ECE594I Notes set 4: More Math: Expectations of 1-2 R.V.'s

Mark Rodwell

University of California, Santa Barbara

rodwell@ece.ucsb.edu 805-893-3244, 805-893-3262 fax

References and Citations:

Sources / Citations :

Kittel and Kroemer : Thermal Physics

Van der Ziel : Noise in Solid - State Devices

Papoulis : Probability and Random Variables (hard, comprehensive)

Peyton Z. Peebles : Probability, Random Variables, Random Signal Principles (introductory)

Wozencraft & Jacobs : Principles of Communications Engineering.

Motchenbaker : Low Noise Electronic Design

Information theory lecture notes : Thomas Cover, Stanford, circa 1982

Probability lecture notes : Martin Hellman, Stanford, circa 1982

National Semiconductor Linear Applications Notes : Noise in circuits.

Suggested references for study.

Van der Ziel, Wozencraft & Jacobs, Peebles, Kittel and Kroemer

Papers by Fukui (device noise), Smith & Personik (optical receiver design)

National Semi. App. Notes (!)

Cover and Williams : Elements of Information Theory

random variables
and
Expectations

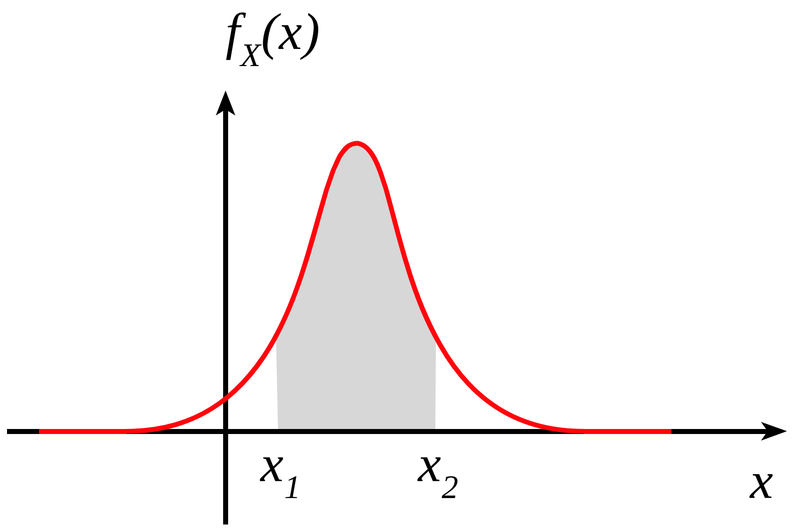
Recall: Distribution Function of Random Variable

During an experiment, a random variable X takes on a particular value x.

The probability that x lies between x_1 and x_2 is

$$P\{x_1 < x < x_2\} = \int_{x_1}^{x_2} f_X(x)dx$$

$f_X(x)$ is the probability distribution function.



Mean values and expectations

Expectation of a function g(X) of the random variable X

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f_X(x)dx$$

Mean Value of X

$$\langle X \rangle = \overline{X} = E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx$$

Expected value of X^2

$$\langle X^2 \rangle = E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx$$

Variance

The variance σ_x^2 of X is its root - mean - square deviation from its average value

$$\sigma_X^2 = \left\langle \left(X - \overline{x}\right)^2 \right\rangle = E\left[\left(X - \overline{x}\right)^2\right] = \int_{-\infty}^{+\infty} \left(x - \overline{x}\right)^2 f_X(x)\,dx$$

