Quick Review of Probability Theory and Stochastic Processes for RF and THz Sensors

• Arguably one of the most useful branches of Applied Mathematics because of its universal application in Engineering, Physical Science, and Life Science

• Applies when an otherwise deterministic measurement or operation is made uncertain because of random events in time or space associated with some physical quantity involved in the measurement or operation.

• Classic example in Electrical Engineering is the current through a resistor at temperature T. From Solid State Physics, we know resistance always involves the collisions (classical viewpoint) or scattering (quantum mechanical viewpoint) between charge carriers and imperfections in the solid, such as impurities or defects. Since these collisions or scattering events are inherently random in time and/or space, the associated electric current varies randomly too. Since there are generally so many charge carriers in any real resistor, it is impossible in practice (although not in principle) to solve for the fluctuating current in the time domain. However, when the number of charge carriers gets so large that nearly every possible type of collision or scattering event is occurring over the sampling time of the experiment (more on this later), then the problem gets relative simple if we know the *statistics.*

• "Knowing the statistics" means that we know the probability that in any given measurement the value of the random variable (e.g., current in the resistor) will have a value in a certain range (e.g., i to i + di for the instantaneous current), or a specific value x_n for a discrete random variable (as occurs often in quantum systems).

• A good way to think about the "statistics" is that you have a "ensemble" of identical copies of the system under question (e.g., the resistor) but in all possible different instantaneous conditions with respect to the random variable of interest. In other words the "ensemble" of systems represents every possible value of current that can possibly occur according to the laws of physics and boundary conditions applicable. Then, the probability that the current has value P(i, i+di) is just the number of systems in the ensemble having current in this range, and P takes on a grand purpose called the probability density function (pdf).

• In THz and most RF sensor and communications system engineering, we assume to know the "statistics" of all random processes in advance. However, in real-world system, we often run into situations where the statistics is unknown and must be measured (e.g., multipath in urban communications system, clutter in radar systems, speckle and coherent scattering in THz imaging systems, etc)

Quick Review of Probability Theory (cont)

- Knowing the statistics and associated pdf, we can perform many useful mathematical operations of great interest to system performance:

(1) Normalization: $\int_{x_{min}}^{x_{max}} P(x)dx = 1$ where x_{min} and x_{max} are the minimum and maximum allowed values for random variable x

(2) Ensemble Averaging: Mean value $< x >= \int_{x_{min}}^{x_{max}} xP(x)dx$

- Mean deviation from mean value $< x - < x >>= \int_{x_{min}}^{x_{max}} (x - < x >)P(x)dx \equiv \Delta x$

(very useful for random processes where <x> = 0)

- Mean square deviation (aka "variance") $< (x - < x >)^2 >= \int_{x_{min}}^{x_{max}} (x - < x >)^2 P(x)dx$

(very useful for random processes where <x> does not = 0)

- Root-mean-square deviation (aka "standard deviation")

$$< (x - < x >)^2 >^{1/2} = < (\Delta x)^2 >^{1/2} = \left(\int_{x_{min}}^{x_{max}} (x - < x >)^2 P(x)dx \right)^{1/2} \equiv \sigma$$

Great example: current fluctuations through a resistor. As discovered by Johnson and explained by Nyquist, and in the absence of saturation, limiting, or other nonlinear effects, the current fluctuations through a resistor at temperature T and under bias voltage V_B are described by the classic Gaussian, or "normal" density function

$$P(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp[-(x - < x >)^2 / 2\sigma^2]$$

To use on a resistor need only replace x by i and <x> by $i_0 = V_B/R$ according to Ohm's law
Then Johnson and Nyquist proved that

