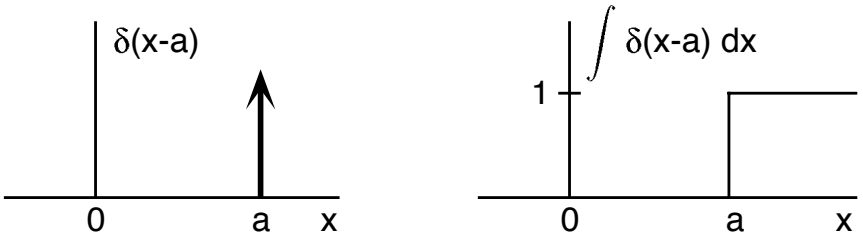


$$\int_{-\infty}^{\infty} f(x)(b\,\delta(x-a))\,dx = bf(a)$$

Note that it is customary to indicate the coefficient of the δ function by a number (or expression) next to its symbol on a graph. The integral of a δ

function is a unit step.

$$\int_{-\infty}^{x} \delta(\zeta - a)\, d\zeta = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$$



. . . . . .

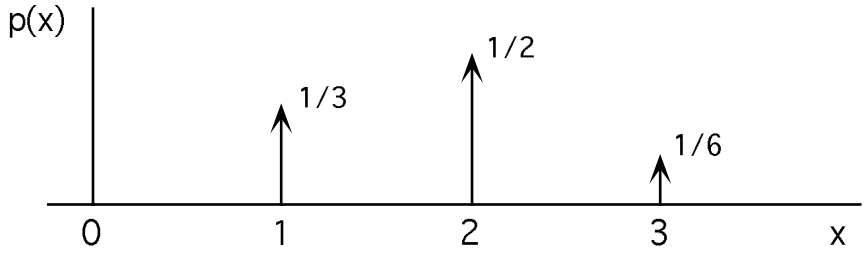
Example: Closely Bound Stars

Given Many "stars" are in fact closely bound groups (not to be confused with more widely spaced aggregations such as globular and open clusters). For illustrative purposes one might use the following probabilities for the number n of stars in a closely bound group.

$p_n(0)$	=	0	by definition
$p_n(1)$	=	1/3	a single star, such as the Sun (unless you believe in Nemisis)
$p_n(2)$	=	1/2	a binary, as in Sirius
$p_n(3)$	=	1/6	a trinary, as in α Centari
$p_n(> 3)$	=	negligible	

Problem Examine and discuss the probabilities.

Solution The probabilities can be displayed on a graph using δ functions.

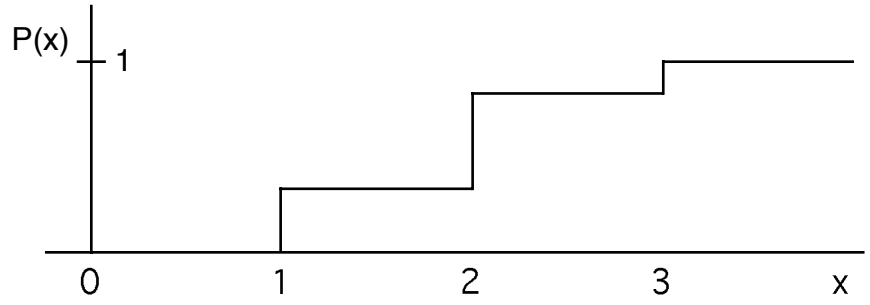


Here x is used as the variable to emphasize the continuous nature of the horizontal axis.

$$p(x) = \frac{1}{3}\delta(x-1) + \frac{1}{2}\delta(x-2) + \frac{1}{6}\delta(x-3)$$

$$\int_{-\infty}^{\infty} p(x)\,dx = \frac{1}{3} + \frac{1}{2} + \frac{1}{6} = 1$$

$$\begin{aligned} P(x) = \int_{-\infty}^{x} p_x(\zeta)\,d\zeta &= 0 && x < 1 \\ &= 1/3 && 1 \le x < 2 \\ &= 5/6 && 2 \le x < 3 \\ &= 1 && 3 \le x \end{aligned}$$



Differentiation of $P(x)$ produces δ functions at each step discontinuity with coefficients equal to the height of the step. Thus $p(x)$ is recovered.

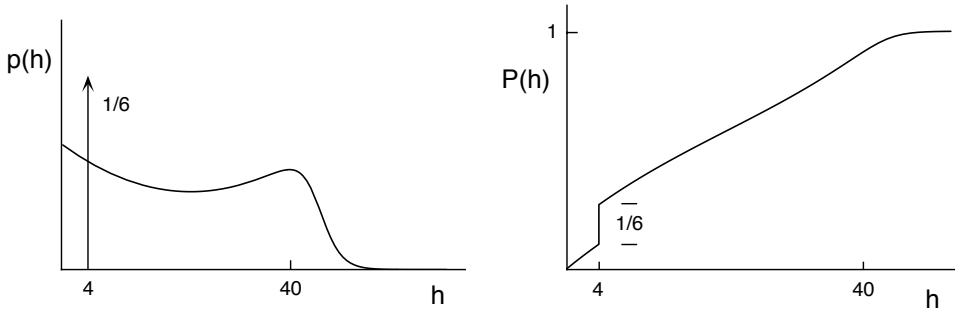
. . . . . .

Example: A Logged Forest

Given Loggers enter a mature forest and cut down one sixth of the trees, leaving stumps exactly 4 feet high.

Problem Sketch $p(h)$ and $P(h)$ for the subsequent tree trunk height h.

Solution This is an example of a random variable having a spectrum of possible values containing both a discrete and a continuous part.

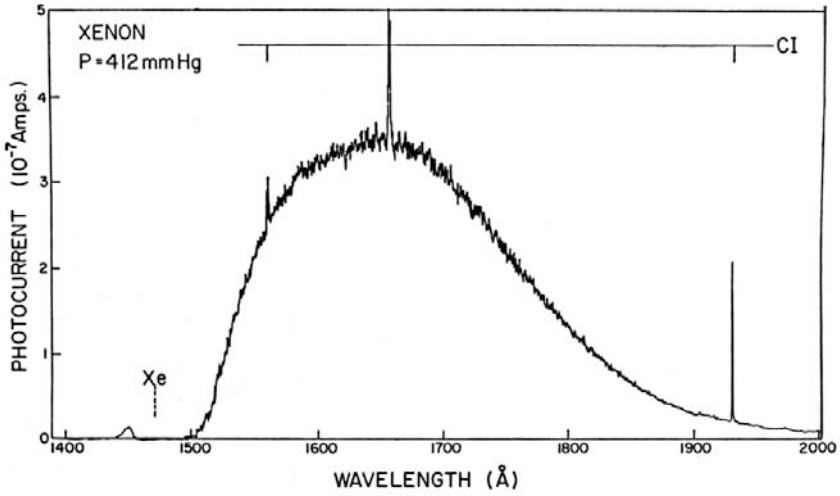


. . . . . .

The above example is admittedly fatuous, but it does provide a simple illustration of a mixed random variable and the associated probability functions. The next example has more physical content.

Example: Xenon Ultraviolet Light Source

Given An electric discharge is maintained in xenon gas. The intensity of the emitted light as a function of wavelength is shown in the figure.



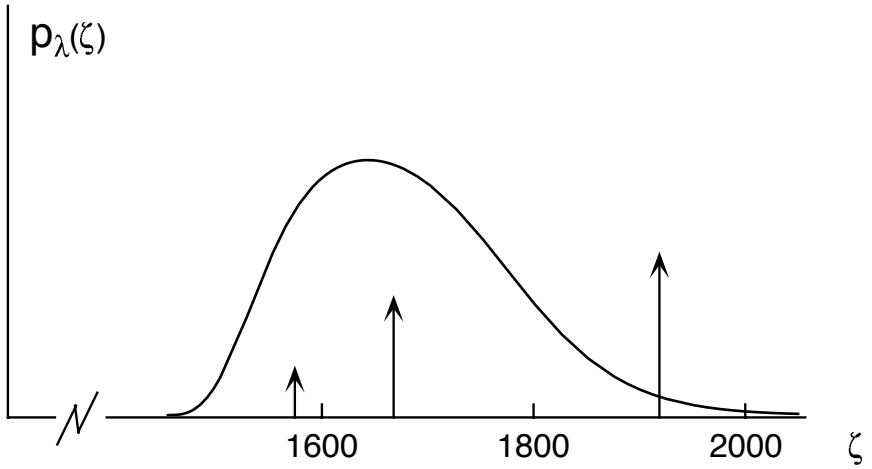
Problem Discuss the nature of the spectrum and cast the measurements in the form of a probability density.

Solution It was mentioned earlier that the bound electron states of atoms have discrete energies and the unbound (ionized) states have a continuum

of energies. Transitions between bound states will give rise to photons with sharply defined energies. Transitions from bound to unbound states, or between unbound ones, give rise to a continuous distribution of photon energies. This explains the presence of both sharp lines and a continuous portion in the xenon data. Note that the discrete and continuous parts of the electron energy spectrum in the atom do not overlap; however, the discrete and continuous parts of the photon energy spectrum, arising from differences of electronic energies, do overlap.

A careful study of the figure reveals some interesting features. The vertical axis displays the current obtained at the output of photo-multiplier tube. This current is directly proportional to the number of photons from the incident light beam which fall on the photocathode. If one assumes that the sensitivity of the photocathode is independent of the wavelength of the photons in the region of interest, then the amplitude of the displayed curve can be taken to be proportional to the number of detected photons. (In modern experiments, in particular when the incoming light is very weak, one often counts and records the number of photoelectron events directly.) The sharp lines in the figure are not delta functions; rather, they have a finite width. This width could be intrinsic, due to the finite lifetime of the excited electronic states of the atom, or it could be instrumental, due to the finite resolving power of the spectrometer. The small scale fluctuations in the trace (the "grass" whose height increases with increasing continuum intensity) is not an intrinsic feature of the continuum emission. It is an artifact of the detection process known as shot noise. Shot noise will be discussed later in the course.

The horizontal axis displays the wavelength λ of the light. The wavelength of the photons emitted by the source can be considered a random variable. The trace in the figure would then be proportional to $p_\lambda(\zeta)$ where $p_\lambda(\zeta)\,d\zeta$ is the probability that a given detected photon would have a wavelength between ζ and $\zeta + d\zeta$.



Often it is more physical to think about the energy of a photon rather than its wavelength. The energy is inversely proportional to the wavelength, $E = 2\pi\hbar c/\lambda$, so a simple linear scaling of the horizontal axis will not give the probability of finding a photon of energy E. A technique will be introduced later, functions of a random variable, that can be used to find the probability density for a variable $y \equiv a/x$ if the probability density for x is known.

. . . . . .

The probability density contains complete information about a random variable. Many times, however, less information is needed for a particular application. The desired information often takes the form of an *ensemble average* of a function $f(x)$ of the random variable x. Such an average can be computed as an integral over the probability density.

$$< f(x) > \equiv \int_{-\infty}^{\infty} f(\zeta) p_x(\zeta)\, d\zeta$$

Some averages occur frequently enough to deserve separate names. The *mean*, $< x >$, is the average value of the random variable itself. It is a measure of the location of $p(x)$ along the x axis.

$$< x > = \int_{-\infty}^{\infty} x p(x)\, dx$$

The *mean square*, $< x^2 >$, is often associated with the energy in a field or power in a signal.

$$< x^2 > = \int_{-\infty}^{\infty} x^2 p(x)\, dx$$

The *variance*, $\mathrm{Var}(x)$, is a quadratic measure of the fluctuations of x about its mean.

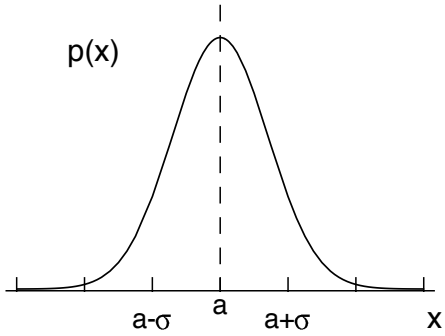
$$\begin{aligned}
\mathrm{Var}(x) &\equiv < (x - < x >)^2 > \\
&= \int_{-\infty}^{\infty} (x - < x >)^2 p(x)\, dx \\
&= \int_{-\infty}^{\infty} x^2 p(x)\, dx - 2 < x > \int_{-\infty}^{\infty} x p(x)\, dx + < x >^2 \int_{-\infty}^{\infty} p(x)\, dx \\
&= < x^2 > - 2 < x >< x > + < x >^2 \\
&= < x^2 > - < x >^2
\end{aligned}$$

The *standard deviation* is defined as the square root of the variance. It is a measure of the width of the probability density.

Example: The Gaussian Density

Given Probability densities of the Gaussian form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(x-a)^2}{2\sigma^2}]$$



are frequently encountered in science and engineering. They occur in thermal equilibrium when the variable makes a quadratic contribution to the energy of the system (the x component of the velocity of an atom in a gas for example). They also occur when the variable is the sum of a large number of small, independent contributions (such as the final displacement after a random walk along a line).

Problem Find the mean, variance, and standard deviation for a random variable with a Gaussian probability density. Sketch $P(x)$.