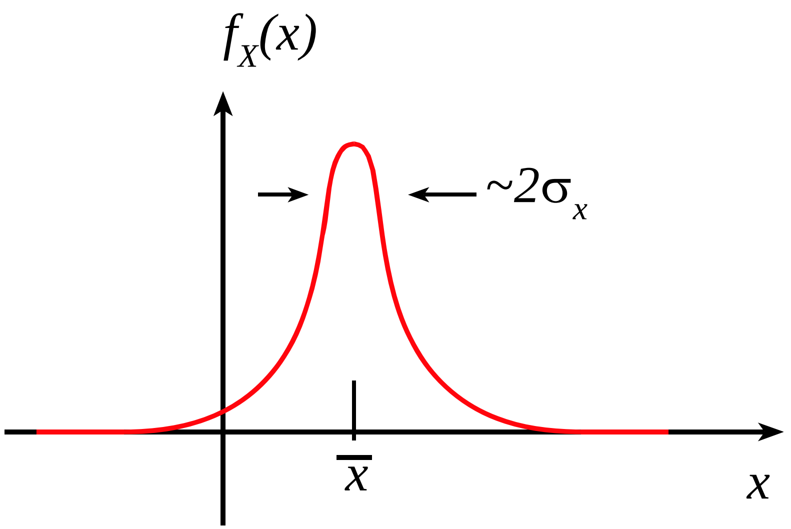
The standard deviation σ_x of X is simply the square root of the variance

Returning to the Gaussian Distribution

The notation describing the Gaussian distribution :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(\frac{-(x-\bar{x})^2}{2\sigma_x^2} \right)$$

should now be clear.



Variance vs Expectation of the Square

$$\sigma_X^2 = \left\langle \left(X - \bar{x}\right)^2 \right\rangle = \left\langle \left(X - \bar{x}\right)\left(X - \bar{x}\right) \right\rangle$$

$$= \left\langle X^2 - 2X \cdot \bar{x} + \left(\bar{x}\right)^2 \right\rangle$$

$$= \left\langle X^2 \right\rangle - 2 \cdot \bar{x} \left\langle X \right\rangle + \left\langle \left(\bar{x}\right)^2 \right\rangle$$

$$= \left\langle X^2 \right\rangle - 2 \cdot \bar{x} \cdot \bar{x} + \left(\bar{x}\right)^2$$

$$\sigma_X^2 = \left\langle X^2 \right\rangle - \left(\bar{x}\right)^2$$

The variance is the expectation of the square minus the square of the expectation.

Example of Expectation: Mean Kinetic Energy

Our particle with a thermal velocity distribution :

$$f_{V_x}(v_x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_v} \exp\left\{ -\frac{v_x^2}{2\sigma_v^2} \right\} \text{ where } \sigma_v = \sqrt{kT/m}$$

$$E[v_x] = \int_{-\infty}^{+\infty} v_x f_{V_x}(v_x) dv_x = 0$$

$$E[v_x^2] = \int_{-\infty}^{+\infty} v_x^2 f_{V_x}(v_x) dv_x$$

(this computes the variance of a Gaussian --- skip proof)

$$= \sigma_v^2 = kT/m$$

So, $E[\text{Kinetic energy}] = E[mv_x^2/2] = kT/2$

$$E[E] = kT/2$$

Example of Expectation: Shot Noise (Bernoulli Trial)

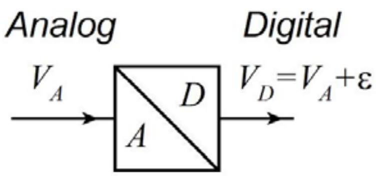
$$X = \begin{cases} 1 & \text{probability } p \\ 0 & \text{probability } q = (1-p) \end{cases}$$

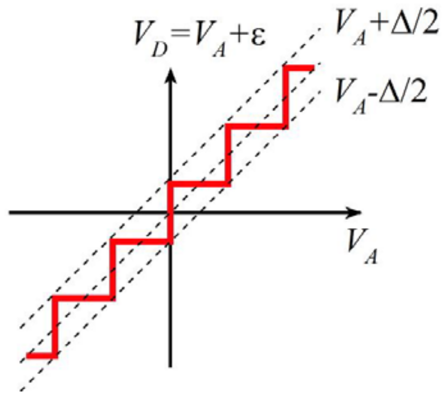
$$E[X] = \sum x p_X(x) = 0 \cdot q + 1 \cdot p = p$$