$$\sigma^2 = <(\Delta i)^2> = 4k_B T\Delta f/R$$

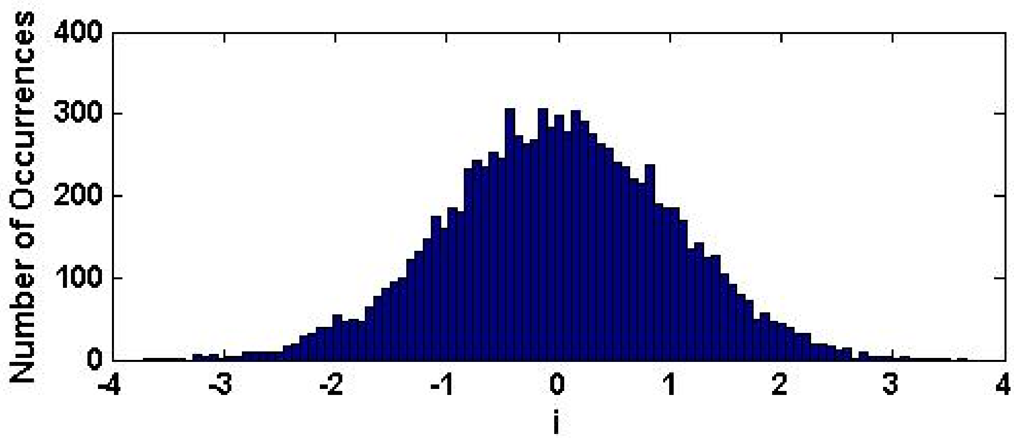
where Δf is the measurement bandwidth. Thus, we can write

$$P(i) = \sqrt{\frac{R}{8\pi k_B T\Delta f}} \cdot \exp[-(i - i_0)^2 \cdot R/(8k_B T)]$$

Quick Review of Probability Theory (cont)

- A great way to visualize this and all pdfs is by the *histogram*. A *histogram is just* a way of plotting the results of a multitude of measurements where the results of each measurement are governed by the pdf of interest. The outcome values of each measurements are plotted along the x axis, and the number of occurrences of that value are plotted in "bins" along the y axis. In this way, the histogram is just a Simpson's rule approximation of the true pdf. As such, the sum over all bins must equal the total number of measurements taken

- Let's take the resistor Johnson-Nyquist (Gaussian) result as an example assuming 10000 samples, 100 bins, and $\sigma = 1$ (for simplicity). We get the histogram below produced by two simple MATLAB commands: (1) x = randn(10000,1) and (2) hist(x,100). Note that the resulting histogram is clearly Gaussian in form.



- The Johnson-Nyquist result also *indicates a very important relationship between random processes and measurement sampling time in electrical engineering* defined by Nyquist's sampling relation: $B \sim f_S/2$
where f_S is the sampling rate (number of measurements per unit time).

- In RF and THz sensors of all sorts, the physical variables of interest are usually dynamic signals, meaning that they are electrical quantities that vary not only in an ensemble sense, but also vary randomly in time.
- The connection between the ensemble sense (represented by the pdf) and the time dependence (implicit in the Johnson-Nyquist theorem) is made by the *ergodic* hypothesis of probability theory:
The ensemble average of x is equivalent to the time average over a large number of sequential samples. i.e.,

$$< x >= \int_{x_{\min}}^{x_{\max}} xP(x)dx = \overline{x(t)} = \frac{1}{T}\int_{t_0}^{t_0+T} x(t)dt$$

where t_0 is an arbitrarily chosen time and T is long enough that a majority of the samples of the ensemble P(x) are sampled during this interval.

Quick Review of Probability Theory (cont)

- Not knowing the statistics in advance, and not knowing how long it takes before the ergodic principle becomes valid, the most rigorous way to deal with random variables is in the time-domain, starting with measurements of the autocorrelation function, leading to calculation of the generating function and ultimately the pdf.
- However, assuming that we know the statistics in advance, we can deal exclusively with the pdf and one other important quantity from probability theory called the power spectral density (or "power spectrum" for short), S(f).

$$S(f) = \frac{|x(f)x^*(f)|^2}{\Delta f}$$

where x(f) is the Fourier transform of x(t), and Δf is chosen small enough that |x(f)| is practically constant from f to f + Δf.
The power spectrum acknowledges the fact that in RF and THz systems of all sorts, it is the received electromagnetic energy, or energy per unit time (i.e., power), that matter most in the transfer of information (communicating) or making of a decision (sensing). This is a fundamental result of *information theory.*

