Lecture Notes on Probability
for
8.044: Statistical Physics I

Thomas J. Greytak

Physics Department

Massachusetts Institute of Technology

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 $_{92}U^{237}$, 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 nth member of the ensemble produces the variable x_n.
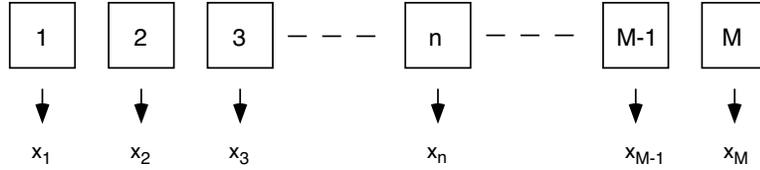
Imagine compiling a histogram indicating the number of systems for which x falls within certain intervals. The *probability density function* $p_x(\zeta)$ is defined in terms of such a histogram:

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

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.

$$p_x(\zeta) \ge 0 \;\; \text{for all } \zeta$$

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

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

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*.

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

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

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

$P_x(\zeta)$ 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

<u>Given</u> [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

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

<u>Problem</u> Examine and discuss $p(t)$ and $P(t)$.

<u>Solution</u> 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.
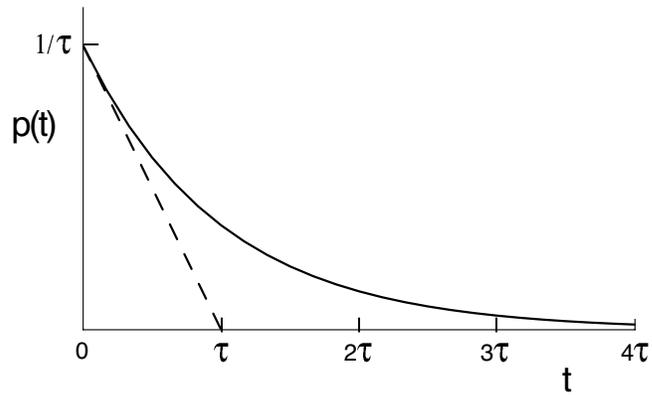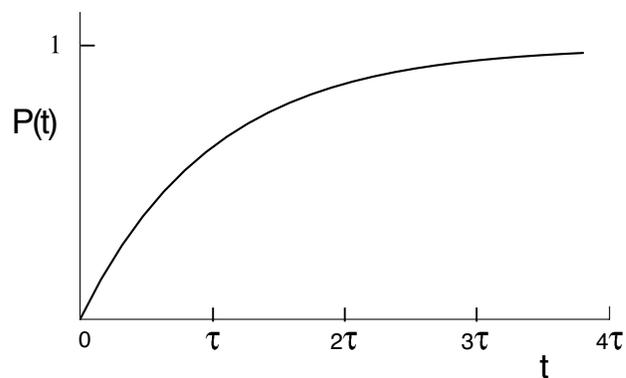
The density is normalized.

$$\int_{-\infty}^{\infty} p(t)\,dt \;\; = \;\; \int_0^\infty e^{-t/\tau}\,\frac{dt}{\tau} = \int_0^\infty e^{-y}\,dy$$

$$= \;\; 1$$

The cumulative function is found by integration.

$$P(t) \;\; = \;\; \int_{-\infty}^{t} p_t(\zeta)\,d\zeta = \int_0^t e^{-\zeta/\tau}\,\frac{d\zeta}{\tau} = \int_0^{t/\tau} e^{-y}\,dy$$

$$= \;\; 1 - e^{-t/\tau} \qquad t \geq 0$$

The density can be recovered from the cumulative by differentiation.

$$p(t) \;=\; \frac{d}{dt}\,P(t) = -(\frac{-1}{\tau})\,e^{-t/\tau}$$

$$=\; \tau^{-1}e^{-t/\tau} \qquad t \geq 0$$

· ·

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.
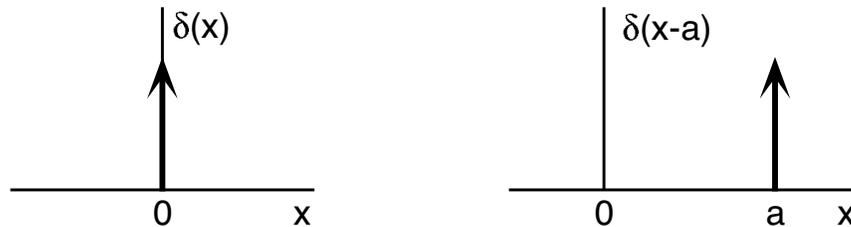
Mathematics: The δ Function

The δ function has the following properties.