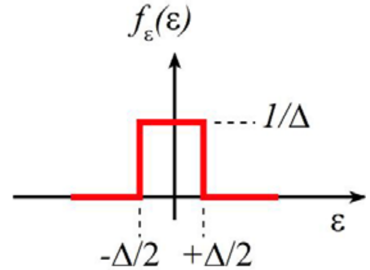
$$E[X^2] = \sum x^2 p_X(x) = 0^2 \cdot q + 1^2 \cdot p = p$$

$$\sigma_X^2 = E[X^2] - \left(E[X]\right)^2 = p^2 - p = pq$$

Example of Expectation: Quantization "Noise"







If we stipulate that V_A is itself an R.V. distributed uniformly over some large range, say $\pm 2^N \cdot \Delta$, then the quantization error ε is uniformly distributed over $[-\Delta/2, +\Delta/2]$

The quantization error is then also an R.V., with

$$E[\varepsilon] = 0, \;\; E[\varepsilon^2] = \sigma_\varepsilon^2 = \int_{-\Delta/2}^{+\Delta/2} \frac{d\varepsilon}{\Delta} = \frac{\Delta^2}{12}$$

Note that ε is an R.V. only if V_A is also an R.V., i.e. we must be cautious in treating quantization error as noise.

Pairs of Random Variables

To understand random processes,
we must first understand pairs of random variables.

In an experiment, a pair of random variables X and Y
takes on specific particular values x and y.

Their joint behavior is described by the joint distribution $f_{XY}(x,y)$

$$P\{A < x < B \text{ and } C < y < D\} = \int_C^D \int_A^B f_{XY}(x,y)dxdy$$

Pairs of Random Variables

Marginal distributions must also be defined

$$P\{A < x < B\} = \int_{-\infty}^{+\infty}\int_{A}^{B} f_{XY}(x,y)dxdy$$

$$= \int_{A}^{B} f_X(x)dx$$

and similarly for Y :

$$P\{C < y < D\} = \int_{C}^{D}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy$$

$$= \int_{C}^{D} f_Y(y)dy$$

Statistical Independence

In the case where

$$f_{XY}(x,y) = f_X(x) f_Y(y),$$

the variables are said to be statistically independent.

This is not generally expected.

Conditional Densities

Arises in revising the distribution of an R.V. after making some observation. Consider the conidtional probability that the R.V. X is less than a specific value x given the occurence of event B.

$$P[(X \leq x) \mid B] \equiv \frac{P[(X \leq x) \cap B]}{P[B]} = F_{X|B}(x \mid B)$$

This is the cumulative distribution of function of X given B.

Distribution function : $f_{X/B}(x \mid B) \equiv \dfrac{d}{dx} F_{X|B}(x \mid B)$

Conditional Densities (II)

Given a pair of random variables (X,Y), what is the distibution of X given that Y has some particular value y?

$$f_{X/Y}(x \mid Y=y) = f_{X/Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

If $X \perp\!\!\!\perp Y$, then $f_{X/Y}(x \mid y) = f_X(x)$

Expectations of a pair of random variables

The expectation of a function g(X, Y) of the random variables Y and Y is

$$E[g(x,y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f_{XY}(x,y)\,dxdy$$

Expectation of X :

$$E[X] = \overline{x} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{XY}(x,y)\,dxdy = \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

Expectation of X^2

$$E[X^2] = \langle X^2 \rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f_{XY}(x,y)\,dxdy = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx$$

...and similarly for Y and Y^2.

Correlation between random variables

The correlation of X and Y is

$$R_{XY} = E[XY] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy \bullet f_{XY}(x,y)\,dx\,dy$$

The covariance of X and Y is

$$C_{XY} = E[(X-\overline{x})(Y-\overline{y})] = E[XY - \overline{x}Y - X\overline{y} + \overline{xy}]$$
$$= R_{XY} - \overline{x}\cdot\overline{y}$$

Note that correlation and covariance are the same if either X or Y have zero mean values.

Correlation versus Covariance

When we are working with voltages and currents, we usually separate the mean value (DC bias) from the time-varying component.

The random variables then have zero mean.

Correlation is then equal to covariance.

