ECE 461: Digital Communications

Lecture 2: Statistical Channel Model

Introduction

We began our study of reliable communication last lecture with a very simple model of the additive noise channel. It works fine, except that one may haver a very conservative value for the noise fluctuations $\pm\sigma_{\mathrm{th}}$. This will lead to a correspondingly poor performance (in terms of number of reliable bits communicated for a given energy constraint). In this lecture, we take a more nuanced look at the additive noise channel model. Our basic goal is to have a *statistical* model of the additive noise. This will allow us to talk about reliable communication with a desired level of reliability (as opposed to the "fully reliable" notion of the previous lecture).

Statistical models can be arrived at by plain experiments of how the additive noise looks like and taking the *histogram* as the statistical model. Based on what this model and a desired reliability level, we could work out the appropriate value of σ_{th}. We could then directly use this choice of σ_{th} to our communication strategies from the previous lecture (transmit voltages as far apart from each other). While this already gives a significant benefit over and above the conservative estimates of the worst-case fluctuation σ_{th}, this may not be the optimal communication strategy (in terms of allowing largest number of bits for a given energy constraint and reliability level). We will see that depending on the exact shape of the histogram one can potentially do better. We will also see when the performance cannot be improved beyond this simple scheme for a wide range of histograms. Finally, we will see that most histograms that arise in nature are indeed of this type. Specifically, it turns out that most interesting noise models have the *same* statistical behavior with just two parameters that vary: the mean (first order statistics) and variance (second order statistics). So, we can design our communication schemes based on this *universal* statistical model and the performance only depends on two parameters: the mean and variance. This streamlines the communication design problem and allows the engineer to get to the heart of how the resources (power and bandwidth) can be used to get maximum performance (rate and reliability).

We start out with a set of properties that most additive noise channels tend to have. Next, we will translate these properties into an appropriate mathematical language. This will allow us to arrive at a robust *universal* statistical model for additive noise: it is *Gaussian* or *normally* distributed. We will see that our understanding of transmission and reception strategies using the deterministic model from the previous lecture extends naturally to one where the model is statistical.

Histogram Models and Reliable Communication Strategies

Suppose we make detailed measurements of the noise values at the location where we expect communication to take place. Suppose we have made N separate measurements, where N is a large value (say, 10,000): $v_1, \ldots, v_N$. The *histogram* of the noise based on the measurements

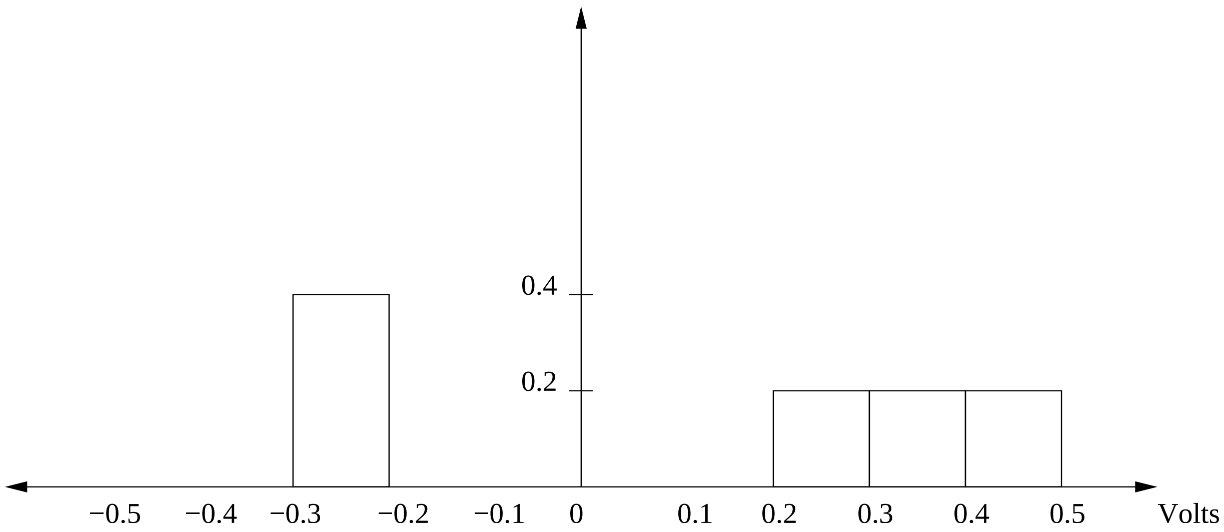

Figure 1: A exemplar histogram.