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

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

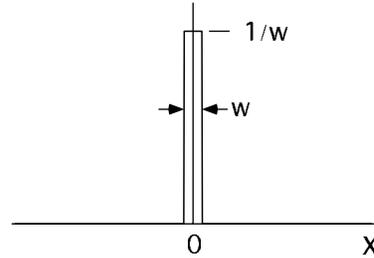$$\int_{-\epsilon}^{\epsilon} \delta(x)\,dx = 1 \qquad \text{for all } \epsilon > 0$$

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.
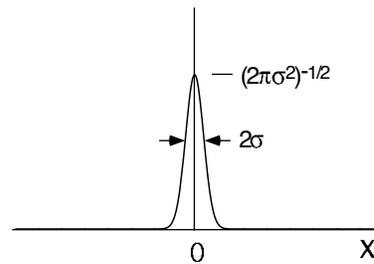
Rectangular Model

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

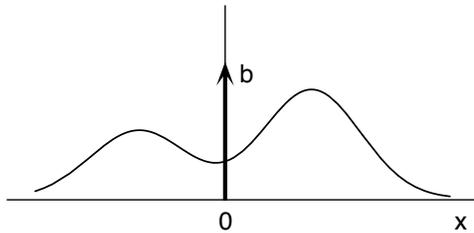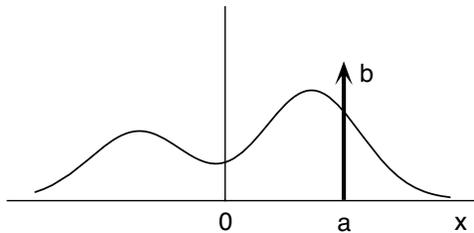Gaussian Model

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

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.

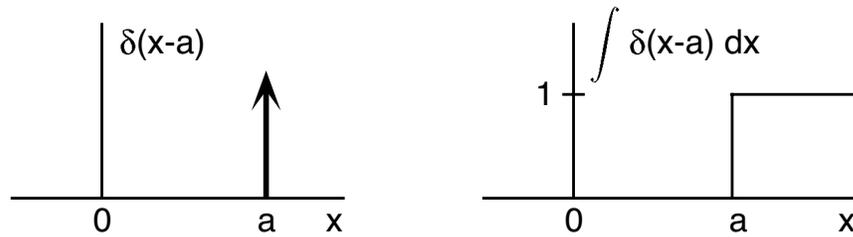$$\int_{-\infty}^{\infty} f(x)(b\,\delta(x))\,dx = bf(0)$$

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

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

<u>Given</u> 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.

$$
\begin{aligned}
p_n(0) &= 0 & \text{by definition} \\
p_n(1) &= 1/3 & \text{a single star, such as the Sun (unless you believe in Nemisis)} \\
p_n(2) &= 1/2 & \text{a binary, as in Sirius} \\
p_n(3) &= 1/6 & \text{a trinary, as in } \alpha \text{ Centari} \\
p_n(> 3) &= \text{negligible}
\end{aligned}
$$

<u>Problem</u> Examine and discuss the probabilities.

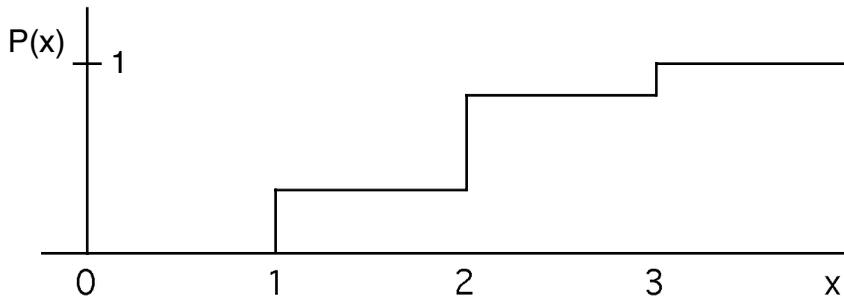<u>Solution</u> The probabilities can be displayed on a graph using δ functions.
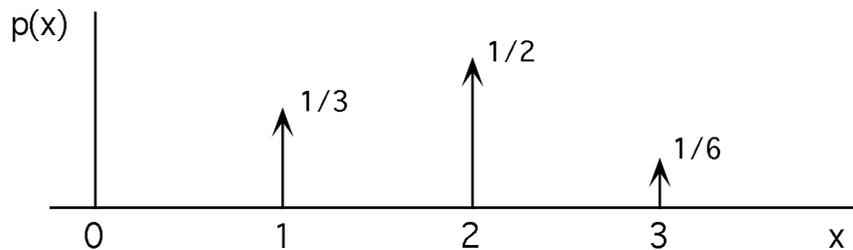
Here x is used as the variable to emphasize the continuous nature of the horizontal axis.