It is therefore common in circuit noise analysis to use the two terms interchangeably.

But, nonzero mean values can return when we e.g. calculate conditional distributions.

Be careful.

Correlation Coefficient

The correlation coefficient of X and Y is

$$\rho_{XY} = C_{XY} / \sigma_X \sigma_Y$$

Note the (standard) confusion in terminology between correlation and covariance.

Sum of TWO Random Variables

Sum of two random variables : $Z = X + Y$

$$E\left[Z^2\right] = E\left[(X+Y)^2\right] = E\left[X^2 + 2XY + Y^2\right]$$
$$= E\left[X^2\right] + E\left[Y^2\right] + 2R_{XY}$$

If X and Y both have zero means

$$E\left[Z^2\right] = E\left[X^2\right] + E\left[Y^2\right] + 2C_{XY}$$

This emphasizes the role of correlation.

Uncorrelated Variables.

Uncorrelated :

$$C_{XY} = 0$$

Statistically independent :

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Independence implies zero correlation.

Zero correlation does not imply independence.

For JGRV's, uncorrelated does imply independence

Summing of Noise (Random) Voltages

Two voltages are applied to the resistor R

The power dissipated in the resistor is a random variable P

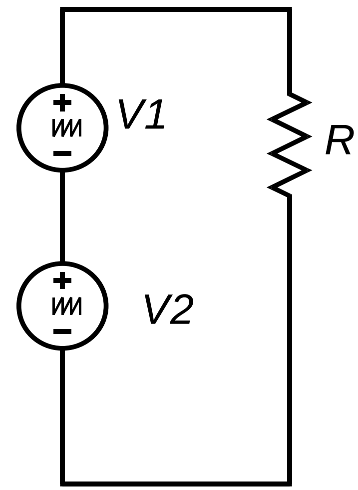
$$E[P] = \langle P \rangle = \frac{1}{R}\left\langle (V_1 + V_2)^2 \right\rangle = \frac{1}{R}\left\langle V_1^{\ 2} + 2V_1V_2 + V_2^{\ 2} \right\rangle$$

$$= \frac{1}{R}\left\langle V_1^{\ 2} \right\rangle + \frac{1}{R} 2C_{V_1V_2} + \frac{1}{R}\left\langle V_2^{\ 2} \right\rangle$$

$$= \frac{1}{R}\left\langle V_1^{\ 2} \right\rangle + \frac{1}{R} 2\sigma_{V_1V_2} + \frac{1}{R}\left\langle V_2^{\ 2} \right\rangle$$

$$= \frac{1}{R}\left\langle V_1^{\ 2} \right\rangle + \frac{1}{R}\left\langle V_2^{\ 2} \right\rangle + \frac{1}{R} 2\rho_{V_1V_2}\sigma_{V_1}\sigma_{V_2}$$

$$= \frac{1}{R}\left\langle V_1^{\ 2} \right\rangle + \frac{1}{R} 2\langle V_1V_2 \rangle + \frac{1}{R}\left\langle V_2^{\ 2} \right\rangle$$



The noise powers of the two random generators do not add -

- a correllati on term must be included.

The instantane ous time values of the random noise voltages do add.

Shot Noise as a Random Variable

The fiber has transmission probability p.
Send one photon, and call the # of received photons N_1.

$$E[N_1] = \overline{N}_1 = p \quad \text{and} \quad E[N_1^2] = p \quad \text{so} \quad \sigma_{N_1}^2 = E[N_1^2] - \overline{N}_1^2 = p - p^2$$

If we now send many photons (M of them), transmission of each is statistically independent, so --- calling the # of received photons N,

$$E[N] = M \cdot E[N_1] = Mp \quad \text{and} \quad \sigma_N^2 = M \cdot \sigma_{N_1}^2 = M(p - p^2)$$

Now suppose $M >> 1$, $p << 1$, and $Mp >> 1$,

$$\rightarrow \sigma_N^2 = \overline{N}$$

The variance of the count approaches the mean value of the count.