Solution

$$
\begin{aligned}
< x > &= \int_{-\infty}^{\infty} xp(x)\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp[-\frac{(x-a)^2}{2\sigma^2}]\,dx \\
&\quad \text{Let } \eta = x - a; \text{then } d\eta = dx \\
&= a \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[-\frac{\eta^2}{2\sigma^2}]\,d\eta}_{1} + \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{\infty} \eta \exp[-\frac{\eta^2}{2\sigma^2}]\,d\eta}_{0} \\
&= a
\end{aligned}
$$

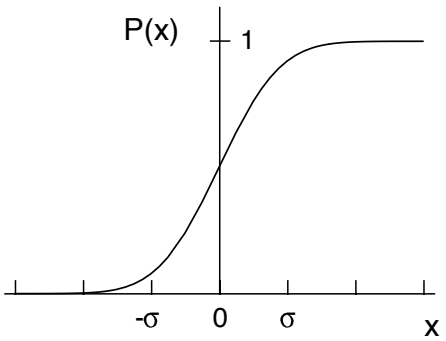
$$
\begin{aligned}
< x^2 > &= \int_{-\infty}^{\infty} x^2 p(x)\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\eta^2 + 2\eta a + a^2) \exp[-\frac{\eta^2}{2\sigma^2}]\,d\eta \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{\infty} \eta^2 \exp[-\frac{\eta^2}{2\sigma^2}]\,d\eta}_{\sqrt{2\pi}\sigma^3} + 0 + a^2 \\
&= \sigma^2 + a^2
\end{aligned}
$$

$$\text{Var}(x) =< x^2 > - < x >^2 = \sigma^2$$

$$\text{Standard Deviation} = \sigma$$

Note that the Gaussian density is completely specified by two parameters, its mean and its variance. In the case where $a = 0$ the cumulative function is

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp[-\frac{\zeta^2}{2\sigma^2}]\, d\zeta$$



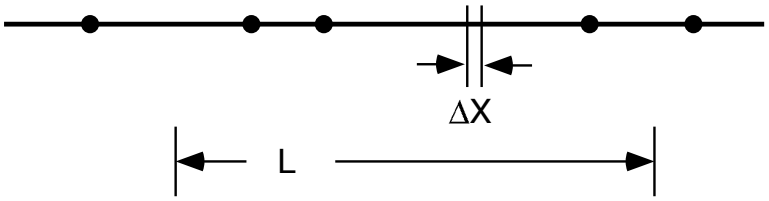
The integral can not be expressed as a closed form expression using simple functions. It can be found tabulated as the "error function".

. . . . . .

Example: The Poisson Density

Given Imagine a situation in which events occur at random along a line X and are governed by the two conditions

- In the limit $\Delta X \to 0$ the probability that one and only one event occurs between X and $X + \Delta X$ is given by $r\Delta X$, where r is a given constant independent of X.
- The probability of an event occurring in some interval ΔX is statistically independent of events in all other portions of the line.



Under these circumstances the probability that exactly n events will occur in an interval of length L is given by the Poisson probability:

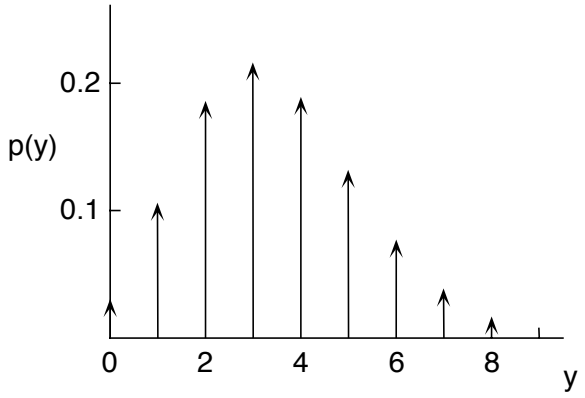
$$p(n) = \frac{1}{n!}(rL)^n e^{-rL} \qquad n = 0, 1, 2, \cdots$$

Although this result is to be taken as a given now, it will be derived in the next section. The Poisson probability can be cast in terms of a continuous

random variable y as follows:

$$p(y) = \sum_{n=0}^{\infty} p(n)\,\delta(y-n)$$

This is shown below for the case were $rL = 3.5$.



The number of decays detected in a lump of radioactive material in a time T will be Poisson. Of course T would have to be short compared to the mean lifetime of a single nucleus, or the mean decay rate would change during the interval T (as the lump is depleted of excited nuclei) and the first condition would be violated.

When the Poisson density is derived later it will become clear that it is not restricted to a one-dimensional geometry. As long as events in different "regions" are statistically independent the results will be Poisson even if the regions are arrayed in a two or three dimensional pattern. The number of ^3He atoms in one square millimeter of a liquid monolayer of naturally occurring helium (0.013% ^3He) would be Poisson. [Note that since two atoms can not occupy the same spot, the second condition above will be violated if the mean distance between ^3He atoms becomes comparable to atomic dimensions.] The number of atoms closer than 100 angstroms to a given atom in a dilute non-interacting 3-dimensional gas similarly would be Poisson.

Problem Check the normalization, find the mean and variance, and show that the density is specified by a single parameter, its mean.

Solution

Normalization

$$\begin{aligned}\int_{-\infty}^{\infty} p(y)\,dy &= \int_{-\infty}^{\infty}\sum_{n=0}^{\infty} p(n)\delta(y-n)\,dy \\ &= \sum_{n=0}^{\infty} p(n)\underbrace{\int_{-\infty}^{\infty}\delta(y-n)\,dy}_{1} = \sum_{n=0}^{\infty} p(n) \\ &= e^{-rL}\sum_{n=0}^{\infty}\frac{1}{n!}(rL)^n = e^{-rL}e^{rL} \\ &= 1\end{aligned}$$

Mean

$$\begin{aligned}< n > &= \int_{-\infty}^{\infty} yp(y)\,dy = \sum_{n=0}^{\infty} p(n)\underbrace{\int_{-\infty}^{\infty} y\delta(y-n)\,dy}_{n} \\ &= \sum_{n=0}^{\infty} np(n) = e^{-rL}\sum_{n=0}^{\infty}\frac{n}{n!}(rL)^n\end{aligned}$$

This sum can be done by differentiating the previous sum used to check the normalization.

$$\frac{d}{dr}[\sum_{n=0}^{\infty}\frac{1}{n!}(rL)^n] = \sum_{n=0}^{\infty}\frac{n}{n!}r^{n-1}L^n = \frac{1}{r}\sum_{n=0}^{\infty}\frac{n}{n!}(rL)^n$$

Alternatively

$$\frac{d}{dr}[\sum_{n=0}^{\infty}\frac{1}{n!}(rL)^n] = \frac{d}{dr}[e^{rL}] = Le^{rL}$$

Equating the two results gives the required sum.

$$\sum_{n=0}^{\infty}\frac{n}{n!}(rL)^n = rLe^{rL}$$

Finally

$$\begin{aligned} < n > &= e^{-rL} rLe^{rL} \\ &= rL \\ &= \text{(rate)} \times \text{(interval)} \end{aligned}$$

Mean Square

$$< n^2 >= \sum_{n=0}^{\infty} n^2 p(n) = e^{-rL} \sum_{n=0}^{\infty} \frac{n^2}{n!} (rL)^n$$

Again, proceed by differentiating the previous sum, the one worked out for the mean.

$$\frac{d}{dr}[\sum_{n=0}^{\infty} \frac{n}{n!}(rL)^n] = \sum_{n=0}^{\infty} \frac{n^2}{n!} r^{n-1} L^n = \frac{1}{r} \sum_{n=0}^{\infty} \frac{n^2}{n!}(rL)^n$$

Alternatively

$$\frac{d}{dr}[\sum_{n=0}^{\infty} \frac{n}{n!}(rL)^n] = \frac{d}{dr}[rLe^{rL}] = Le^{rL} + rL^2 e^{rL}$$

Again, equating the two results gives the required sum.

$$\sum_{n=0}^{\infty} \frac{n^2}{n!}(rL)^n = rLe^{rL} + (rL)^2 e^{rL}$$

Concluding the calculation

$$\begin{aligned} < n^2 > &= e^{-rL}(rL + (rL)^2) e^{rL} \\ &= rL + (rL)^2 \\ &= < n > + < n >^2 \end{aligned}$$

Variance

$$\begin{aligned} \text{Var}(n) &= < n^2 > - < n >^2 \\ &= < n > \end{aligned}$$

The Poisson probability is most conveniently expressed in terms of its mean:

$$p(n) = \frac{1}{n!} < n >^n e^{-<n>} \qquad \text{where} \quad < n >= rL.$$

. . . . . .

2 Two Random Variables

A number of features of the two-variable problem follow by direct analogy with the one-variable case: the *joint probability density*, the *joint probability distribution function*, and the method of obtaining averages.

$$p_{x,y}(\zeta,\eta)d\zeta d\eta \equiv \text{prob.}(\zeta < x \leq \zeta + d\zeta \text{ and } \eta < y \leq \eta + d\eta)$$

$$\begin{aligned} P_{x,y}(\zeta,\eta) &\equiv \text{prob.}(x \leq \zeta \text{ and } y \leq \eta) \\ &= \int_{-\infty}^{\zeta}\int_{-\infty}^{\eta} p_{x,y}(\zeta',\eta')d\zeta' d\eta' \end{aligned}$$

$$p_{x,y}(\zeta,\eta) = \frac{\partial}{\partial\zeta}\frac{\partial}{\partial\eta}P_{x,y}(\zeta,\eta)$$

$$< f(x,y) >= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\zeta,\eta)p_{x,y}(\zeta,\eta)d\zeta d\eta$$

The discussion of two random variables does involve some new concepts: reduction to a single variable, conditional probability, and statistical independence. The probability density for a single variable is obtained by integrating over all possible values of the other variable.

$$p_x(\zeta) = \int_{-\infty}^{\infty} p_{x,y}(\zeta,\eta)d\eta$$

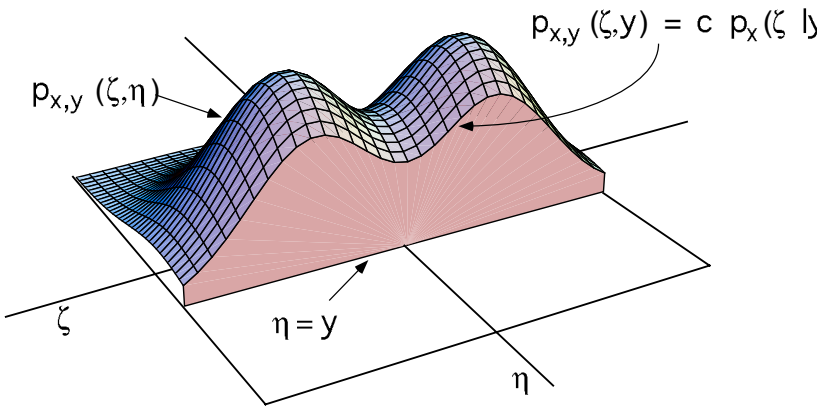
$$p_y(\eta) = \int_{-\infty}^{\infty} p_{x,y}(\zeta,\eta)d\zeta$$

These expressions arise because p_x refers to the probability density for x regardless of the value of y.

At the other extreme of knowledge (or lack of it) is the *conditional probability density* $p_x(\zeta|y)$, defined to be the probability density for x given that y is known to have the indicated value.

$$p_x(\zeta|y)d\zeta \equiv \text{prob.}(\zeta < x \leq \zeta + d\zeta \text{ given that } \eta = y)$$

Note that in the expression $p_x(\zeta|y)$, ζ is a variable but y is simply a parameter. $p_x(\zeta|y)$ has all the properties of a probability density function of a single random variable, ζ. The following picture may be helpful in understanding the connection between the joint probability density $p_{x,y}(\zeta,\eta)$ and the conditional probability density $p_x(\zeta|y)$.