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

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

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

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.
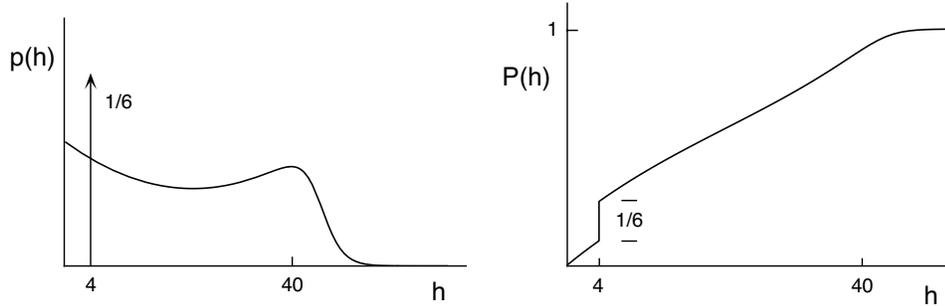
· ·

Example: A Logged Forest

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

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

<u>Solution</u> 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

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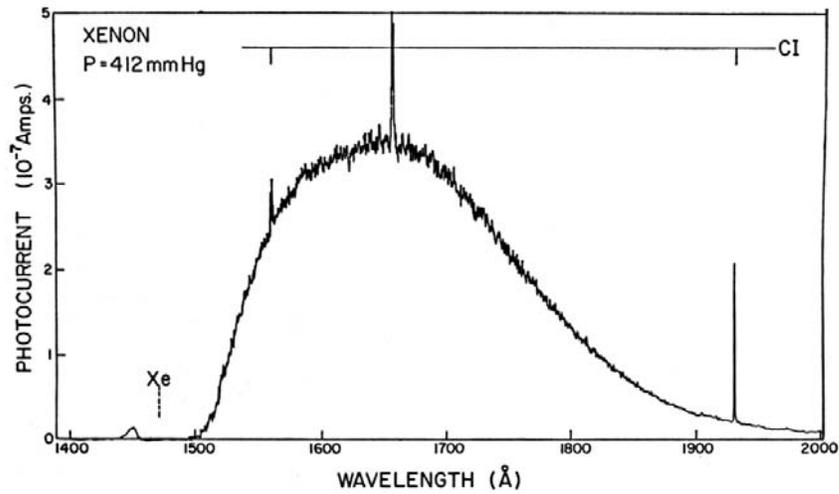

<u>Problem</u> Discuss the nature of the spectrum and cast the measurements in the form of a probability density.

<u>Solution</u> 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 <u>electron</u> energy spectrum in the atom do not overlap; however, the discrete and continuous parts of the <u>photon</u> 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 $\zeta + d\zeta$.
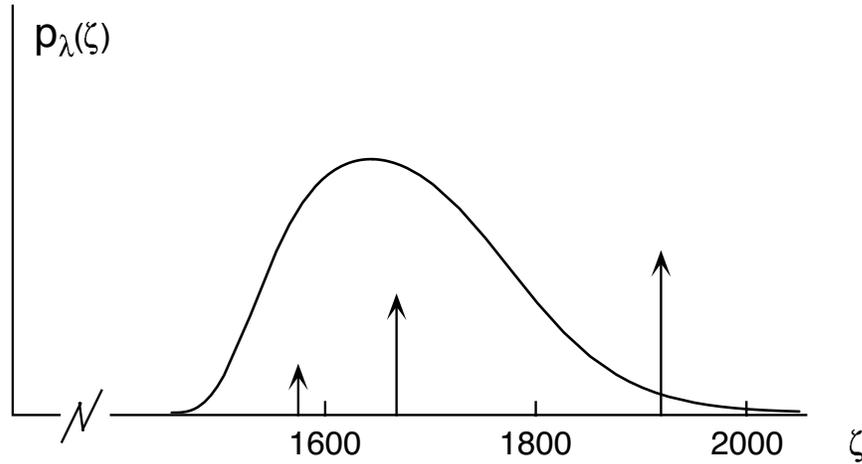
Often it is more physical to think about the energy of a photon rather than its wavelength. The energy is inversely proportional to the wavelength, $E = 2\pi\hbar c/\lambda$, so a simple linear scaling of the horizontal axis will not give the probability of finding a photon of energy E. A technique will 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*, $\text{Var}(x)$, is a quadratic measure of the fluctuations of x about its mean.

