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.

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

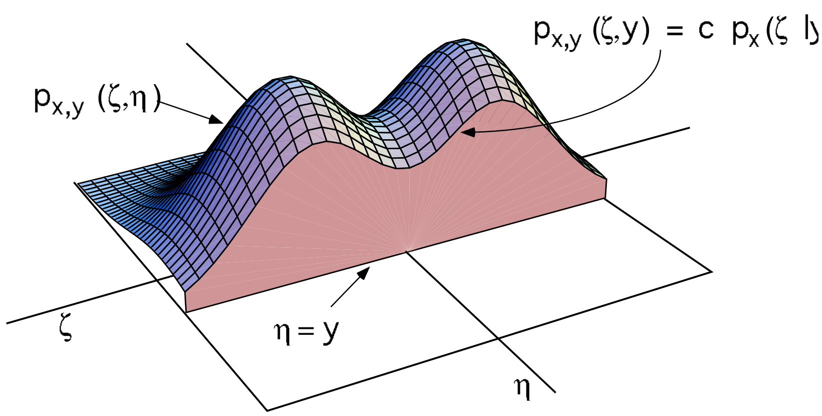
$$p_y(\eta) = \int_{-\infty}^{\infty} p_{x,y}(\zeta,\eta)d\zeta$$

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 $\zeta - \eta$ 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.

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

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

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

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

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

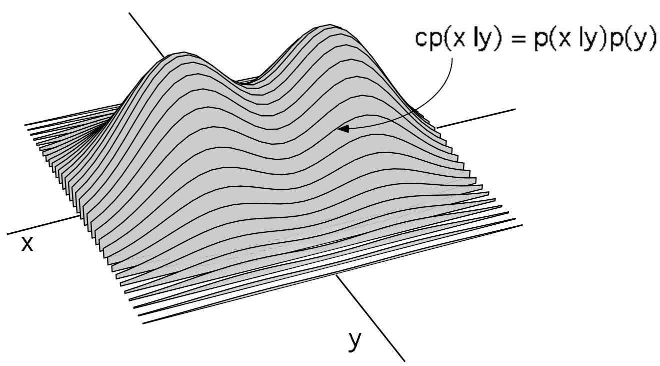
The result can be viewed in two ways. It can be interpreted as a way of finding the conditional density from the joint density (and the density of the conditioning event which can be recovered from the joint density):

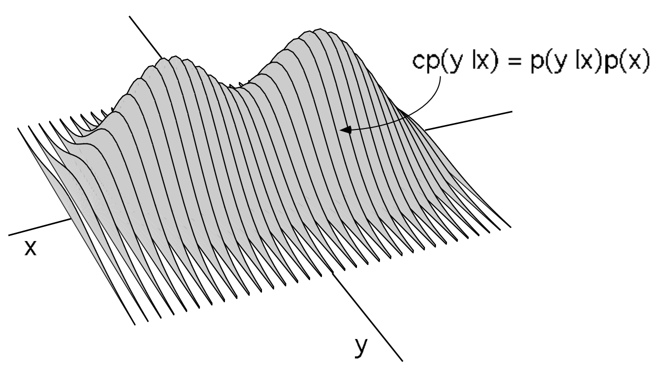
$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

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:

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

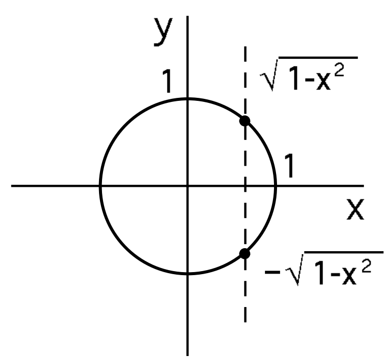
Physically, two variables are S.I. if knowledge of one gives no additional information about the other beyond that contained in its own unconditioned probability density:

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

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.