- The seminal work of Shannon and Wiener in the 1940s proved the utility of the power-spectral approach to RF systems, and relegated time-domain probability to special cases and rigorous theoretical analysis. The capstone in this transition was the Wiener-Khinchin theorem, which states that the power spectral density is the Fourier transform of the autocorrelation function:

$$S(f) = \int_{-\infty}^{\infty} R(t) e^{j\omega t} dt$$ where R is the autocorrelation function

$$R(t) = \int_{-\infty}^{\infty} x(\tau)x(\tau+t)d\tau \approx \int_{-T/2}^{T/2} x(\tau)x(\tau+t)d\tau$$ for large T

- There are several assumptions behind the WK theorem, such as linearity, time-invariance, and memory-free (i.e., Markovian). But they are satisfied by most RF and THz systems
- Other factors that make the power-spectral approach so useful in systems:
  (a) Linear, band-limited components such as amplifiers do not change the statistics of electrical noise, just the variance, the bandwidth, and possibly the mean.
  (b) The transfer function, $h_x(f)$, for linear components is the same for the signal as the noise. Since noise is naturally spread over a very broad range in frequency, we are usually concerned about the equivalent noise bandwidth B_N of all components

$$B_N = \frac{\int_0^{\infty} |h_x(f)|^2 \, df}{|h_x|^2_{\max}}$$

Where $|h_x(f)|_{max}$ is the maximum value of the magnitude of the transfer function

Quick Review of Probability Theory (cont)

(c) The autocorrelation function for a truly Gaussian, Markovian process is the product of the variance, the observation time T (assuming it is long enough) and a Dirac delta function:

$$R(t) \approx \sigma^2 T\delta(t\,).$$

Physically, this says that a Gaussian random variable only correlates with itself over infinitisimal time offset. We can approximate the observation time in the time domain as the inverse bandwidth Δf in the frequency domain.
By the WK theorem, we then get

$$S(f) = \int_{-\infty}^{\infty} R(t)e^{j\omega t}dt = \int_{-\infty}^{\infty} \sigma^2 T\delta(t)e^{j\omega t}dt \approx \frac{\sigma^2}{\Delta f}$$

So, for example, form the Johnson-Nyquist theorem of current noise in a resistor, $\sigma^2 = <(\Delta i)^2> = 4k_B T\Delta f/R$, and

$S_i(f) = 4k_B T/R$, which is independent of frequency (i.e., *white*)

- These characteristics lead to a natural metric for system performance using the power spectral density method, the power signal-to-noise ratio:

$$\frac{P_S}{P_N} \equiv SNR = \frac{T^{-1}\int_{t_0}^{t_0+T} |x(t)|^2\, dt}{S_X(f)\cdot B_N}$$

where P_S is the average signal power and P_N is the average noise power

(important practical point: the dimension of P_S and P_N are not necessarily power, per se. For example, if x is a current, i, then the dimension of P_S is A^2, S_X is A^2/Hz, and P_N is A^2, so SNR is unitless as it should be !)

- In principle, the SNR can be defined uniquely at each and every point in any RF or THz system from the “front-end” (or RF band) to the “back-end” (or baseband).

- SNR is an important metric in any RF or THz system, but it is not specific enough for systems that are transmitting specific information (e.g., digital communications) or systems that have to help make decisions (e.g. radars). Such systems must generally provide *threshold* levels in terms of signal amplitude or power (e.g., a threshold between a “zero” and a “one” in a binary communications system.

Quick Review of Probability Theory (cont)

Cumulative Distribution Functions

• In threshold systems, we must deal with the noise at the ensemble level, knowing (or assuming) the pdf. For example, if the threshold for a particular system between a “zero” and a “one” (or between a “no target” and “yes target”) is set at Xt, then the probability that the noise itself stays below the threshold is

$$P_0 = \int_{X_{\min}}^{X_T} P(x)dx$$

and the probability that the noise fluctuations above the threshold is

$$P_1 = \int_{X_T}^{X_{\max}} P(x)dx$$

Since P(x) is normalized, we see that P_0 and P_1 are “complementary”, i.e., $P_0 + P_1 = 1$

• Because of the widespread utility of these partial integrals of pdf functions, they have a special name in probability theory called cumulative distribution functions (cdfs)