Thermal Noise as a Random Variable

A capacitor C is connected to a resistor R .
The resistor is in equilibrium with a "reservoir" (a warm room) at temperature T

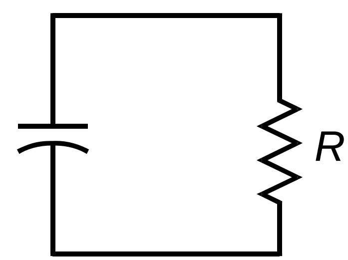
R can exchange energy with the room in the form of heat.
C can dissipate no power : it establishes thermal equilibrium with the room via the resistor.

From thermodynamics, any independent degree of freedom of a system at temperature T has mean energy kT/2, hence

$$\langle E \rangle = kT/2$$

$$\langle CV^2/2 \rangle = kT/2$$

$$\langle V^2 \rangle = kT/C$$



The noise voltage has variance kT/C.

Distribution of Sums and Jointly Gaussian RV's

Distribution of a Sum of 2 Independent Random Variables

Sum of two *indendent* random variables : $Z = X + Y$

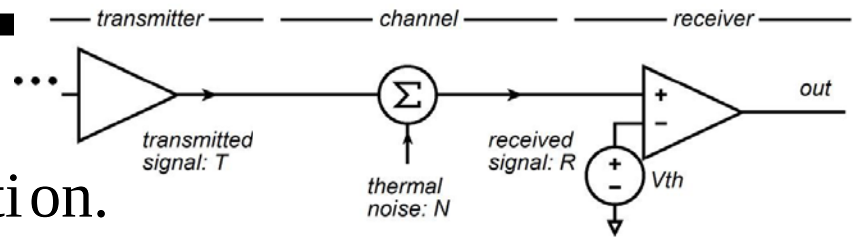
$$F_Z(z) = P[X + Y \le z] = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-Y} f_{XY}(x, y)dx \right) \cdot dy$$

$$= \int_{-\infty}^{+\infty} f_Y(y) \int_{-\infty}^{z-Y} f_X(x)dx \cdot dy$$

But $f_Z(z) = \dfrac{d}{dz} F_Z(z)$, so

$f_Z(z) = \int_{-\infty}^{+\infty} f_Y(y) f_X(z - y)dy$, the convolution of $f_Y(y)$ & $f_X(z)$.

Example: Digital Transmission

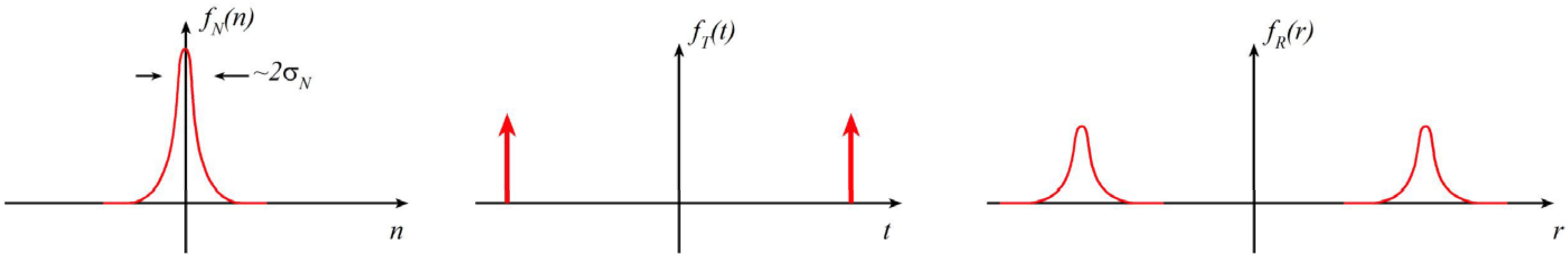


N = thermal noise, Gaussian distribution.