at a resolution level of Δ is simply a function from voltage levels to the real numbers: for every $a \in (m\Delta, (m+1)\,\Delta)$,

$$f_\delta(a) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_{v_k \in (m\Delta,(m+1)\Delta)}, \tag{1}$$

where we have denoted the *indicator function*

$$\mathbf{1}_{\cdot} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if the parameter is true} \\ 0 & \text{else.} \end{cases} \tag{2}$$

One important property of the histogram function is that the area under the histogram curve is equal to unity. For example, with $N = 5$ and $v_1 = 0.2V, v_2 = -0.25V, v_3 = 0.45V, v_4 = -0.27V, v_5 = 0.37V$, the histogram at a resolution of $\Delta = 0.1V$ is depicted in Figure 1. In the limit of very large number of samples N and very small resolution Δ, the histogram function is called the *density* of the noise. Henceforth we will use the term density to denote the histogram created from the noise measurements. As any histogram, the density function is always non-negative and the area under it is unity. The density function of a noise that takes any voltage value in the range $[-0.5V, 0.5V]$ equally likely is depicted in Figure 2.

Now suppose we are willing to tolerate errors in communication a fraction η of the time. Then we can pick the smallest value of σ_{th} such that the area under the density function over the range $\pm\sigma_{\text{th}}$ is at least $1 - \eta$. This ensures that the noise is within σ_{th} at least a fraction $1 - \eta$ of the time. For the density function in Figure 2, a value of $\eta = 0.1$ means that $\sigma_{\text{th}} = 0.45V$; a pictorial depiction is available in Figure 3.

We can now pick the transmission and reception schemes as in Lecture 1 using this new value of $\sigma_{\text{th}} = 0.45V$. We are now guaranteed reliable communication at a level of tolerable unreliability $\eta = 0.1$. This corresponds to a saving in energy of a fraction

$$\frac{0.05V}{0.5V} = 10\%. \tag{3}$$

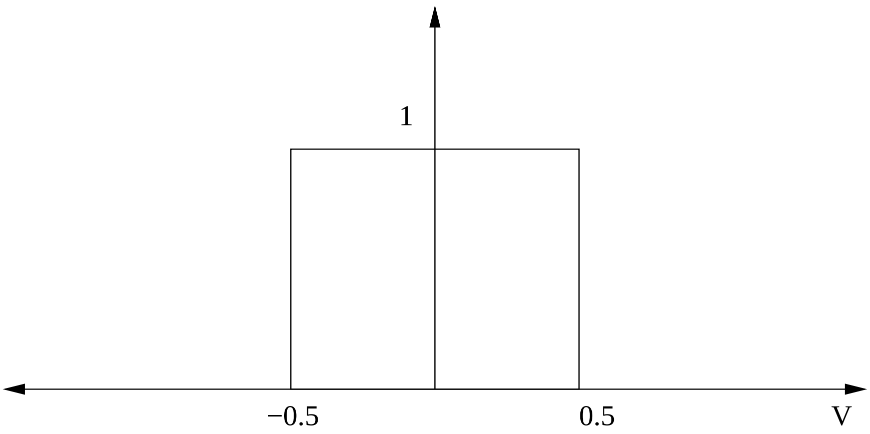

Figure 2: A uniform density function.

