EE 422G - Signals and Systems Laboratory

Lab 2 Noise Analysis

Written by
Kevin D. Donohue
Department of Electrical and Computer Engineering
University of Kentucky
Lexington, KY 40506
September 15, 2011

Objectives:

- Use Matlab to simulate noise signals.
- Use the Histogram, Autocorrelation, and Power Spectral Density to analyze noise.

1. Background

Noise is a process that cannot be predicted without error (using explicit functions/formulae). Noise is therefore characterized as a random variable (RV), which *is a function that maps an event into a real number*. The behavior of the RV associated with the noise can be described with a probability density function (PDF). Statistics are often used to describe noise amplitude fluctuations, such as means, variances, and root mean square (RMS) values. The most common noise model is a zero-mean and independent Gaussian process. The power of the noise signal is equivalent to the variance for the zero mean case (RMS equivalent to the standard deviation). If noise is uncorrelated from sample to sample, it does not influence other samples, and therefore cannot be estimated or predicted from neighboring samples. If, on the other hand, correlation exists between samples, then the correlation statistic can be used to predict a sample from its neighbors. So in addition to describing the signal amplitude fluctuations, correlation between time samples must also be characterized. This is done with the autocorrelation (AC) function in the time domain or the power spectral density (PSD) in the frequency domain. This lab focuses on simulating and analyzing noise signals with characteristics based on the PSD and PDF.

Probability Density Function:

Consider the event of making a measurement when no signal is present. Whatever is measured in this case is noise. Let $v[n]$ be a voltage value corresponding to a sample noise signal measurement whose amplitude distribution is described by the PDF $p_V(v)$. In the case of Gaussian distributed noise, the PDF is given by:

$$p_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(v-\mu)^2}{2\sigma^2}\right) \qquad (1)$$

where μ is the mean and σ is the standard deviation (σ^2 is the variance). The mean or expected value of v is computed from the PDF by:

$$\mu = \mathbf{E}[v] = \int_{-\infty}^{\infty} v p_v(v) dv \qquad (2)$$

where $\mathbf{E}[\cdot]$ is the expected value operator. If estimated from the data, the mean is computed from a simple average, often referred to as the sample mean:

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} v[n] \qquad (3)$$

Matlab's function *mean*() implements the formula in Eq. (3). The variance is computed from the PDF by:

$$\sigma^2 = \mathbf{E}\left[(v-\mu)^2\right] = \int_{-\infty}^{\infty} (v-\mu)^2 p_v(v) dv \qquad (4)$$

Equation (4) can be expressed in several other useful forms by applying the linearity properties of the expected value operator:

$$\sigma^2 = \mathbf{E}\left[\left(v^2 - 2\mu v + \mu^2\right)\right] = \mathbf{E}\left[v^2\right] - 2\mu\mathbf{E}[v] + \mu^2 = \mathbf{E}\left[v^2\right] - \mu^2 \qquad (5)$$

The variance can be computed from data (sample variance) with the following formula:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (v[n] - \hat{\mu})^2 = \frac{1}{N} \sum_{n=0}^{N-1} (v[n])^2 - \hat{\mu}^2 \qquad (6)$$

Matlab's function *std*() implements the above formula and then takes it square root to result in the standard deviation estimate (or equivalently the RMS value of a zero-mean signal). Note that the above equation is equivalent to the formula for power if the signal is zero-mean.

The probably that a RV occurs within a range of values is computed from the integral of the PDF. The probably of $a \le v < b$ is given by:

$$\Pr[a \le v < b] = \int_a^b p_V(v) dv \qquad (7)$$

The Gaussian distribution does not have a closed-form expression for its PDF integral (the cumulative distribution function (CDF)); however, tables and pre-computed functions exist from which these numbers can be obtained. The most famous is the error function, which Matlab has an implementation of it given by function *erf*(). This has the computed values for a Gaussian RV with zero mean and a standard deviation equal to the reciprocal of $\sqrt{2}$. The PDF is integrated from *-x* to *x*, where *x* is the argument of the function. So for a zero-mean Gaussian noise process with $\sigma = 1/\sqrt{2}$ the probability that $-b \le v < b$ can be computed from:

$$\Pr[-b \le v < b] = \operatorname{erf}(b) = \frac{2}{\sqrt{\pi}} \int_0^b \exp(-v^2) dv \qquad (8)$$

Instead of integrating from –b, the lower limit is set to 0 and the integral doubled to exploit the symmetry of the Gaussian PDF. For a general problem, the RV must be shifted to zero-mean and scaled down so its standard deviation becomes $\sqrt{2}$ so the error function can be applied. This scaling is applied to the original limits of integration and then strategically substituted into the error function of Eq. (8). For example, the probability that a Gaussian RV with mean 2 and standard deviation 3.5 is between a and b is given by:

$$\Pr[a \le v < b] = \frac{1}{2}\left(\mathrm{erf}\left(\frac{b-2}{3.5\sqrt{2}}\right) - \mathrm{erf}\left(\frac{a-2}{3.5\sqrt{2}}\right)\right). \tag{9}$$

Eq. (9) is equivalent to integrating the Gaussian PDF of mean 2 and standard deviation 3.5 over the interval from a to b.

The PDF can be estimated from sample data using the histogram and normalization. Recall the histogram finds the number of sample values occurring between a sequence of contiguous intervals referred to as bins. If enough samples fall in each bin, then dividing by the total number of samples collected provides a statistically stable estimate of the probability that a sample value occurs in that bin, which is directly related to the PDF. So with enough sample data, the PDF can be estimated by computing the histogram with sufficiently small bin intervals and dividing each bin count by the total number of samples. The Matlab functions hist() and histc() are useful functions in performing PDF estimation. It should be note however that bins containing only a few samples are unreliable estimates of the PDF. So less accuracy can be expected in the tails or low probability regions of the PDF, because fewer samples will occur there than in other parts of the PDF.

Power Spectral Density and the Autocorrelation

If a process has inertia/memory resulting in limited bandwidth, correlation will exist between time samples. Most practical random processes are correlated to some degree. A noise processes that is uncorrelated, is referred to as ***white noise***. If correlation exists between samples, the noise process is referred to as ***colored noise***. White noise has a flat average spectrum (all frequencies have the same expected power). This is analogous to white light, which contains all frequencies in the optical spectrum. Colored noise has a frequency emphasis and typically results from filtering colored noise with a band-limited filter.

The correlation between 2 signals is defined as the expected value between them, given by:

$$\sigma_{xy} = \mathbf{E}[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\,p_{XY}(x,y)\,dx\,dy, \tag{10}$$

where p_{xy} is the joint PDF between RVs x and y. The autocorrelation (AC) function describes the correlation between a signal's own samples over time. If the signal statistics do not change over time (i.e. a Gaussian process with a constant time-invariant mean and variance), it is referred to as a stationary process. The AC function of a stationary process is only a function of the time interval between the samples, independent of where in the time sequence they occur. The AC of a stationary signal is given by:

$$R(k) = \mathbf{E}[v[n]v[n-k]] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v[n]v[n-k]\,p_{v[n]v[n-k]}(v[n],v[n-k])\,dv[n]\,dv[n-k] \tag{11}$$