R = received signal = $T + N$

$f_T(t) = (1/2)\delta(t+1) + (1/2)\delta(t-1)$ (t is transmitted signal, not time)

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-n^2}{2\sigma_n^2}\right)$$



Example of Convolution of 2 Distributions: Communication

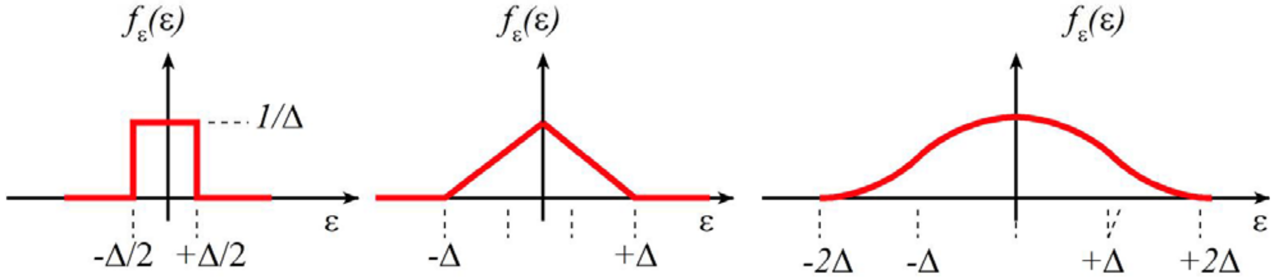
Sum of two *indendent * random variables : $Z = X + Y$

$$F_Z(z) = P[X + Y \le z] = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-Y} f_{XY}(x, y) dx \right) \cdot dy$$

$$= \int_{-\infty}^{+\infty} f_Y(y) \int_{-\infty}^{z-Y} f_X(x) dx \cdot dy$$

But $f_Z(z) = \dfrac{d}{dz} F_Z(z)$, so

Distribution of a Sum of Many Independent RV's



Given $f_Z(z) = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y)\,dy$, the convolution of $f_Y(y)$ & $f_X(z)$,

we can see that convolving 1,2,4, ... identical uniform distributions will slowly lead to a form similar to a Gaussian.

This gives some crude sense of the central limit theorem.

Distribution of a Sum of a Few Random Variables

Sum of two * perhaps not indendent * random variables : $Z = X + Y$

$$F_Z(z) = P[X + Y \le z] = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-Y} f_{XY}(x,y)dx \right) \cdot dy \quad \text{and} \quad f_Z(z) = \frac{d}{dz} F_Z(z)$$

This is a major difficulty for circuit & system design. Given a random process (random function of time) $V_{in}(t)$, and a linear filter

$$V_{out}(t) = a_0 V_{in}(t - 0 \cdot \tau) + a_1 V_{in}(t - 1 \cdot \tau) + a_2 V_{in}(t - 2 \cdot \tau) + \ldots$$

the output V_{out} is a sum of random variables, and to find its distribution function requires computing convolution integrals.

Jointly Gaussian random variables (next) avoid this difficulty.

Pairs of Jointly Gaussian Random Variables

If X and Y are Jointly Gaussian :

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

$$\times\exp\left[-\frac{1}{2(1-\rho_{XY}^2)}\cdot\left(\frac{(x-\bar{x})^2}{\sigma_X^2}+\frac{(x-\bar{x})(y-\bar{y})}{\sigma_X\sigma_Y}+\frac{(y-\bar{y})^2}{\sigma_Y^2}\right)\right]$$

This definition can be extended to a larger # of variables.

In general, we can have a Jointly Gaussian random vector

$(X_1, X_2, \cdots, X_n)$

which is specified by a set of

means $\bar{x}_i$, variances $E[x_i x_i]$, and covariances $E[x_i x_j]$

Linear Operations on JGRV's

If X and Y are Jointly Gaussian, and if we define

$V = aX + bY$ and $W = cX + dY$

Then V and W are also Jointly Gaussian.

This is stated without proof; the result arises because convolution of 2 Gaussian functions produces a Gaussian function.

The result holds for JGRVs of any number.