The specification that y is known restricts the possibilities to those lying on the line $\eta = y$ in the ζ – η plane. Therefore, $p_x(\zeta|y)$ must be proportional to $p_{x,y}(\zeta, y)$:

$$p_{x,y}(\zeta, y) = c\, p_x(\zeta|y).$$

The constant of proportionality, c, can be found by integrating both sides of the above equality over all ζ.

$$\int_{-\infty}^{\infty} p_{x,y}(\zeta, y)\, d\zeta = p_y(\eta = y) \qquad \{\text{reduction to a single variable }\}$$

$$c \underbrace{\int_{-\infty}^{\infty} p_x(\zeta|y)\, d\zeta}_{1} = c \qquad \{\text{normalization of } p_x(\zeta|y)\ \}$$

$$\Rightarrow \quad c = p_y(\eta = y)$$

This result is known as *Bayes' Theorem* or the *fundamental law of conditional probability*:

$$p_{x,y}(\zeta, y) = p_x(\zeta|y) p_y(\eta = y)$$

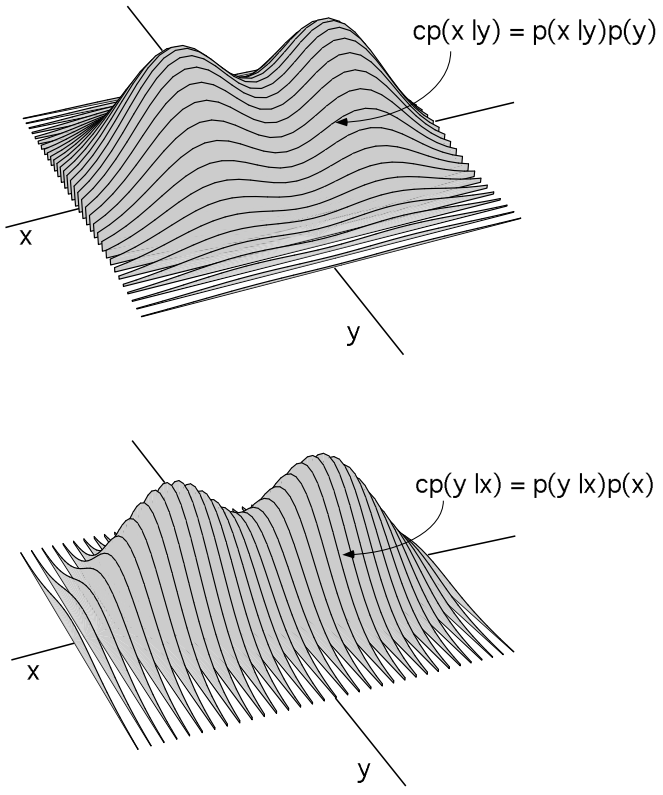
The result can be viewed in two ways. It can be interpreted as a way of finding the conditional density from the joint density (and the density of the conditioning event which can be recovered from the joint density):

$$p(x|y) = \frac{p(x,y)}{p(y)}.$$

This view is illustrated in the previous figure where $p(x|y)$ is exposed by 'slicing through' $p(x,y)$. Alternatively Bayes' theorem can be interpreted as a way of constructing the joint density from a conditional density and the probability of the conditioning variable:

$$p(x,y) = p(x|y)p(y).$$

This is illustrated below for the two possible choices of conditioning variable. Here, as with store-bought bread, one can reassemble the loaf by stacking the individual slices side by side.



The two random variables are said to be *statistically independent* (S.I.) when their joint probability density factors into a product of the densities of the two individual variables:

$$p_{x,y}(\zeta,\eta) = p_x(\zeta)p_y(\eta) \qquad \text{if } x \text{ and } y \text{ are S.I.}$$

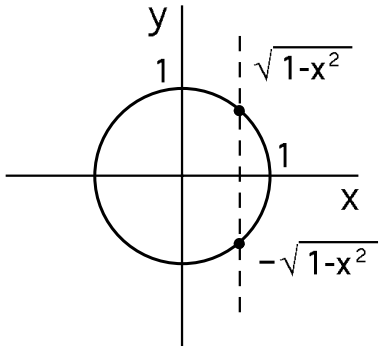
Physically, two variables are S.I. if knowledge of one gives no additional information about the other beyond that contained in its own unconditioned probability density:

$$p(x|y) = \frac{p(x,y)}{p(y)} = p(x) \qquad \text{if } x \text{ and } y \text{ are S.I.}$$

Example: Uniform Circular Disk

Given The probability of finding an event in a two dimensional space is uniform inside a circle of radius 1 and zero outside of the circle.

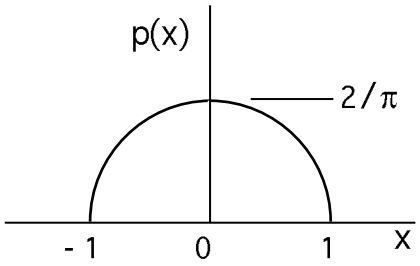
$$\begin{aligned} p(x,y) &= 1/\pi \quad x^2+y^2 \le 1 \\ &= 0 \quad x^2+y^2 > 1 \end{aligned}$$



Problem Find $p(x)$, $p(y)$, and $p(x|y)$. Are x and y S.I.?

Solution

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} p(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy = \frac{2}{\pi}\sqrt{1-x^2} \qquad |x| \le 1 \\ &= 0 \qquad |x| > 1 \end{aligned}$$

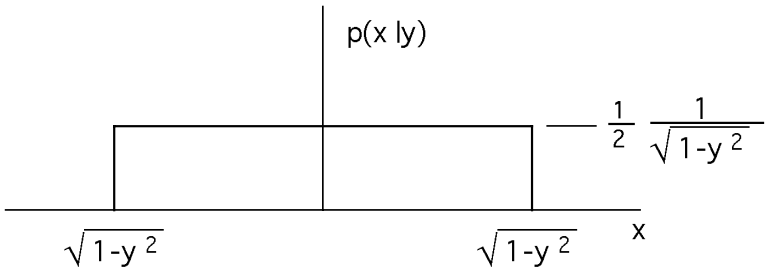


By symmetry, the functional form for $p(y)$ will be identical.

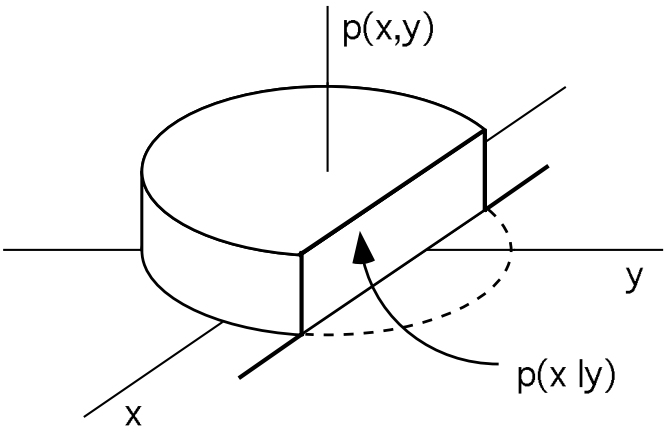
$$\begin{aligned} p(y) &= \frac{2}{\pi}\sqrt{1-y^2} && |y| \le 1 \\ &= 0 && |y| > 1 \end{aligned}$$

It is apparent that the product of $p(x)$ and $p(y)$ does not equal $p(x,y)$, so the random variables x and y are not S.I. The conditional probability is found from Bayes' theorem.

$$\begin{aligned} p(x|y) &= \frac{p(x,y)}{p(y)} = \frac{(1/\pi) \quad \{\text{when } x^2 \le 1-y^2\}}{(2/\pi)\sqrt{1-y^2} \quad \{\text{when } y^2 \le 1\}} \\ &= \frac{1}{2\sqrt{1-y^2}} \qquad |x| \le \sqrt{1-y^2} \\ &= 0 \qquad\qquad \text{elsewhere} \end{aligned}$$



It is not surprising that $p(x|y)$ is a constant when one considers the following interpretation.



. . . . . .

Example: Derivation of the Poisson Density

Given Events occurring at random alone a line X are governed by the following two conditions:

- In the limit $\Delta X \to 0$ the probability that one and only one event occurs between X and $X + \Delta X$ is given by $r\Delta X$, where r is a given constant independent of X.
- The probability of an event occurring in some interval ΔX is statistically independent of events in all other portions of the line.

Problem

a) Find the probability $p(n = 0; L)$ that no events occur in a region of length L. Proceed by dividing L into an infinite number of S.I. intervals and calculate the joint probability that none of the intervals contains an event.

b) Obtain the differential equation

$$\frac{d}{dL}p(n; L) + rp(n; L) = rp(n - 1; L)$$

as a recursion relation governing the p(n;L).

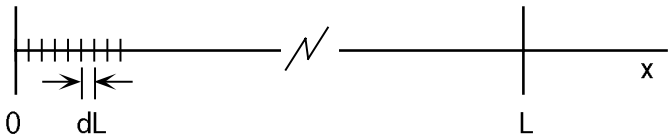
c) Show that the Poisson density

$$p(n;L) = \frac{1}{n!}(rL)^n e^{-rL}$$

is a solution of the equation. Is it unique?

Solution

a) To find $p(n=0;L)$ divide L into intervals each of length dL:

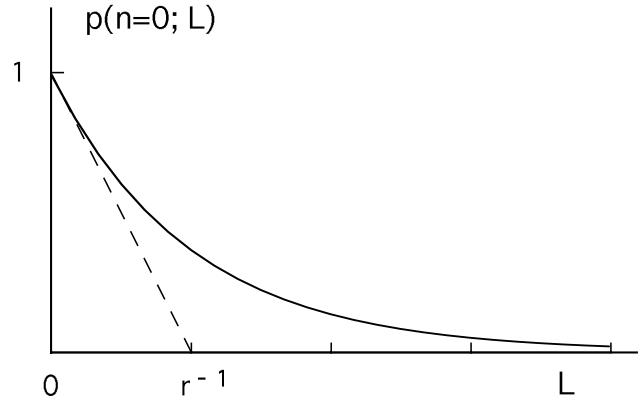


Consider dL so short that $p(0) >> p(1) >> p(n>1)$ in dL. But the probabilities must sum to unity, $p(0)+p(1)+p(2)+\cdots = 1$, so one can find an approximation to $p(0)$ which will be valid in the limit of small dL.

$$p(0) \approx 1 - p(1) = 1 - r(dL)$$

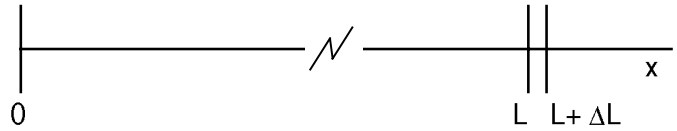
The probability of an event in any sub-interval is S.I. of the events in every other sub-interval, so

$$\begin{aligned}
p(n=0;L) &= \prod_{m=1}^{m=L/dL} (1 - r(dL)) \\
\ln p(n=0;L) &= \sum_m \ln(1 - r(dL)) \\
\ln p(n=0;L) &\approx \sum_{m=1}^{m=L/dL} -r(dL) \\
&= -\left(\frac{L}{dL}\right) r(dL) = -rL \\
p(n=0;L) &= e^{-rL}
\end{aligned}$$



Note that $\int_0^\infty p(n=0;L)\,dL \neq 1$ since $p(n=0;L)$ is not a probability density for L, rather it is *one element* of a discrete probability density for n which depends on L as a parameter.

b) Now consider the span $X = 0$ to $X = L + \Delta L$ to be composed of the finite length L and an infinitesimal increment ΔL.



The two intervals are S.I. so one may decompose $p(n; L + \Delta L)$ in terms of two mutually exclusive events.

$$\begin{aligned} p(n; L + \Delta L) &= p(n; L)p(0; \Delta L) + p(n-1; L)p(1; \Delta L) \\ &= p(n; L)(1 - r\Delta L) + p(n-1; L)(r\Delta L) \end{aligned}$$

Rearranging

$$\frac{p(n; L + \Delta L) - p(n; L)}{\Delta L} = rp(n-1; L) - rp(n; L)$$

Passing to the limit $\Delta L \to 0$ gives

$$\frac{d\,p(n; L)}{dL} = rp(n-1; L) - rp(n; L)$$

c) To show that the Poisson density satisfies this equation, take its derivative with respect to L and compare the result with the above expression.

$$\begin{aligned} p(n;L) &= \frac{1}{n!}(rL)^n e^{-rL} \\ \frac{d}{dL}p(n;L) &= r\,\frac{n}{n!}(rL)^{n-1}e^{-rL} - r\frac{1}{n!}(rL)^n e^{-rL} \\ &= r\frac{1}{(n-1)!}(rL)^{n-1}e^{-rL} - r\frac{1}{n!}(rL)^n e^{-rL} \\ &= rp(n-1;L) - rp(n;L) \end{aligned}$$

This solution is unique when the differential recursion relation is supplemented by the boundary conditions

$$p(0;L) = e^{-rL}$$

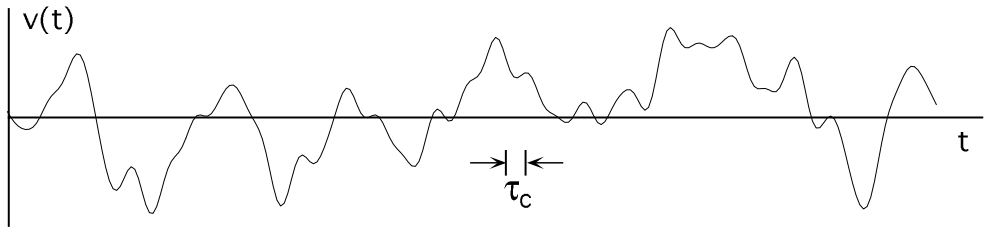
$$p(n;0) = 0 \qquad n \neq 0.$$

. . . . . .

Extended Example: Jointly Gaussian Random Variables

Introduction The purpose of this example is to examine a particular joint probability density and the information that can be extracted from it. We will focus our attention on a physical example that might be encountered in the laboratory. However, the origin of the effect is not of concern to us now. We are interested instead in understanding and manipulating a given probability density.