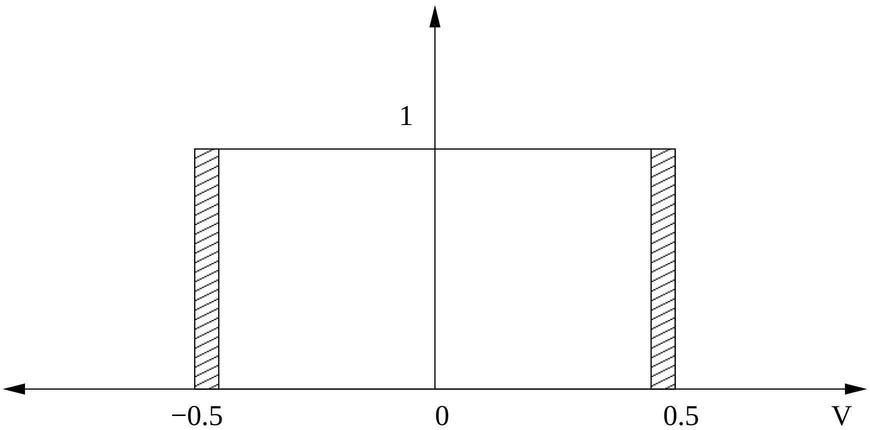

Figure 3: Choosing a threshold based on the reliability level.

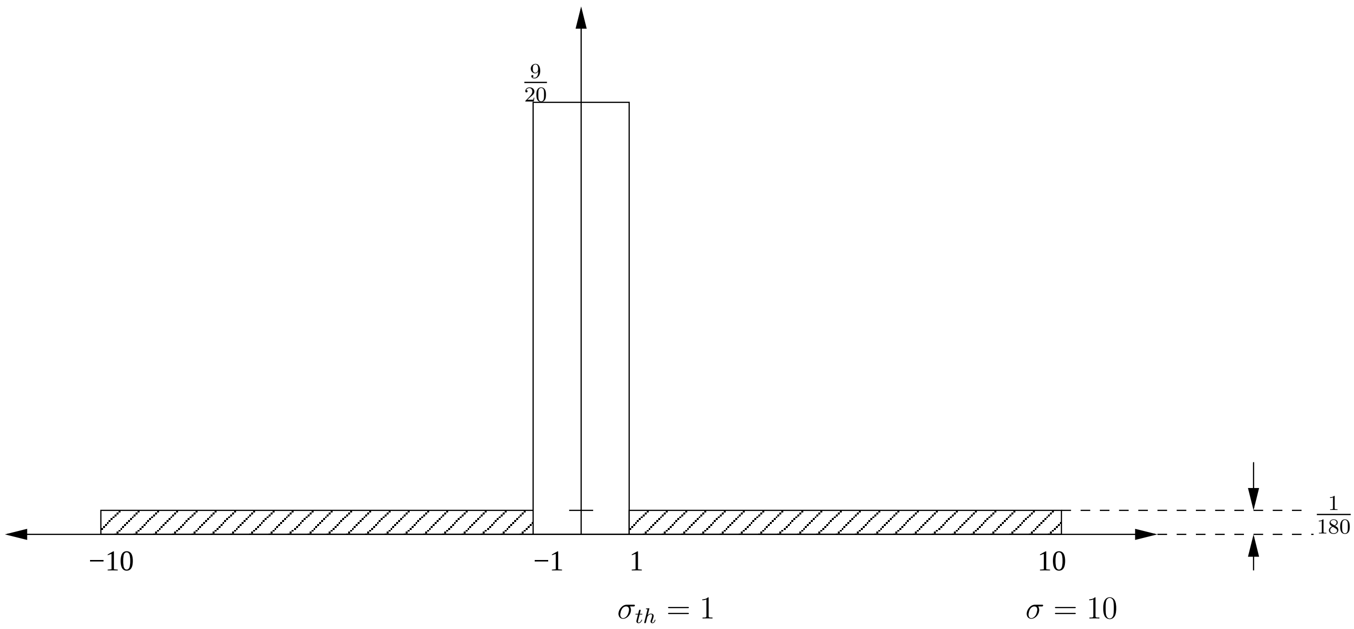

Figure 4: The threshold based on the reliability level can be significantly smaller than one based on worst-case.