• The most common cdfs, widely used in RF and THz systems, are those that occur when the noise statistics is Gaussian

$$P_0 \equiv \int_{-\infty}^{X_T} P(x)dx = \frac{1}{2}[1+erf(X_T)] = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{X_T} e^{-x^2} dx$$

$$P_1 \equiv \int_{X_T}^{\infty} P(x)dx \equiv 1 - P_0 = \frac{1}{2}[1-erf(X_T)] = \frac{1}{2} erfc(X_T) \equiv \frac{1}{\sqrt{\pi}} \int_{X_T}^{\infty} e^{-x^2} dx$$

where *erf* is the “error function” and *erfc* is the “complementary error function”. Notice that both the *erf* and *erfc* are defined with respect to a one-sided Gaussian (centered and starting at x=0), whereas P_0 and P_1 are generally defined with respect to a two-sided Gaussian since electromagnetic signals are generally bipolar.

Quick Review of Probability Theory (cont)

Other Important Probability Density Functions

- In addition to the Gaussian, we will need a few other pdfs that commonly occur in RF and THz systems of all sorts.

(A) Complex, or two-dimensional Gaussian. This occurs when we have a random variable with two degrees of freedom each of which is individually Gaussian. A good example is the spatial electric field propagating far from a thermal source (where we can assume the radiation is approximately in the form of a plane wave) From fundamental electromagnetics, we know that we can vector decompose the total electrical field into two perpendicular components, say along x and y. But in thermal radiation, these components are statistically independent, and the component electric fields can be described by independent Gaussian pdfs. If they come from a source having a uniform emissivity and temperature, we expect the variance s along each axis to be the same so we can write the *joint pdf:*

$$P(x\,|\,y) \equiv \frac{1}{2\pi\sigma^2}\exp\{[-(x-x_0)^2-(y-y_0)^2]/2\sigma^2\}$$

Assuming x_0 and y_0 = 0, we can transform this to a more useful form through the change of variables: $r=\sqrt{x^2+y^2}$ $\quad \theta=\tan^{-1}(y/x)$

Statistical independence then dictates $P(\theta)=\frac{1}{2\pi}$ (i.e., all phase angles equally likely)

And a Jacobian transform yields $P(r)=\frac{r}{\sigma^2}\exp(-r^2/2\sigma^2)$

This P(r) is called the Rayleigh pdf. It comes up in evaluating the noise at the output of an envelope-type I-Q demodulator in RF and THz receivers.

Quick Review of Probability Theory (cont)

(B) Exponential $P(x)=\alpha^{-1}\cdot\exp(-x/\alpha),\quad x=0\to\infty$

<x> = α , variance = α^2 , standard deviation = α

(this is the general form of the Boltzmann or "canonical" pdf of statistical physics)
Note that the exponential is the only common pdf for which the standard deviation is equal to the mean

(C) Chi-Squared: $P(x)=\dfrac{x^{(k/2)-1}}{2^{k/2}\Gamma(k/2)}\exp(-x/2),\quad k=1,2,3...$ # "degrees of freedom"

where Γ is the gamma function – a common special function in statistical physics and probability theory. Its values for integral k can be gotten through the following rules: $\Gamma(1/2) = \pi^{1/2}$, $\Gamma(n) = (n-1)!$ for integer n (recalling that 0! = 1), and $\Gamma(y+1) = y\Gamma(y)$
Note that for k = 2, we get the exponential pdf as a special case

The form of this for k = 1 up in non-linear rectification, namely square-law detection

(D) Rice: $P(r\,|\,x_0)=\dfrac{r}{\sigma^2}I_o\left(\dfrac{r\cdot x_0}{\sigma^2}\right)\exp\left(-\dfrac{r^2+x_0^2}{2\sigma^2}\right)$

where I_0 is the modified Bessel function of order zero. The Rice pdf is a generalization of the Rayleigh in the presence of a bias x_0 along the x axis, where the bias is usually the baseband output of an RF or THz signal. As we shall see later, this is the pdf for describing a deterministic signal in the presence of noise at the output of an envelope-type I-Q demodulator. Note that it reduces to the Rayleigh joint pdf (as it should) when $x_0 = 0$ since $I_0(0) = 1$ (just like the ordinary Bessel function J_0)