The System Consider an electronic circuit with all sources (power supplies and signal inputs) turned off. If one looks at a given pair of terminals with an oscilloscope, the voltage appears to be zero at low gain, but at high gain there will be a fluctuating random voltage that might look as follows:



The origin of this "noise" voltage is the random thermal motion of electrons in the components. It is referred to as "thermal noise" or "Johnson noise" and is different from the "shot noise" associated with the quantization of charge. This noise is still present when the sources are turned on and may complicate the detection of a weak signal. Later in the course quantitative expressions will be derived for the amplitude of this type of noise. For the present, observe the following features of the voltage:

1) It has zero mean.

2) Its average magnitude $|v|$ seems relatively well defined and excursions too far above this magnitude are unlikely.

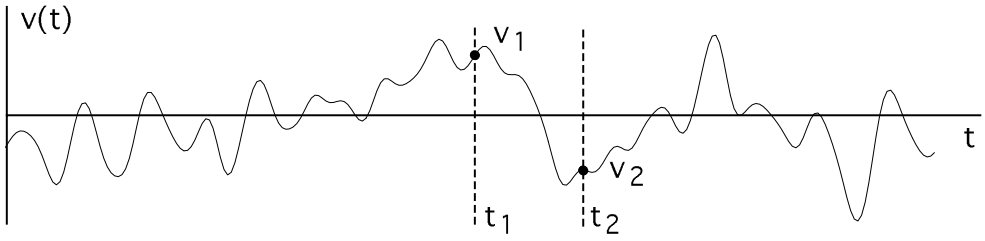
3) The "statistics" do not seem to change with time.

4) There is a "correlation time" τ_c such that over time intervals much less than τ_c the signal does not change appreciably.

5) The voltages at times separated by much more than τ_c seem to be statistically independent.

The noise voltage described above evolves in time and is an example of a *random process*. The study of random processes is a separate field of its own and we will not get involved with it here. Rather, we will simply note that by evaluating the random process at two separate times we can define a pair of random variables. For an important and frequently occurring class of random processes the two variables thus defined will be described by a *jointly Gaussian (or bivariate Gaussian) probability density*. It is this probability density that we will examine here.

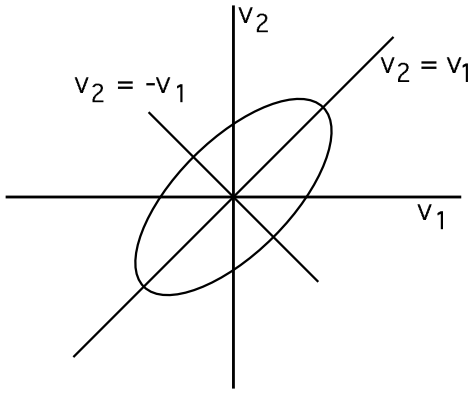
The Joint Probability Density

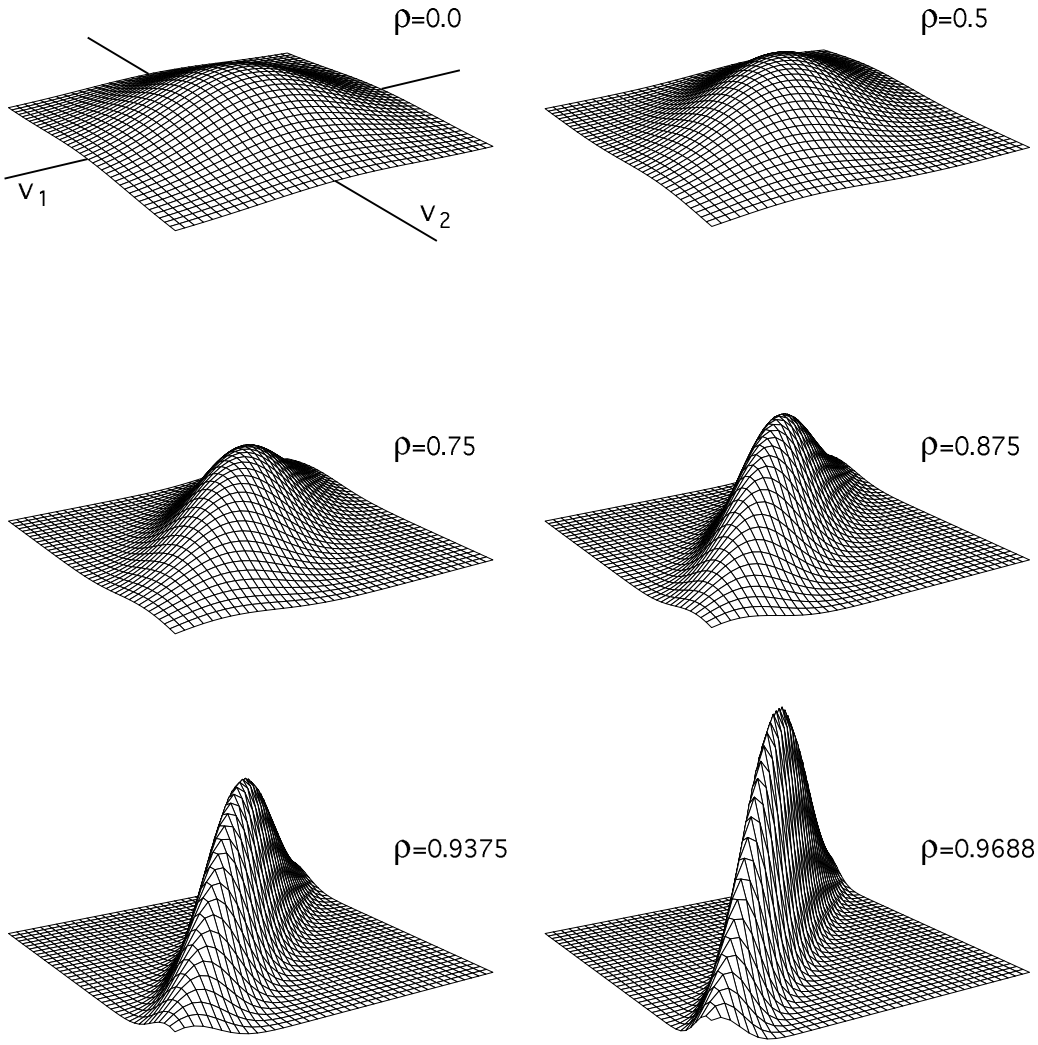


$$p(v_1, v_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp[-\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2\sigma^2(1-\rho^2)}]$$

In the above joint probability density σ and ρ are parameters. σ is a constant, independent of time, which governs the amplitude of the variables. ρ is a function of the time interval between the measurements, $|t_2 - t_1|$; it determines how strongly the two variables are correlated and is referred to as the *correlation coefficient*. The magnitude of ρ is less than or equal to one: $|\rho| \leq 1$. Physically one expects that ρ will be close to one for very small values of $|t_2 - t_1|$ and will decrease to zero for large time separations. We will take this joint probability density as given and examine its properties.

The variables v_1 and v_2 appear as a quadratic form in the exponent. Thus lines of constant probability are ellipses in the v_1, v_2 plane; when $\rho > 0$ the major axis will be along $v_1 = v_2$ and the minor axis will be along $v_1 = -v_2$; for $\rho < 0$ the location of the major and minor axes is reversed. The ellipses are long and narrow for $|\rho| \cong 1$; they become circles when $\rho = 0$.




ρ=0.0
v 1
v 2
ρ=0.5
ρ=0.75
ρ=0.875
ρ=0.9375
ρ=0.9688

Reduction to a Single Variable

$$
\begin{aligned}
p(v_1) &= \int_{-\infty}^{\infty} p(v_1, v_2)\, dv_2 \\
&= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp[-\frac{(v_2^2 - 2\rho v_1 v_2 + \rho^2 v_1^2) + (v_1^2 - \rho^2 v_1^2)}{2\sigma^2(1-\rho^2)}]\, dv_2 \\
&= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp[-\frac{v_1^2}{2\sigma^2}] \underbrace{\int_{-\infty}^{\infty} \exp[-\frac{(v_2 - \rho v_1)^2}{2\sigma^2(1-\rho^2)}]\, dv_2}_{\sqrt{2\pi\sigma^2(1-\rho^2)}} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{v_1^2}{2\sigma^2}] \qquad \{\text{a zero mean gaussian with variance } \sigma^2 \}
\end{aligned}
$$

A similar result is found for $p(v_2)$.

$$p(v_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{v_2^2}{2\sigma^2}]$$

The probability densities for v_1 and v_2 are identical in form, so one concludes that the single time probability densities are independent of time.

Statistical Independence

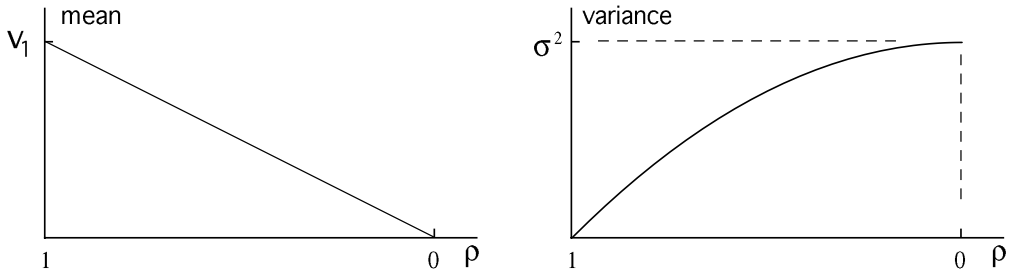
$$p(v_1, v_2) = p(v_1)p(v_2) \qquad \text{only when } \rho = 0$$

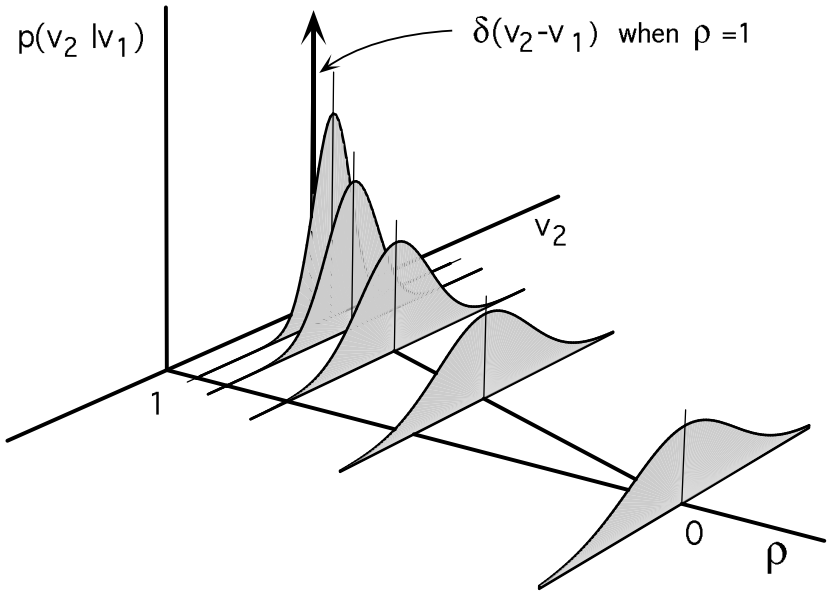
This implies that v_1 and v_2 are not statistically independent unless $\rho = 0$, that is, at large time separations between t_1 and t_2.

Conditional Probability

$$
\begin{aligned}
p(v_2|v_1) &= \frac{p(v_1, v_2)}{p(v_1)} \\
&= \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp[-\left(\frac{(v_2^2 - 2\rho v_1 v_2 + v_1^2)}{2\sigma^2(1-\rho^2)} - \frac{v_1^2(1-\rho^2)}{2\sigma^2(1-\rho^2)}\right)] \\
&= \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp[-\frac{(v_2 - \rho v_1)^2}{2\sigma^2(1-\rho^2)}]
\end{aligned}
$$

This is a Gaussian with mean ρv_1 and variance $\sigma^2(1-\rho^2)$.





Compare these plots of $p(v_2|v_1)$ with an imaginary cut of one of the plots of $p(v_1, v_2)$ by a vertical plane at constant v_1. This will allow you to picture the

relation

$$p(v_2|v_1) \propto p(v_1, v_2).$$

The exact dependence of ρ on $|t_2 - t_1|$ depends on the details of the circuit in which the voltage is measured.

The Correlation Function The correlation function for a random process such as the noise voltage we are discussing is defined as

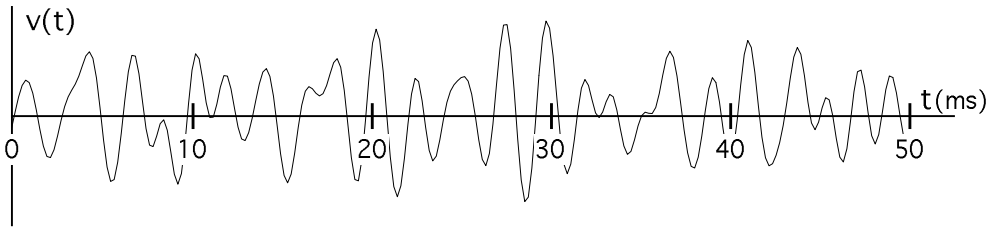
$$R(\tau) \equiv < v(t)v(t+\tau) > .$$

Here we have assumed that the statistics of the process do not change with time so that the correlation function depends only on the time difference, not the actual times themselves. In our notation then $\tau = t_2 - t_1$ and $R(\tau) = R(t_2 - t_1)$. We can now find the correlation function in terms of the parameters appearing in the joint probability density.