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

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

$$
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(x-a)^2}{2\sigma^2}]
$$
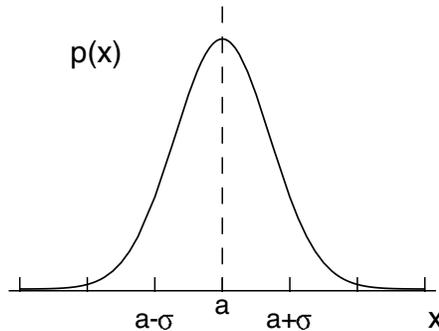
are frequently encountered in science and engineering. They occur in thermal equilibrium when the variable makes a quadratic contribution to the energy of the system (the x component of the velocity of an atom in a gas for example). They also occur when the variable is the sum of a large number of small, 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 $P(x)$.
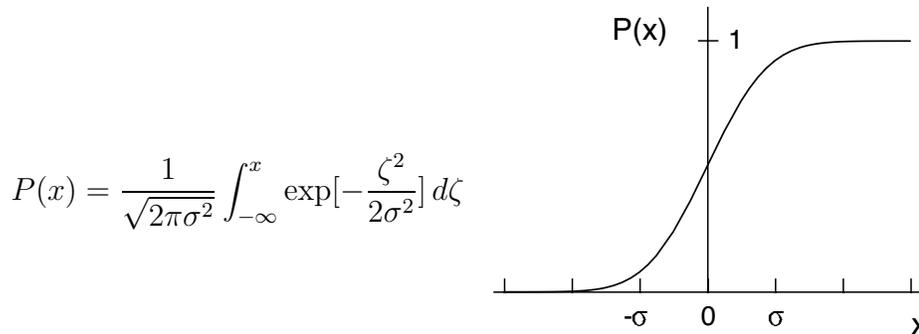
<u>Solution</u>

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

$$= \; \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp[-\frac{(x-a)^2}{2\sigma^2}] \, dx$$

Let $\eta = x - a$; then $d\eta = dx$

$$= \; a \; \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp[-\frac{\eta^2}{2\sigma^2}] \, d\eta}_{1} + \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \eta \exp[-\frac{\eta^2}{2\sigma^2}] \, d\eta}_{0}$$

$$= \; a$$

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

$$= \; \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\eta^2 + 2\eta a + a^2) \exp[-\frac{\eta^2}{2\sigma^2}] \, d\eta$$

$$= \; \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{\infty} \eta^2 \exp[-\frac{\eta^2}{2\sigma^2}] \, d\eta}_{\sqrt{2\pi\sigma^3}} + 0 \; + a^2$$

$$= \; \sigma^2 + a^2$$

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

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

<u>Given</u> 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.
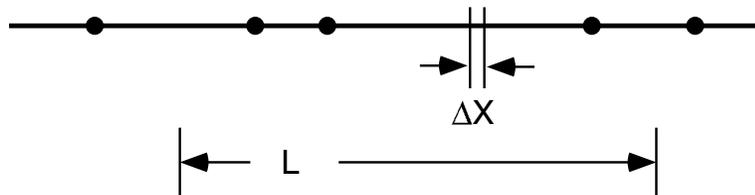
Under these circumstances the probability that exactly n events will occur in an interval of length L is given by the Poisson probability:
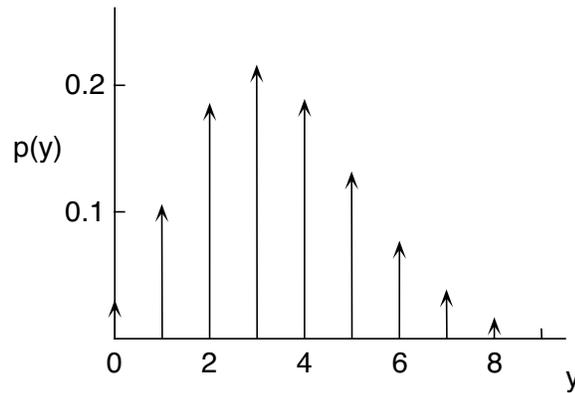
$$p(n) = \frac{1}{n!}(rL)^n e^{-rL} \qquad n = 0, 1, 2, \cdots$$

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

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

$$= \sum_{n=0}^{\infty} p(n) \underbrace{\int_{-\infty}^{\infty} \delta(y-n)\, dy}_{1} = \sum_{n=0}^{\infty} p(n)$$

$$= e^{-rL} \sum_{n=0}^{\infty} \frac{1}{n!}(rL)^n = e^{-rL}e^{rL}$$

$$= 1$$

Mean

$$<n> = \int_{-\infty}^{\infty} yp(y)\, dy = \sum_{n=0}^{\infty} p(n) \underbrace{\int_{-\infty}^{\infty} y\delta(y-n)\, dy}_{n}$$

$$= \sum_{n=0}^{\infty} np(n) = e^{-rL} \sum_{n=0}^{\infty} \frac{n}{n!}(rL)^n$$

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

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

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.

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

Alternatively

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

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:

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

· ·

MIT OpenCourseWare
http://ocw.mit.edu

8.044 Statistical Physics I
Spring 2013

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.