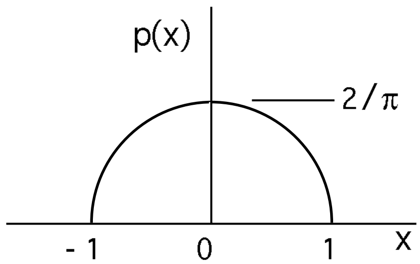
$$\begin{aligned} p(x,y) &= 1/\pi \quad x^2+y^2 \leq 1 \\ &= \;\; 0 \quad\;\; x^2+y^2 > 1 \end{aligned}$$



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

Solution

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

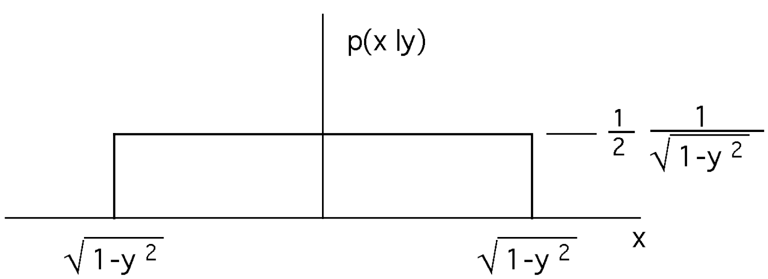


By symmetry, the functional form for $p(y)$ will be identical.

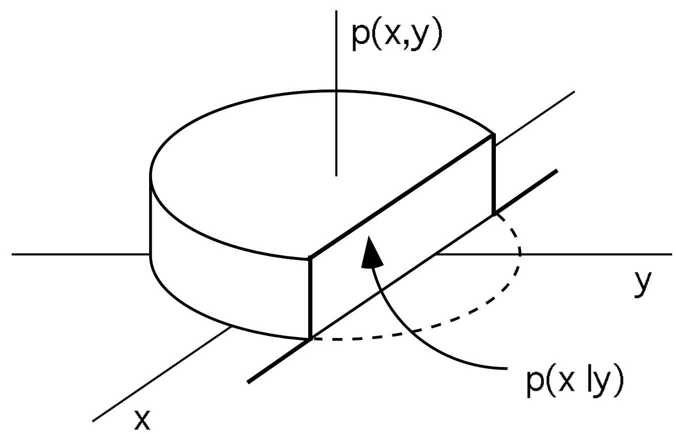
$$\begin{aligned} p(y) &= \frac{2}{\pi}\sqrt{1-y^2} & \qquad |y| \le 1 \\ &= 0 & \qquad |y| > 1 \end{aligned}$$

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.

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



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).

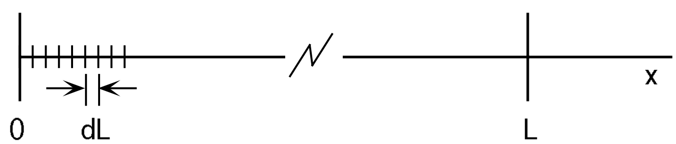
c) Show that the Poisson density

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

is a solution of the equation. Is it unique?

<u>Solution</u>

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

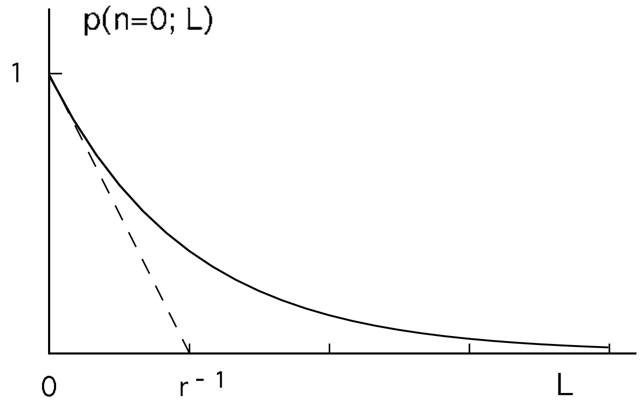


Consider dL so short that $p(0) >> p(1) >> p(n>1)$ in dL. But the probabilities must sum to unity, $p(0)+p(1)+p(2)+\cdots = 1$, so one can find an approximation to $p(0)$ which will be valid in the limit of small dL.