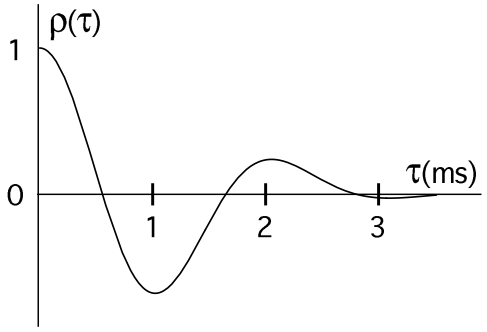
$$\begin{aligned}
< v_1 v_2 > &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1 v_2 p(v_1, v_2)\, dv_1\, dv_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1 v_2 p(v_2|v_1) p(v_1)\, dv_1\, dv_2 \\
&= \int_{-\infty}^{\infty} v_1 p(v_1) \underbrace{\int_{-\infty}^{\infty} v_2 p(v_2|v_1)\, dv_2}_{\text{conditional mean} = \rho(\tau)v_1} dv_1 \\
&= \rho(\tau) \underbrace{\int_{-\infty}^{\infty} v_1^2 p(v_1)\, dv_1}_{<v^2>=\sigma^2}
\end{aligned}$$

Thus the correlation function for the random process can be written in the simple form $R(\tau) = \sigma^2\rho(\tau)$ and the correlation coefficient ρ can be interpreted as the normalized correlation function for the process.

In the figures presented above ρ has been displayed, for simplicity, as positive. However for some random processes ρ may become negative, or even oscillate, as it decays toward zero. Consider the random process that generates the following output.



The frequencies which contribute to the process seem to be peaked around 500 Hz. Thus if the signal were positive at a given time one might expect that it would be negative half a "period" later (1 ms) and, more likely than not, positive again after a delay of 2 ms. This physically expected behavior is reflected in the τ dependence of the correlation coefficient shown below.



One possible random process with these characteristics is the noise voltage of an electronic circuit that is resonant near a single frequency. If the circuit had a very high Q, the correlation function might oscillate many times before falling away to zero.

. . . . . .

3 Functions of a Random Variable

Consider a gas in thermal equilibrium and imagine that the probability density for the speed of an atom $p_v(\zeta)$ is known. The variable of interest, for a particular application, is an atom's kinetic energy, $T \equiv \frac{1}{2}mv^2$. It would be a simple matter to compute the mean value of T, a single number, by using $p_v(\zeta)$ to find $< v^2 >$. But perhaps more information about the random variable T is needed. To completely specify T as a random variable, one needs to know its own probability density $p_T(\eta)$. Finding $p_T(\eta)$ given $p_v(\zeta)$ is a branch of probability known as *functions of a random variable*.

Let $f(x)$ be some known, deterministic function of the random variable x (as $T(v)$ is a function of v above). There are several ways of obtaining the probability density for f, $p_f(\eta)$, from the probability density for x, $p_x(\zeta)$. They are covered in textbooks on random variables. Only one of these methods will be introduced here, one which always works and one which proves to be the most useful in situations encountered in physics. The method consists of a three step process:

A. Sketch $f(x)$ verses x and determine that region of x in which $f(x) < \eta$.

B. Integrate p_x over the indicated region in order to find the cumulative distribution function for f, that is $P_f(\eta)$.

C. Differentiate $P_f(\eta)$ to obtain the density function $p_f(\eta)$.

At first sight it may appear that the integration in step B could lead to computational difficulties. This turns out not to be the case since in most instances one can avoid actually computing the integral by using the following mathematical result.

Mathematics: Derivative of an Integral Expression

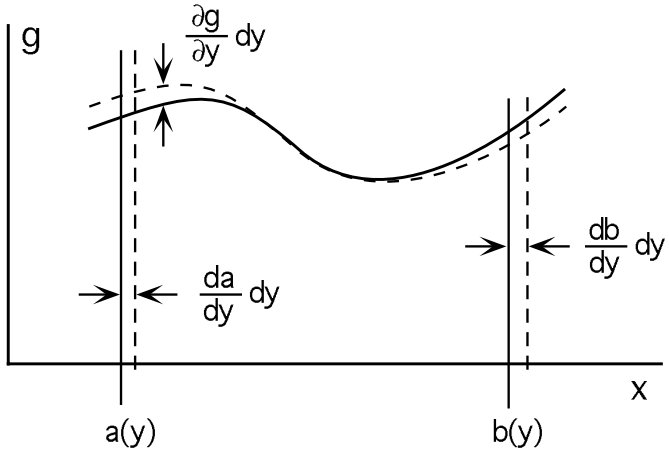
If

$$G(y) \equiv \int_{a(y)}^{b(y)} g(y,x)\,dx,$$

then

$$\frac{dG(y)}{dy} = g(y, x = b(y))\frac{db(y)}{dy} - g(y, x = a(y))\frac{da(y)}{dy} + \int_{a(y)}^{b(y)} \frac{\partial g(y,x)}{\partial y}\,dx.$$

This result follows from examining how the value of the integral changes as the upper limit $b(y)$, the lower limit $a(y)$, and the kernel $g(y, x)$ are separately varied as indicated in the following figure.



. . . . . .

Example: Classical Intensity of Polarized Thermal Light

Given The classical instantaneous intensity of a linearly polarized electromagnetic wave is proportional to the square of the electric field amplitude $\mathcal{E}$.

$$I = a\mathcal{E}^2$$

The fact that the power or intensity in some process is proportional to the square of an amplitude is a common occurrence in nature. For example, it is also found in electrical and acoustic systems. When the radiation field is in thermal equilibrium with matter at a given temperature it is called thermal radiation and the electric field amplitude has a Gaussian density.

$$p(\mathcal{E}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\mathcal{E}^2/2\sigma^2}$$

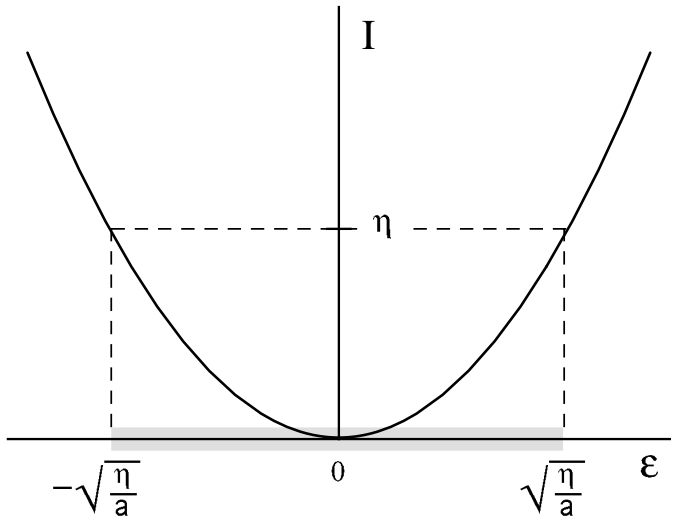
Problem Find the probability density for the instantaneous intensity $p(I)$.

Solution Note that finding the mean intensity is straight forward.

$$< I >= a < \mathcal{E}^2 >= a\sigma^2$$

To find $p(I)$ use the procedure outlined above.

[Step A]



The shaded region of the $\mathcal{E}$ axis shows those values of $\mathcal{E}$ which result in intensities less than η.

[Step B]

$$P_I(\eta) = \int_{-\sqrt{\eta/a}}^{\sqrt{\eta/a}} p_{\mathcal{E}}(\zeta)\, d\zeta$$

It is <u>not</u> necessary to evaluate this integral! Setting it up with the proper (η dependent) limits of integration is sufficient.

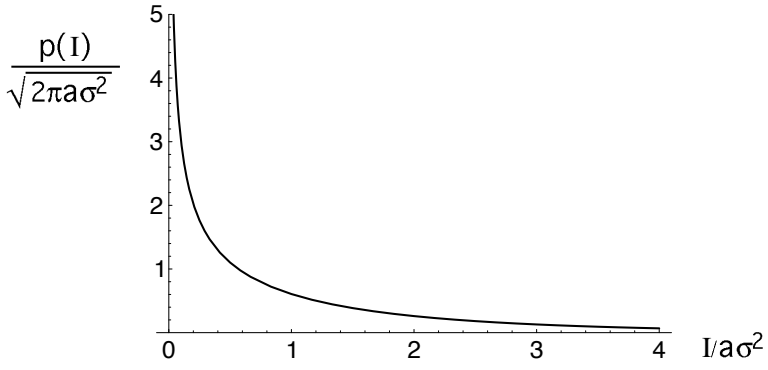
[Step C]

$$\begin{aligned} p_I(\eta) &= \frac{d}{d\eta} P_I(\eta) \\ &= \frac{d}{d\eta} \int_{-\sqrt{\eta/a}}^{\sqrt{\eta/a}} p_{\mathcal{E}}(\zeta)\, d\zeta \\ &= \frac{1}{2}\frac{1}{\sqrt{\eta a}}\, p_{\mathcal{E}}(\sqrt{\eta/a}) + \frac{1}{2}\frac{1}{\sqrt{\eta a}}\, p_{\mathcal{E}}(-\sqrt{\eta/a}) \end{aligned}$$

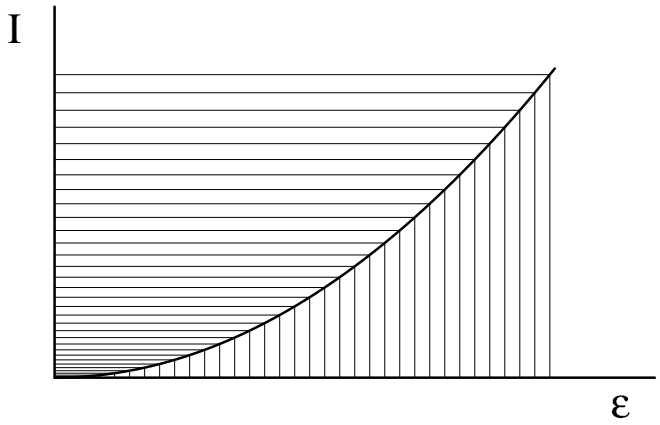
$$= \frac{1}{2}\frac{1}{\sqrt{\eta a}}\left[p_{\mathcal{E}}(\sqrt{\eta/a}) + p_{\mathcal{E}}(-\sqrt{\eta/a})\right]$$

This general result applies to any probability density for the electric field, $p(\mathcal{E})$. For the specific case under consideration, where $p(\mathcal{E})$ is Gaussian, one can proceed to an analytic result.

$$\begin{aligned} p(I) &= \frac{1}{\sqrt{aI}} p_{\mathcal{E}}(\sqrt{I/a}) \qquad \text{since } p_{\mathcal{E}}(\zeta) \text{ is even} \\ &= \frac{1}{\sqrt{2\pi a\sigma^2 I}} \exp[-\frac{I}{2a\sigma^2}] \qquad I > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad I < 0 \end{aligned}$$



The integrable singularity at $I = 0$ is a consequence of two features of the problem. The parabolic nature of $I(\mathcal{E})$ emphasizes the low $\mathcal{E}$ part of $p(\mathcal{E})$. Imagine a series of equally spaced values of $\mathcal{E}$. The corresponding values of I would be densest at small I.



This type of mapping will be treated quantitatively later in the course when the concept of "density of states" is introduced. The Gaussian nature of $p(\mathcal{E})$ gives a finite weight to the values of $\mathcal{E}$ close to zero. The result of the two effects is a divergence in $p(I)$ at $I = 0$.

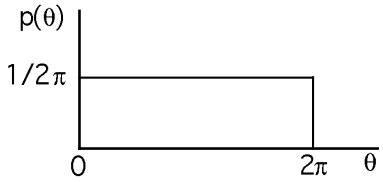
. . . . . .

Example: Harmonic Motion

Given

$$p(\theta) = \frac{1}{2\pi} \qquad 0 \le \theta < 2\pi$$

$$= 0 \qquad \text{elsewhere}$$

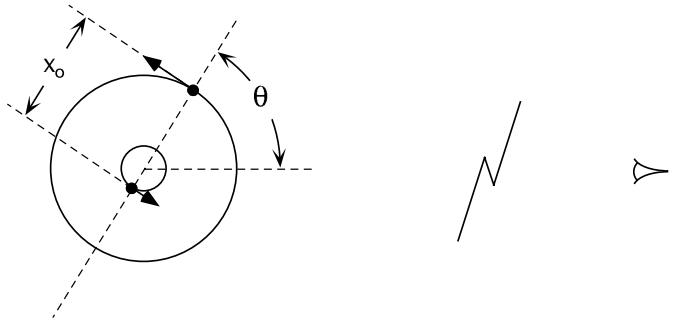


$$x(\theta) = x_0 \sin(\theta)$$

One physical possibility is that x represents the displacement of a harmonic oscillator of fixed total energy but unknown phase or starting time:

$$x = x_0 \sin(\underbrace{\omega t + \phi}_{\theta}).$$

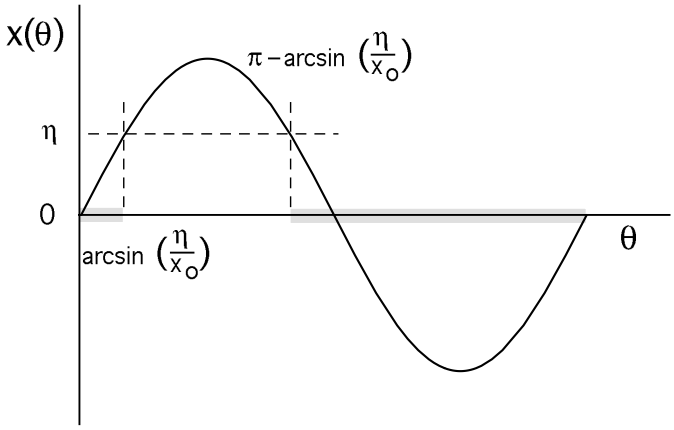
Another possibility is a binary star system observed from a distant location in the plane of motion. Then x could represent the apparent separation at an arbitrary time.