Probability distribution after a Linear Operation on JGRV's

$$\bar{V} = E[V] = E[aX + bY] = a\bar{X} + b\bar{Y} \qquad \text{and} \quad \bar{W} = c\bar{X} + d\bar{Y}$$

$$\sigma_V^2 = E[V^2] - \bar{V}^2 = a^2 E[X^2] + b^2 E[Y^2] + 2ab \cdot E[XY] - (a\bar{X} + b\bar{Y})^2$$

$$\sigma_W^2 = E[W^2] - \bar{W}^2 = c^2 E[X^2] + d^2 E[Y^2] + 2cd \cdot E[XY] - (c\bar{X} + d\bar{Y})^2$$

$$\begin{aligned} C_{VW} &= E[VW] - \bar{V}\bar{W} = E[(aX + bY)(cX + dY)] - \bar{V}\bar{W} \\ &= E[acX^2 + (ad + bc)XY + bdY^2] - \bar{V}\bar{W} \\ &= acE[X^2] + (ad + bc)E[XY] + bd \cdot E[Y^2] - (a\bar{X} + b\bar{Y})(c\bar{X} + d\bar{Y}) \end{aligned}$$

tedious details

We can now calculate the joint distribution of V and W.

$$f_{VW}(v,w) = \frac{1}{2\pi\sigma_V\sigma_W\sqrt{1-\rho_{VW}^2}} \times \exp\left[-\frac{1}{2(1-\rho_{VW}^2)} \cdot \left(\frac{(v-\bar{v})^2}{\sigma_V^2} + \frac{(v-\bar{v})(w-\bar{w})}{\sigma_V\sigma_W} + \frac{(w-\bar{w})^2}{\sigma_W^2} \right) \right]$$

Jointly Gaussian Random Variables in N Dimensions

Define a set of random variables (a random vector) $\mathbf{X}$ and a covariance matrix $\mathbf{C_X}$:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \qquad \mathbf{C_X} = E\left[\left(\mathbf{X} - \overline{\mathbf{X}}\right)\left(\mathbf{X} - \overline{\mathbf{X}}\right)^T\right] = \begin{bmatrix} C_{X_1X_1} & C_{X_1X_2} & \cdots & C_{X_1X_N} \\ C_{X_2X_1} & C_{X_2X_2} & & \\ \vdots & & \ddots & \\ C_{X_NX_1} & & & C_{X_NX_N} \end{bmatrix}$$

$\mathbf{C_X}$ is a matrix of correlation coefficients.

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\left(\det[\mathbf{C_X}]\right)^{1/2}} \exp\left(-\frac{1}{2}\cdot\left(\mathbf{X} - \overline{\mathbf{X}}\right)^T \mathbf{C_X^{-1}}\left(\mathbf{X} - \overline{\mathbf{X}}\right)\right)$$

Key point : linear operations on jointly Gaussian RV's, such as $\mathbf{Y} = \mathbf{LX}$, lead to jointly Gaussian R.V.'s, with $\mathbf{C_Y} = \mathbf{LC_XL^T}$.

Why are JGRV's Important ?

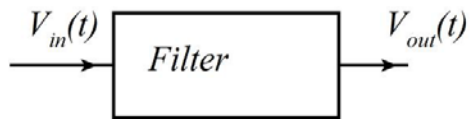
The math on the last slide was tedious but there is a clear conclusion :

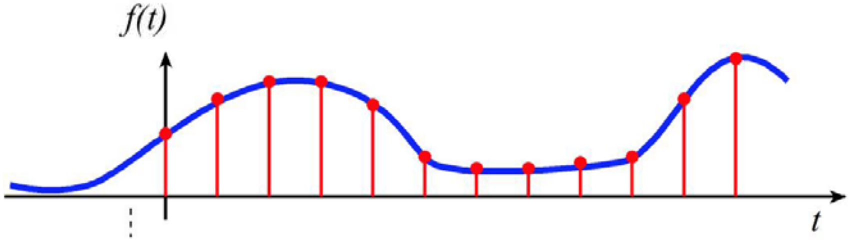
With JGRV's subjected to linear operations, it is sufficient to keep track of means, correlations, and variances.