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

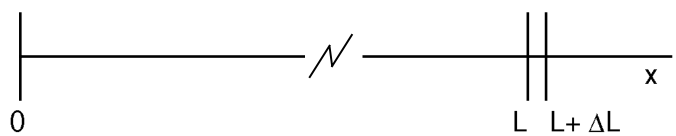
The probability of an event in any sub-interval is S.I. of the events in every other sub-interval, so

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



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 $\Delta L \rightarrow 0$ 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

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

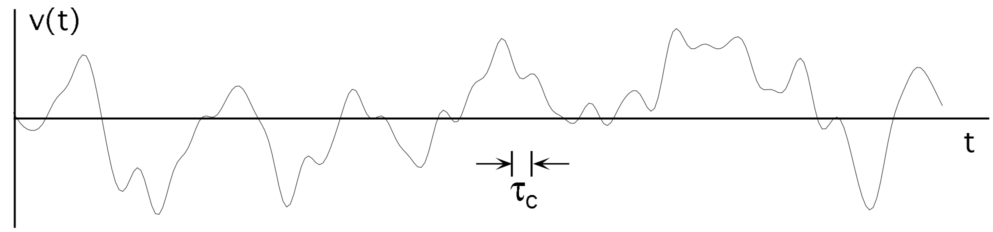
$$p(n;0) = 0 \qquad n \neq 0.$$

. . . .

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.

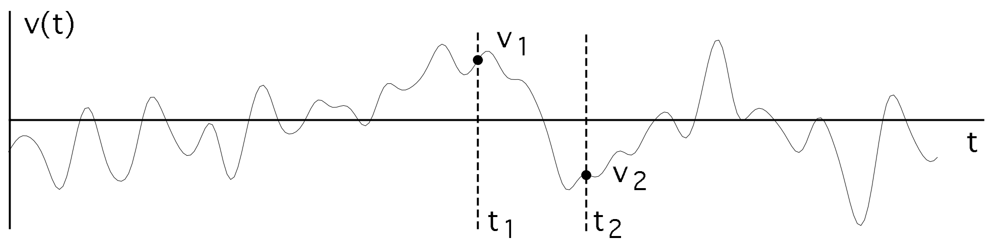
3) The "statistics" do not seem to change with time.

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.