Problem Find $p(x)$.

Solution Consider separately the two regions $x > 0$ and $x < 0$. For $x > 0$

[Step A]



[Step B]

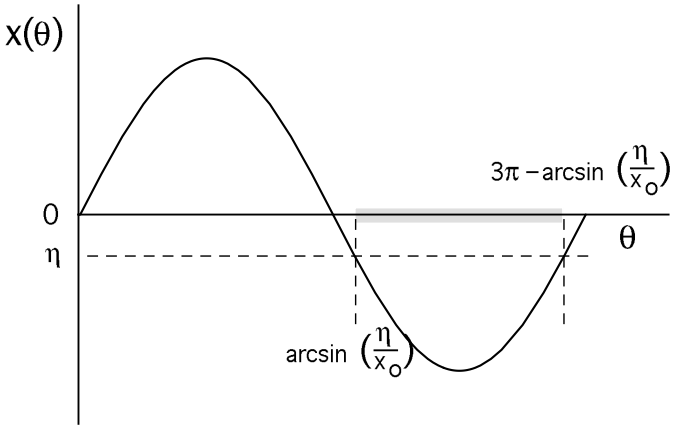
$$\begin{aligned} P_x(\eta) &= \int_{\text{shaded}} p_\theta(\zeta)\, d\zeta = 1 - \int_{\text{unshaded}} p_\theta(\zeta)\, d\zeta \\ &= 1 - \int_{\arcsin(\eta/x_0)}^{\pi - \arcsin(\eta/x_0)} (\frac{1}{2\pi})\, d\zeta \\ &= \frac{1}{2} + \frac{1}{\pi} \arcsin(\eta/x_0) \end{aligned}$$

[Step C]

$$\begin{aligned} p_x(\eta) &= \frac{d}{d\eta} P_x(\eta) \\ &= \frac{1}{\pi} \frac{1}{\sqrt{1-(\eta/x_0)^2}} \frac{1}{x_0} \\ &= \frac{1}{\pi} \frac{1}{\sqrt{x_0^2-\eta^2}} \qquad\qquad 0 \le \eta < x_0 \end{aligned}$$

For $x < 0$

[Step A]



[Step B]

$$\begin{aligned} P_x(\eta) &= \int_{\arcsin(\eta/x_0)}^{3\pi-\arcsin(\eta/x_0)} (\frac{1}{2\pi})\, d\zeta \\ &= \frac{3}{2} - \frac{1}{\pi} \arcsin(\eta/x_0) \end{aligned}$$

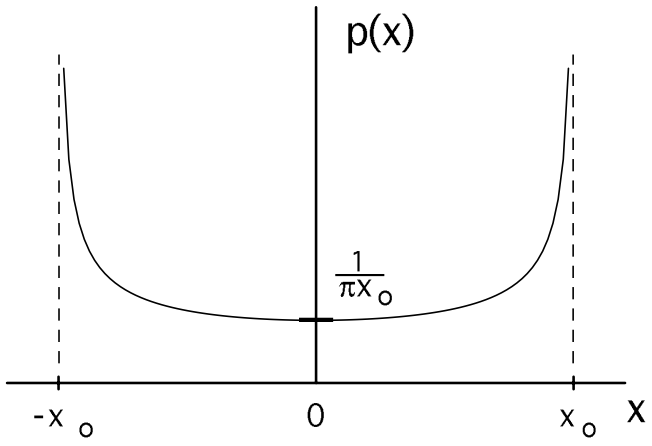
[Step C]

$$p_x(\eta) = \frac{d}{d\eta} P_x(\eta)$$

$$= -\frac{1}{\pi}\,\frac{-1}{\sqrt{1-(\eta/x_0)^2}}\,\frac{1}{x_0}$$

$$= \frac{1}{\pi}\,\frac{1}{\sqrt{x_0^2-\eta^2}} \qquad -x_0 < \eta \le 0$$

Note: the extra -1 in the second line comes about since the derivative of $\arcsin(\zeta)$ is negative in the region $\pi/2 < \arcsin(\zeta) < 3\pi/2$. One could also have obtained the result for $x < 0$ from that for $x > 0$ by symmetry.



The divergence of $p(x)$ at the "turning points" $x = \pm x_0$ can be demonstrated visually by a simple experiment with a pencil as explained in a homework problem.

. . . . . .

The method of functions of a random variable can can also be applied in cases where the function in question depends on two or more random variables with known probabilities, that is one might want to find $p(f)$ for a given $f(x, y)$ and $p(x, y)$.

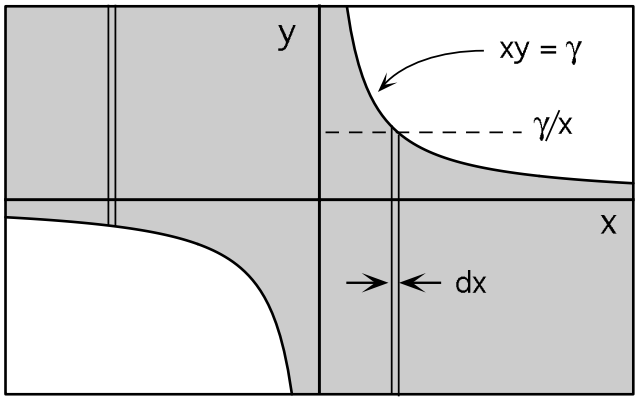
Example: Product of Two Random Variables

Given x and y are defined over all space and $p_{x,y}(\zeta, \eta)$ is known.

Problem Find $p(z)$ where $z \equiv xy$.

Solution Consider the regions $z > 0$ and $z < 0$ separately.
For $z > 0$

[Step A]



z is less than the specific positive value γ in the shaded regions of the x, y plane.

[Step B]

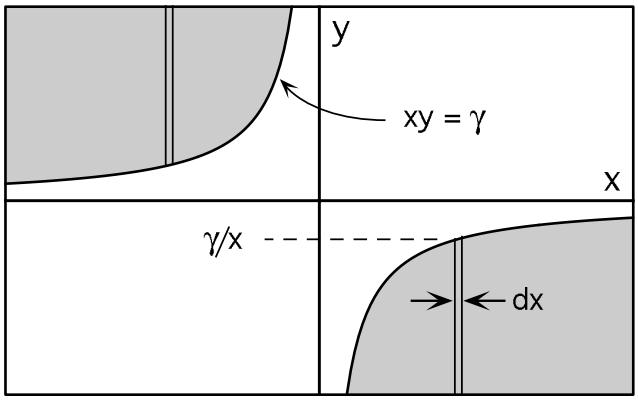
$$P_z(\gamma) = \int_{-\infty}^{0} d\zeta \int_{\gamma/\zeta}^{\infty} d\eta \, p_{x,y}(\zeta, \eta) + \int_{0}^{\infty} d\zeta \int_{-\infty}^{\gamma/\zeta} d\eta \, p_{x,y}(\zeta, \eta)$$

[Step C]

$$\begin{aligned} p_z(\gamma) &= \frac{d}{d\gamma} P_z(\gamma) \\ &= -\int_{-\infty}^{0} \frac{d\zeta}{\zeta} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) + \int_{0}^{\infty} \frac{d\zeta}{\zeta} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) \\ &= \int_{-\infty}^{\infty} \frac{d\zeta}{|\zeta|} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) \qquad\qquad z > 0 \end{aligned}$$

For $z < 0$

[Step A]



Now z is less than a specific negative value γ in the shaded regions of the x, y plane.

[Step B]

$$P_z(\gamma) = \int_{-\infty}^{0} d\zeta \int_{\gamma/\zeta}^{\infty} d\eta \, p_{x,y}(\zeta, \eta) + \int_{0}^{\infty} d\zeta \int_{-\infty}^{\gamma/\zeta} d\eta \, p_{x,y}(\zeta, \eta)$$

[Step C]

$$\begin{aligned} p_z(\gamma) &= \frac{d}{d\gamma} P_z(\gamma) \\ &= -\int_{-\infty}^{0} \frac{d\zeta}{\zeta} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) + \int_{0}^{\infty} \frac{d\zeta}{\zeta} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) \\ &= \int_{-\infty}^{\infty} \frac{d\zeta}{|\zeta|} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) \qquad\qquad z < 0 \end{aligned}$$

This is the same expression which applies for positive z, so in general one has

$$p_z(\gamma) = \int_{-\infty}^{\infty} \frac{d\zeta}{|\zeta|} \, p_{x,y}(\zeta, \frac{\gamma}{\zeta}) \qquad \text{for all } z < 0.$$

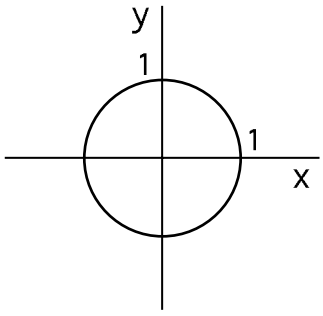
. . . . . .

A common use of functions of a random variable is to carry out a change in variables.

Example: Uniform Circular Disk Revisited

Given The probability of finding an event in a two dimensional space is uniform inside a circle of radius 1 and zero outside of the circle.

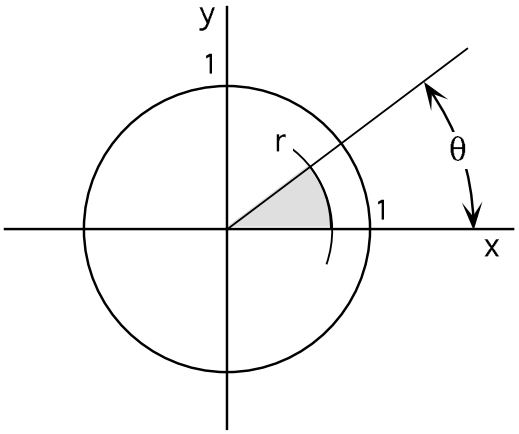
$$\begin{aligned} p(x,y) &= 1/\pi \quad x^2+y^2 \le 1 \\ &= 0 \quad x^2+y^2 > 1 \end{aligned}$$



Problem Find the joint probability density for the radial distance r and the polar angle θ.

Solution

[Step A]



The shaded area indicates the region in which the radius is less than r and the angle is less than θ.

[Step B]

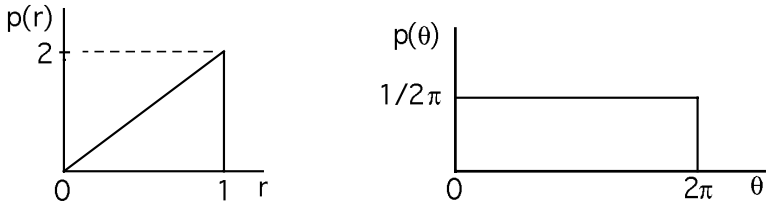
$$\begin{aligned} P(r,\theta) &= \int_{\text{shaded area}} p(x,y)\,dx\,dy \\ &= \underbrace{\frac{1}{\pi}}_{p(x,y)} \underbrace{\pi r^2}_{\text{area of disk}} \underbrace{\frac{\theta}{2\pi}}_{\text{fraction shaded}} = \frac{\theta r^2}{2\pi} \end{aligned}$$

[Step C]

$$p(r,\theta) = \frac{\partial}{\partial r}\frac{\partial}{\partial \theta} P(r,\theta) = \frac{r}{\pi}$$

$$\begin{aligned} p(r) &= \int_0^{2\pi} p(r,\theta)\,d\theta = \int_0^{2\pi} \frac{r}{\pi}\,d\theta = 2r \qquad r < 1 \\ &= 0 \qquad\qquad r > 1 \end{aligned}$$

$$\begin{aligned} p(\theta) &= \int_0^1 p(r,\theta)\,dr = \int_0^1 \frac{r}{\pi}\,dr = \frac{1}{2\pi} \qquad 0 \le \theta < 2\pi \\ &= 0 \qquad\qquad \text{elsewhere} \end{aligned}$$



Note that r and θ are statistically independent (which is not the case for x and y) since

$$p(r)p(\theta) = (2r)(\frac{1}{2\pi}) = \frac{r}{\pi} = p(r,\theta).$$

. . . . . .