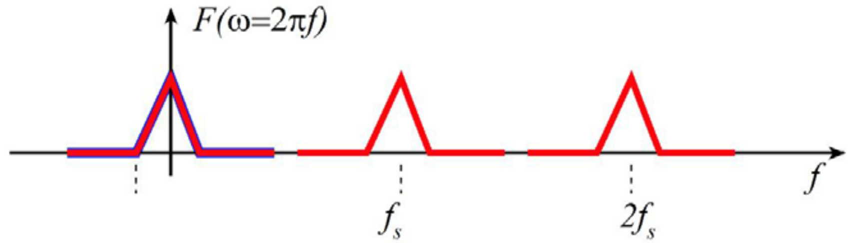
With this information, distribution functions can always be simply found.

This vastly simplifies calculations of noise propagation in linear systems (linear circuits).

Linear Filtering Operations







From Nyquist's sampling theorem, a bandlimited waveform is uniquely determined by its samples at time points $(n \cdot \tau)$

We can therefore analyze a linear filter in discrete time

$$V_{out}(t) = a_0 V_{in}(t - 0 \cdot \tau) + a_1 V_{in}(t - 1 \cdot \tau) + a_2 V_{in}(t - 2 \cdot \tau) + ...$$

We write the signals as vectors :

$$\mathbf{V_{out}} = [V_{out}(0 \cdot \tau), V_{out}(1 \cdot \tau), ..., V_{outN}(n \cdot \tau)]^T = [V_{out1}, V_{out2}, ..., V_{outN}]^T$$

$$\mathbf{V_{in}} = ... = [V_{in1}, V_{in2}, ..., V_{in,N}]^T$$

Hence in some finite time window, $\mathbf{V_{out}} = \mathbf{M V_{in}}$

Time waveforms are vectors. Filters are linear transformations.

Linear Filtering Operations

Noise propagating through linear circuits & systems undergoes linear transformations. Sums of R.V.'s are formed.

If the R.V.'s are Jointly Gaussian, then the filtered R.V.'s are also jointly Gaussian. Hence, we must only calculate means & variances and covariances to determine probability distributions.

If the R.V's are not jointly Gaussian, probability distributions of the filtered R.V.'s must be determined by convolution integrals.

It is fortunate that the central limit theorem causes many random processes to be jointly Gaussian.

Estimation

Conditional Densities Again

$$f_{X/Y}(x \mid Y = y) = f_{X/Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Conditional Expectation

$$E[g(X) \mid y] \equiv \int_{-\infty}^{+\infty} g(x) \cdot f_{X/Y}(x \mid y) dx$$

Expectation of X given a particular value of Y :

$$E[X \mid y] \equiv \int_{-\infty}^{+\infty} x \cdot f_{X/Y}(x \mid y) dx$$

Expected value of the RV X, given that we have observed that the RV Y has taken on value y.

...important for estimation.

Estimation

Given an RV Y, we wish to estimate its value $\rightarrow \hat{Y}$.

Possible measure of quality of this estimate : mean - square error

$$\text{M.S.E.} = E[(Y - \hat{Y})^2] = E[Y^2] + \hat{Y}^2 - 2E[Y] \cdot \hat{Y}$$

Minimum mean squared error estimate MMSEE :

$\hat{Y} = E[Y]$ picked to minimize $E[(Y - \hat{Y})^2]$

Minimum mean squared error MMSE

$$\text{MMSE} = E[Y^2] + (E[Y])^2 - 2E[Y] \cdot E[Y] = \sigma_Y^2$$

Estimation

If we observe the R.V. X, and use this observation to estimate the value of the (presumably correlated) RV Y, then

$\hat{Y}(x) = \text{MMSE estimate of } Y \text{ given that we observe } X = x.$

$$\hat{Y}(x) = E[Y \mid x]$$

$$\text{MMSE} = \sigma^2_{Y|x} = E[(\hat{Y}(x) - Y)^2 \mid x]$$

Estimation with JGRV's

For a pair of JGRV's X and Y the analysis simplifies :

$$E[X \mid y] = \overline{X} + \frac{C_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right)$$