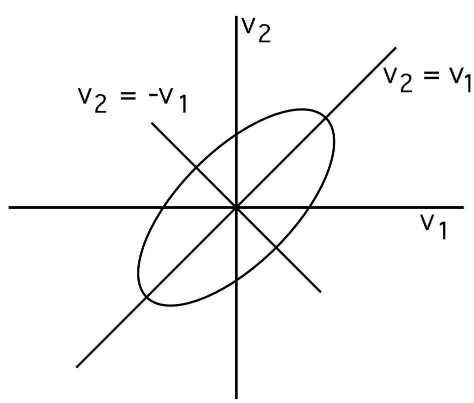
The Joint Probability Density

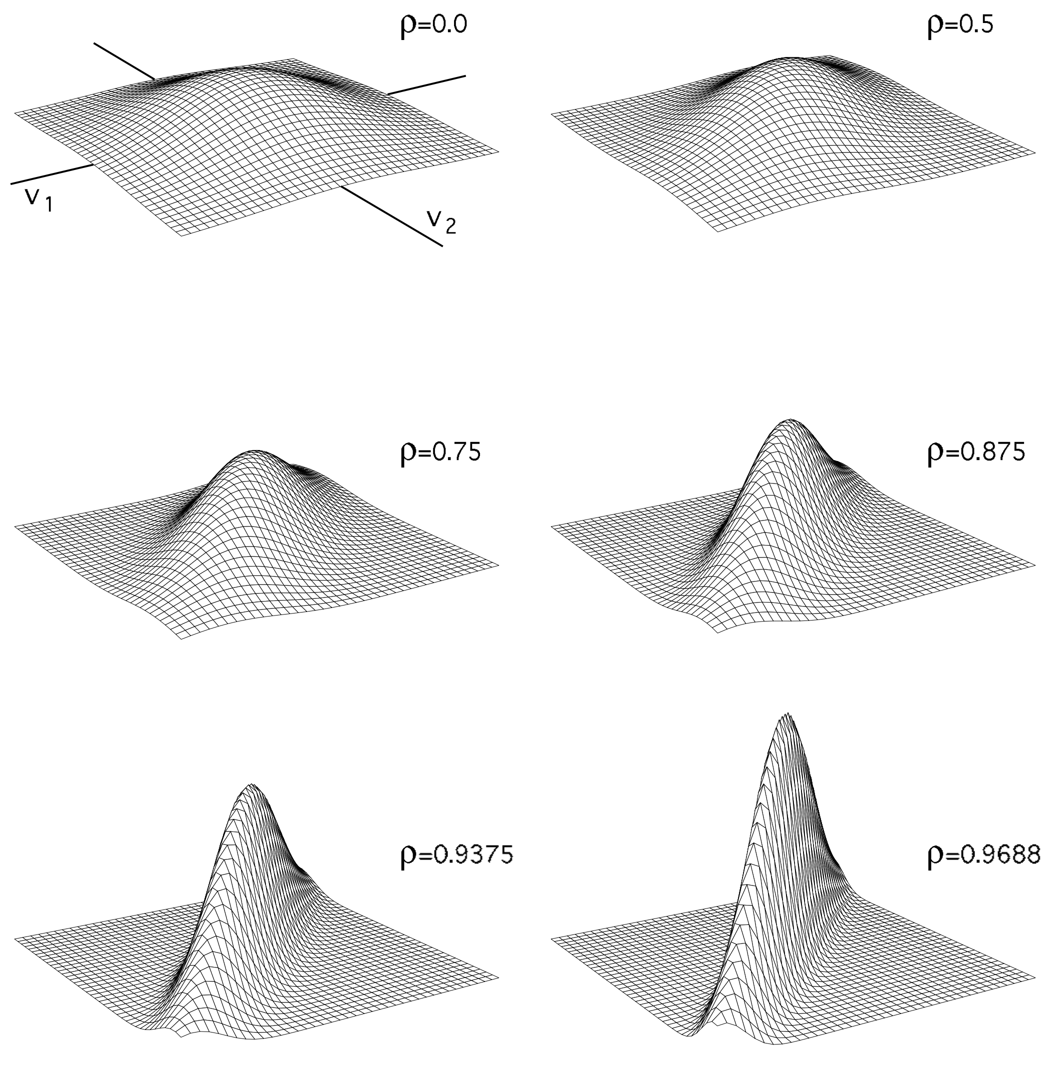


$$p(v_1, v_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp[-\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2\sigma^2(1-\rho^2)}]$$

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: $|\rho| \le 1$. 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

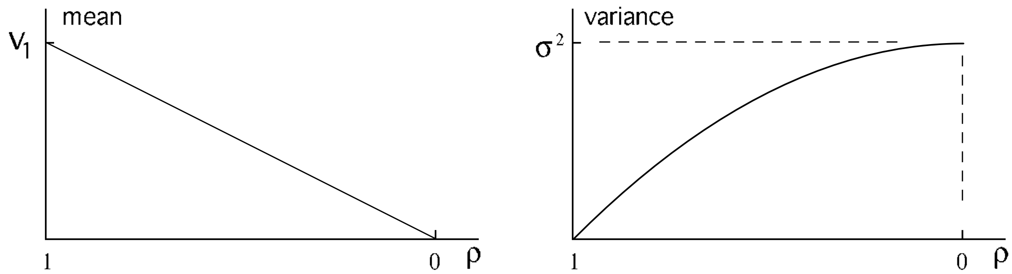
$$
p(v_1, v_2) = p(v_1)p(v_2) \qquad \text{only when } \rho = 0
$$