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:

$$\begin{aligned} < S_n > &= \int (x_1 + x_2 + \cdots + x_n) \underbrace{p(x_1, x_2, \ldots, x_n)}_{p_1(x_1)p_2(x_2)\cdots p_n(x_n)} \, dx_1 \, dx_2 \cdots dx_n \\ &= \sum_{i=1}^{n} [\underbrace{\int x_i p_i(x_i) \, dx_i}_{E_i}] [\prod_{i \neq j} \underbrace{\int p_j(x_j) \, dx_j}_{1}] \\ &= \sum_{i=1}^{n} E_i \end{aligned}$$

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

$$\begin{aligned} Var(S_n) &= < (S_n - < S_n >)^2 > \\ &= < (x_1 - E_1 + x_2 - E_2 + \cdots + x_n - E_n)^2 > \end{aligned}$$

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

$$\begin{aligned} < (x_i - E_i)(x_j - E_j) > &= < x_i x_j > - < x_i > E_j - < x_j > E_i + E_i E_j \\ &= < x_i x_j > - E_i E_j \\ &= \begin{cases} < x_i^2 > - E_i^2 = \sigma_i^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \end{aligned}$$

Therefore,

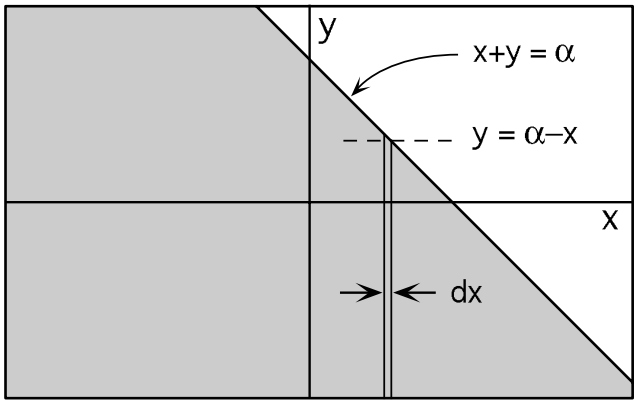
$$Var(S_n) = \sum_{i=1}^{n} \sigma_i^2 .$$

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

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

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,

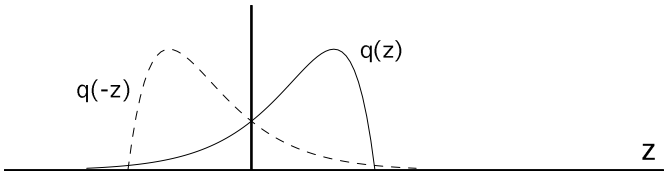
$$a \otimes (b + c) = a \otimes b + a \otimes c,$$

and associative,

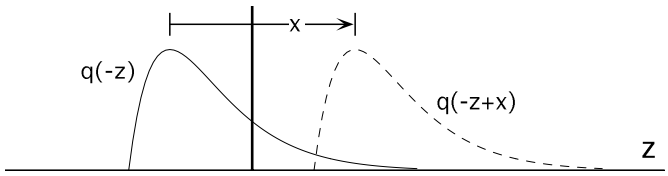
$$a \otimes (b \otimes c) = (a \otimes b) \otimes c.$$

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

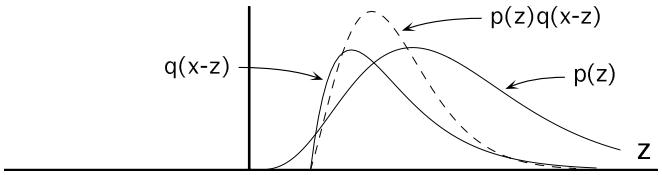
1. FLIP $q(z)$ ABOUT THE ORIGIN OF ITS ARGUMENT TO FORM THE MIRROR IMAGE $q(-z)$.



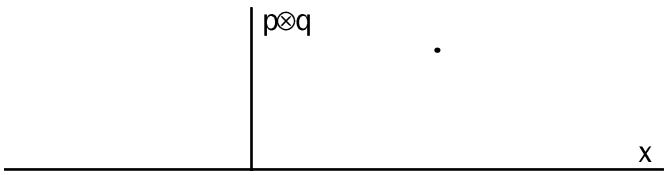
2. SHIFT $q(-z)$ TO THE RIGHT BY AN AMOUNT x TO FORM $q(-z+x)$.



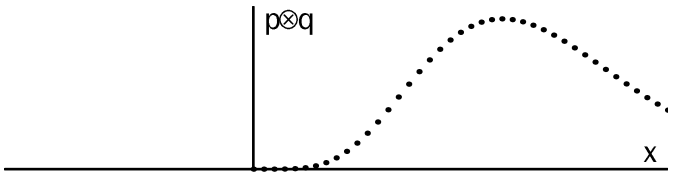
3. MULTIPLY $q(x-z)$ BY $p(z)$.



4. INTEGRATE THE PRODUCT AND PLOT THE RESULT.



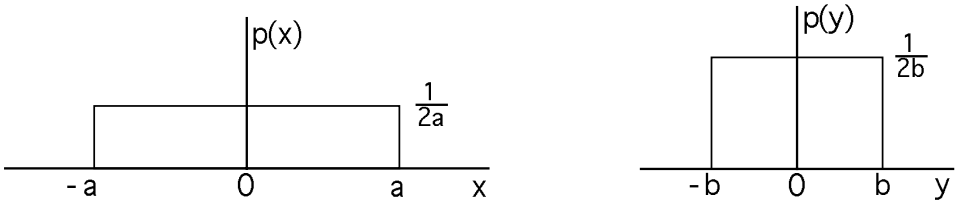
5. REPEAT 2 THROUGH 4 FOR DIFFERENT VALUES OF x.



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

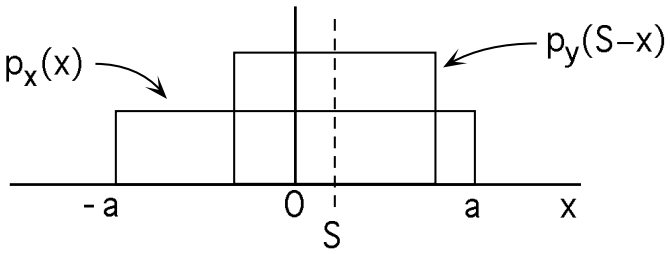
Given x and y are two statistically independent random variables, uniformly distributed in the regions $|x| \leq a$ and $|y| \leq b$.



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

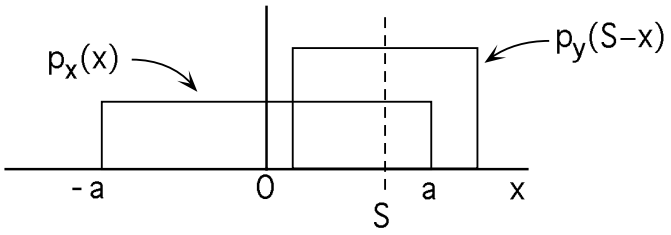
Solution The form of the integral will depend on the value of S. Three separate regions need to be considered. Assume as indicated above that $b < a$.

$0 < S < a - b$



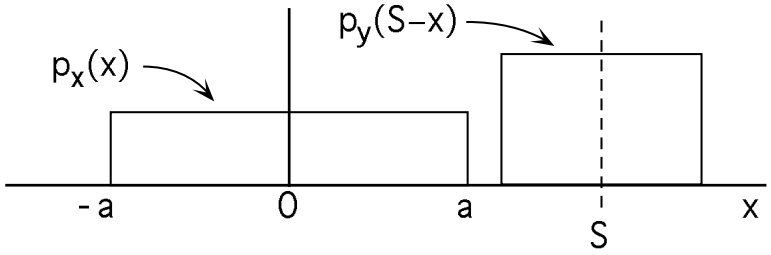
$$p(S) = \int_{S-b}^{S+b} (\frac{1}{2a})(\frac{1}{2b})\, dx = \frac{1}{2a}$$

$a - b < S < a + b$



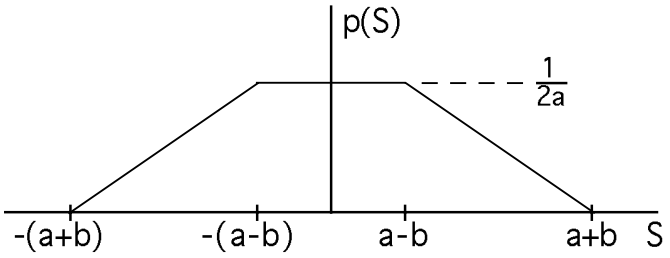
$$p(S) = \int_{S-b}^{a} (\frac{1}{2a})(\frac{1}{2b})\, dx = \frac{a + b - S}{(2a)(2b)}$$

$a + b < S$

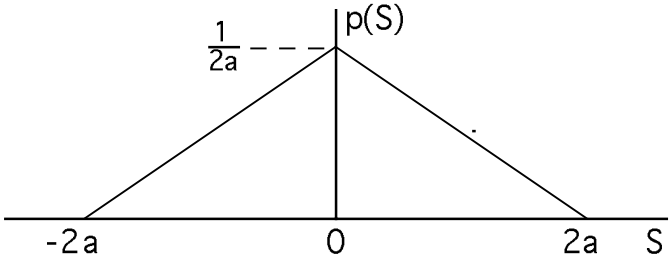


$p(S) = 0$ since there is no overlap

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

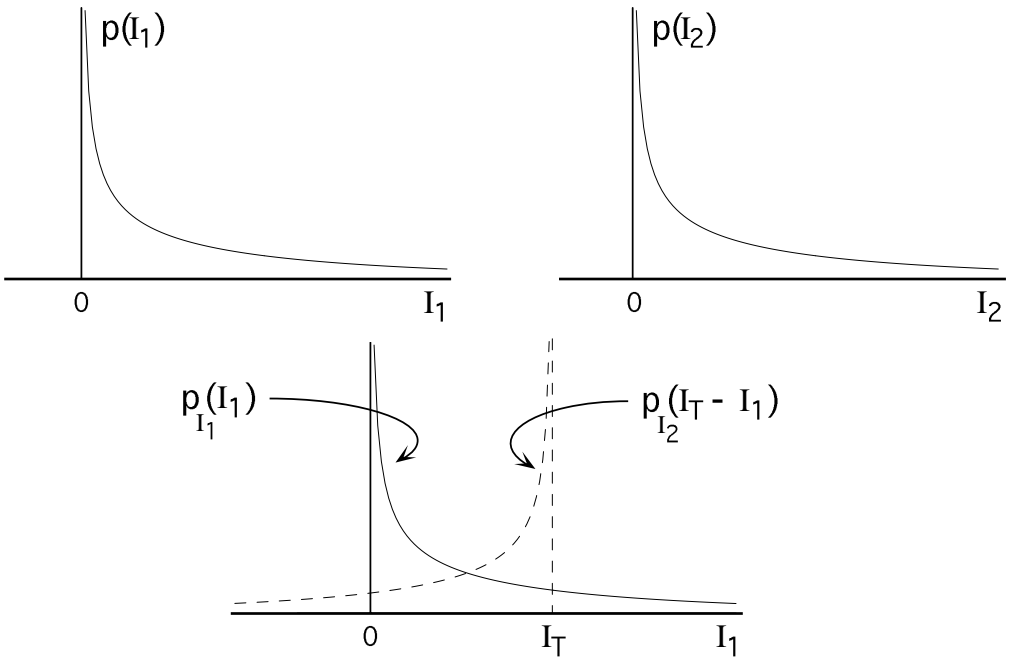
Given An unpolarized beam of thermal light can be viewed as two statistically independent beams of equal average intensity, polarized at right angles to each other. The total intensity I_T is the sum of the intensities of each of the two polarized components, I_1 and I_2.

Problem Find the probability density for I_T using the result from a previous example that

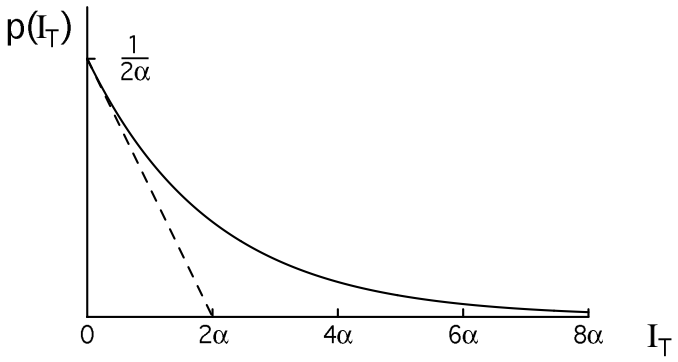
$$\begin{aligned} p(I_1) &= \frac{1}{\sqrt{2\pi\alpha I_1}} \exp[-I_1/2\alpha] \qquad I_1 > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad I_1 < 0 \end{aligned}$$

where $\alpha =< I_1 >$.