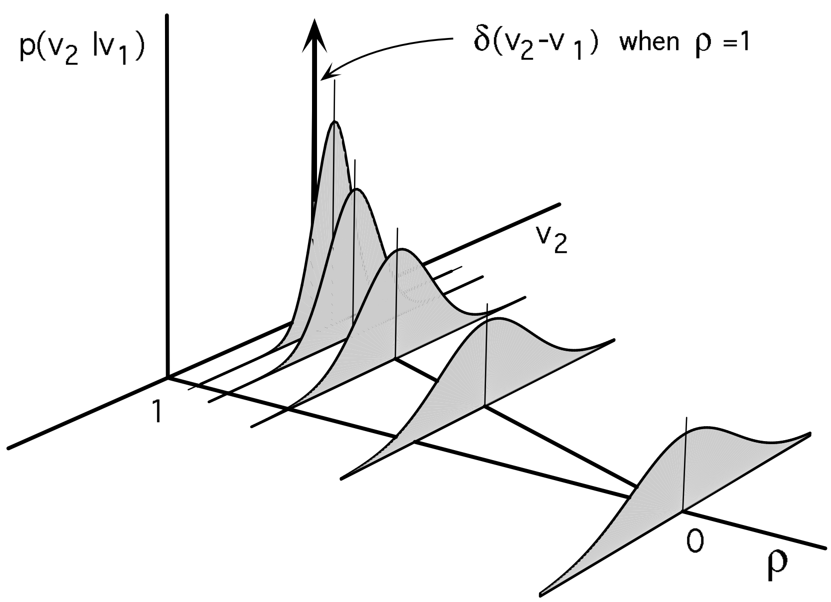
This implies that v_1 and v_2 are not statistically independent unless $\rho = 0$, that is, at large time separations between t_1 and t_2.

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.

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

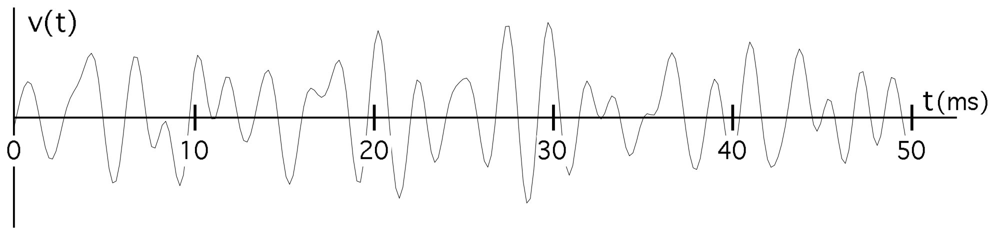
$$R(\tau) \equiv < v(t)v(t+\tau) > .$$

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.

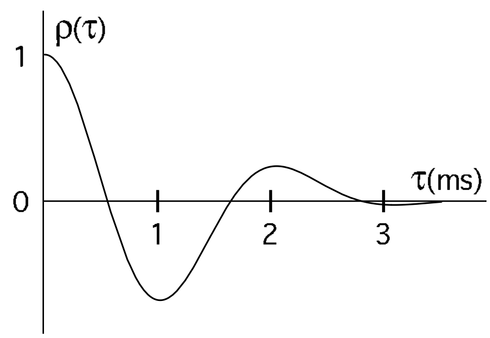
$$\begin{aligned} < v_1 v_2 > &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1 v_2 p(v_1, v_2)\, dv_1\, dv_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v_1 v_2 p(v_2|v_1) p(v_1)\, dv_1\, dv_2 \\ &= \int_{-\infty}^{\infty} v_1 p(v_1) \underbrace{\int_{-\infty}^{\infty} v_2 p(v_2|v_1)\,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

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.

. . . .