Solution Since the light is thermal and unpolarized, the intensity in each of the two polarization directions has the same density: $p_{I_1}(\zeta) = p_{I_2}(\zeta)$.



$$
\begin{aligned}
p(I_T) &= \int_{-\infty}^{\infty} p_{I_1}(I_1) p_{I_2}(I_T - I_1)\, dI_1 \\
&= \int_0^{I_T} \left(\frac{\exp[-I_1/2\alpha]}{\sqrt{2\pi\alpha I_1}} \right) \left(\frac{\exp[-(I_T - I_1)/2\alpha]}{\sqrt{2\pi\alpha(I_T - I_1)}} \right) dI_1 \\
&= \frac{1}{2\pi\alpha} \exp[-I_T/2\alpha] \int_0^{I_T} [I_1(I_T - I_1)]^{-1/2}\, dI_1 \\
&= \frac{1}{2\pi\alpha} \exp[-I_T/2\alpha] \underbrace{\int_{-I_T/2}^{I_T/2} (\frac{1}{4} I_T^2 - x^2)^{-1/2}\, dx}_{\pi} \qquad \text{using} \quad x \equiv I_1 - \frac{1}{2} I_T \\
&= \frac{1}{2\alpha} \exp[-I_T/2\alpha] \qquad I_T \geq 0 \\
&= 0 \qquad\qquad\qquad\qquad I_T < 0
\end{aligned}
$$



. . . . . .

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.

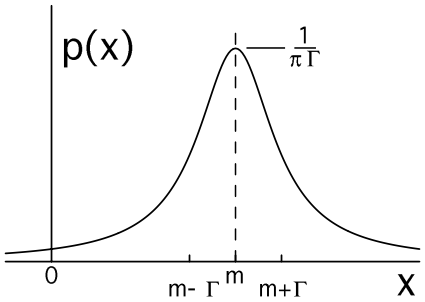
The sum of two S.I. Poisson random variables is also Poisson. Here again, knowing that the result is Poisson allows one to determine the parameters in the sum density. Recall that a Poisson density is completely specified by one number, the mean, and the mean of the sum is the sum of the means.

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

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

$$m =< x >$$

Γ = 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.

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

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

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

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

The second of the two terms in { } brackets in the last line produces a constant which factors out of the integral. The integral over ζ 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.

$$\begin{aligned}p(S) &= (2\pi\sigma_x\sigma_y)^{-1}\ \exp[-\frac{a^2}{2(\sigma_x^2+\sigma_y^2)}]\ \left(2\pi\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}\right)^{1/2}\\ &= \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}}\ \exp[-\frac{(S-(E_x+E_y))^2}{2(\sigma_x^2+\sigma_y^2)}]\end{aligned}$$

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.

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

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

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

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

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

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 σ_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.

- σ_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 sum of δ functions cannot converge to a continuous distribution unless it is averaged appropriately. What actually happens in such cases is that the envelope of the comb of δ 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.

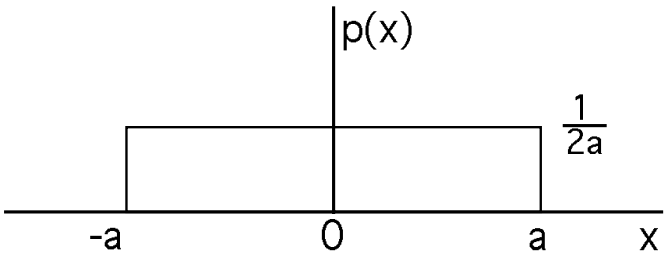
- The Central Limit Theorem may work even if the individual members of the sum are not identically distributed. A prerequisite is that no individual member should dominate the sum. The necessity for this restriction can be visualized easily. Consider adding 6 variables, each uniformly distributed between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

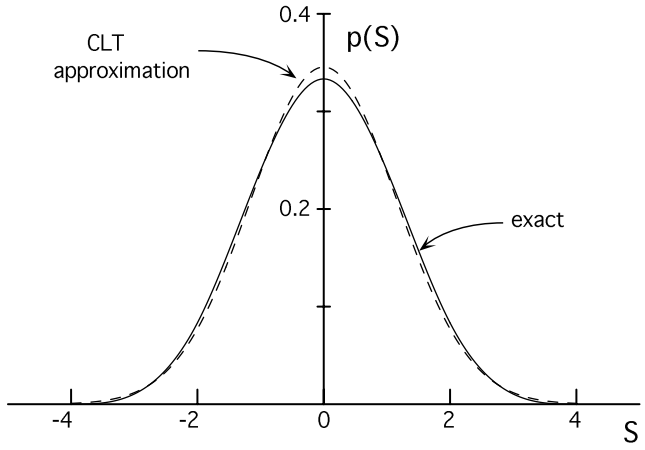
Consider a random variable x which is uniformly distributed between $-a$ and a.



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.

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

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.

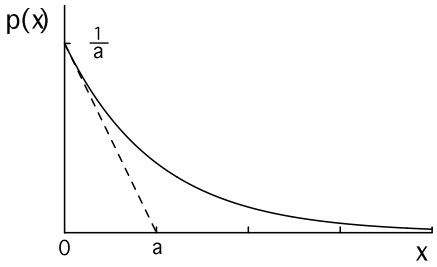


. . . . . .

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.

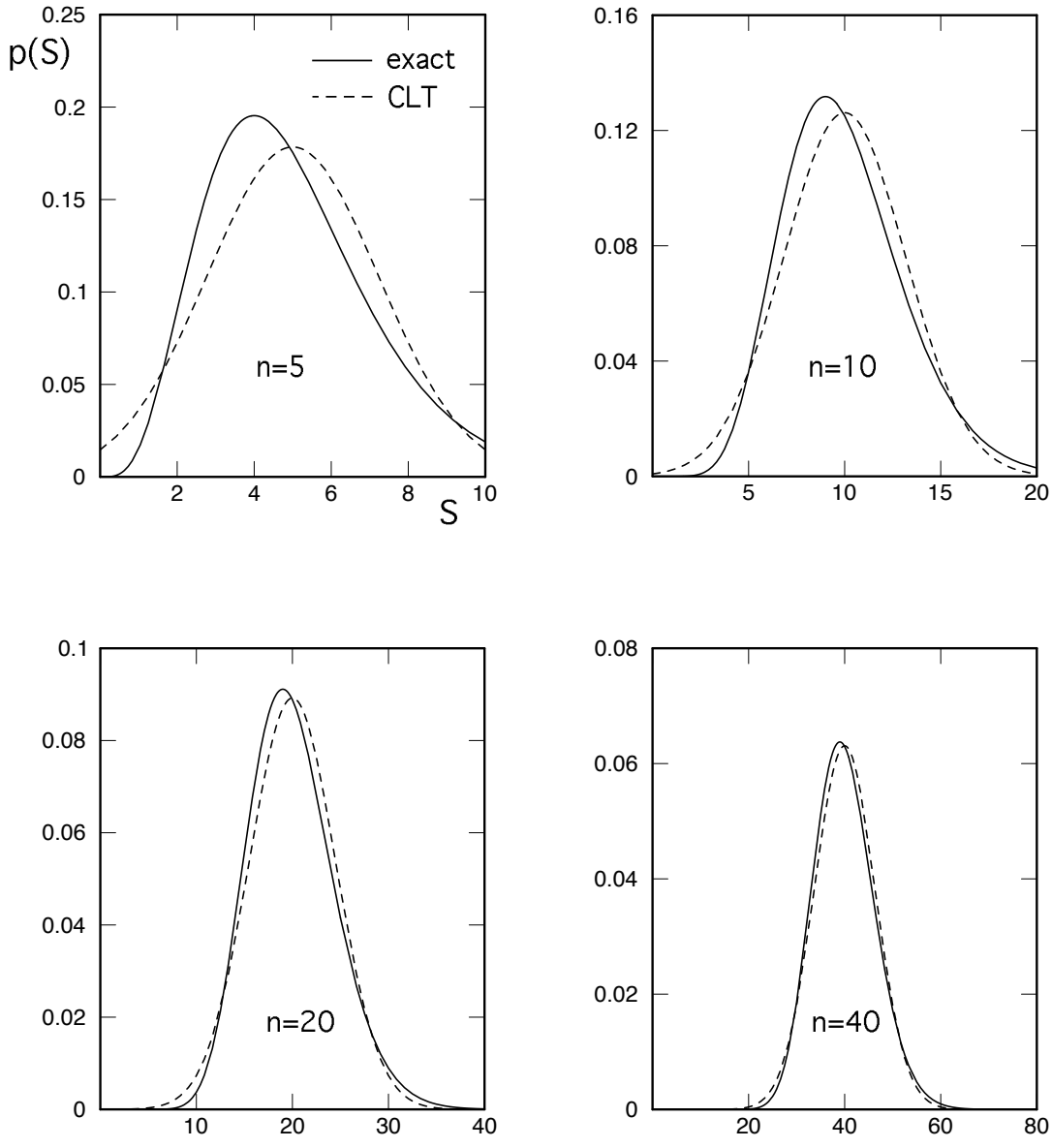
$$\begin{aligned} p(x) &= \frac{1}{a}\, e^{-x/a} \qquad x > 0 \\ &= 0 \qquad\qquad\quad x < 0 \end{aligned}$$



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.

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

This result is shown below for $a = 1$ and several values of n. The Gaussian approximation is also shown for comparison.


p(S)
exact
CLT
n=5
n=10
n=20
n=40
S

. . . . . .

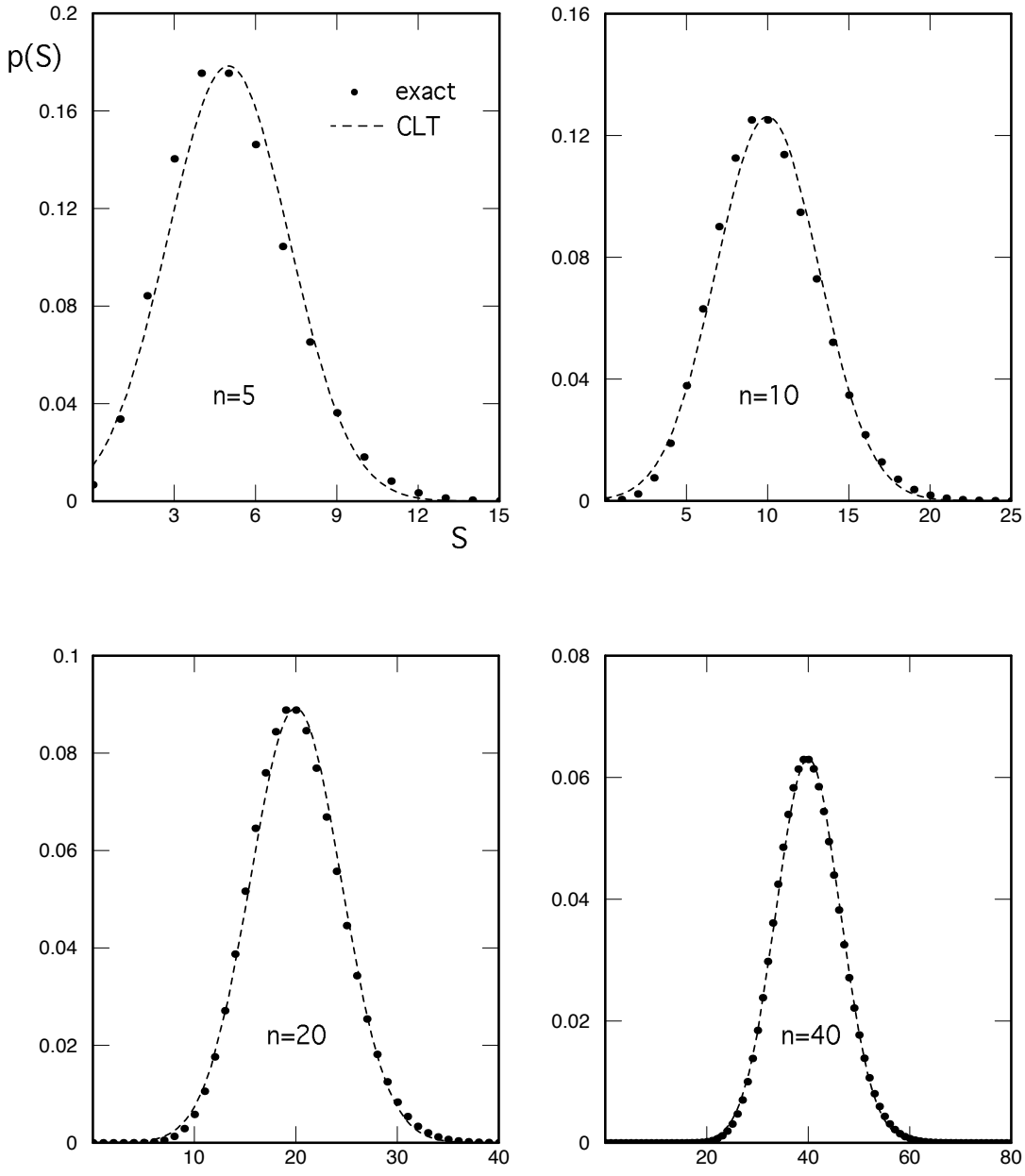
Example: Poisson Density for Large $< n >$

The Central Limit Theorem applied to sums of discrete random variables states that the envelope of the train of δ 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 single 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 $S = 20$.


p(S)
exact
CLT
n=5
n=10
n=20
n=40
S