While this might seem modest, consider the density function in Figure 4, where σ_{th} in the usual sense of Lecture 1 would be $10V$. On the other hand with $\eta = 0.1$, the new value of σ_{th} is only $1.0V$. This corresponds to a savings in energy of a fraction

$$\frac{9V}{10V} = 90\%, \tag{4}$$

a remarkably large fraction!

In the transmission scheme of Lecture 1, we picked the different possible transmit voltage levels to be spaced by at least $2\sigma_{\text{th}}$. This seems reasonable since we know a bound for how much the noise can fluctuate. But we have more knowledge about how the noise fluctuates based on the density function. This provokes us to think along the following natural thought process:

> **Question**: Given the noise density function and energy and reliability constraints, is the scheme of keeping the different transmit voltages apart by $2\sigma_{\text{th}}$ the best one, in terms of giving maximum number of bits?

It turns out that the answer is *no*. A homework question explores this subject in detail; there we see that it might be rather sub-optimal to keep the spacing between different transmit voltages as large as $2\sigma_{\text{th}}$, even when σ_{th} is chosen appropriately based on the density function and the reliability constraint. But for a large class of density functions, this is not the case: the natural approach of extracting the appropriate σ_{th} from the density function to use in the design of Lecture 1 suffices. Interestingly, it turns out that most density functions for *additive* noise have this property. In the rest of this lecture, we will study some canonical properties of the density of additive noise; we start with some simple physical properties.

Physical Properties of Additive Noise

An enumeration of some reasonable properties we may anticipate the additive forms of noise to take is the following.

1. The noise is the overall result of many additive "sub-noises". Typical sub-noises could be the result of thermal noise, device imperfections and measurement inaccuracies. additive noise w can be written as
2. These sub-noises typically have little correlation with respect to each other. We suppose the stronger statement: they are statistically *independent* of each other.
3. No sub-noise is particularly dominant over the other. In other words, they all contribute about the same to the total noise.
4. Finally, there are many sources of sub-noises.

We will work to convert these physical properties into more precise mathematical statements shortly.

Representation of Additive Noise

Using some notation, we can write the total additive noise w as

$$w = n_1 + n_2 + \cdots + n_m, \tag{5}$$

the sum of m sub-noises $n_1, \ldots, n_m$. Furthermore, the sub-noises $n_1, \ldots, n_m$ are statistically independent of each other. Let us denote the densities of the sub-noises as $f_{n_1}(\cdot), \ldots, f_{n_m}(\cdot)$, respectively. An important result from your prerequisite probability class is the following result:

> The density of the total noise w is the *convolution* of the densities of the sub-noises.

This result is best understood in the Laplace or Fourier domain. Specifically, the Laplace transform of a density function $f_w(\cdot)$ is defined as

$$F_w(s) = \int_{-\infty}^{\infty} e^{-sa} f_w(a)\, da \quad \forall s \in \mathbb{C}. \tag{6}$$

Here $\mathbb{C}$ is the complex plane. In terms of the Laplace transforms of the densities of each of the sub-noises,

$$F_w(s) = \prod_{k=1}^{m} F_{n_k}(s), \quad \forall s \in \mathbb{C}. \tag{7}$$

We know what the density function of a noise is from an engineering stand point: it is simply the histogram of a lot of noise measurements at a fine enough resolution level. How does one understand the Laplace transform of the density function from an engineering and physical view point? We can do a Taylor series expansion around $s = 0$ to get a better view of the Laplace transform of a density function:

$$F_w(s) = F_w(0) + sF_w^{'}(0) + \frac{s^2}{2}F_w^{''}(0) + o(s^2), \tag{8}$$

where the function $o(s^2)$ denotes a function of s^2 that when divided by s^2 goes to zero as s approaches zero itself. The first term

$$\begin{aligned} F_w(0) &= \int_{-\infty}^{\infty} f_w(a)\, da && (9) \\ &= 1, && (10) \end{aligned}$$

since the area under a density function is unity. The second term can be calculated as

$$\begin{aligned} \frac{d}{ds}F_w(s) &= \int_{-\infty}^{\infty} -ae^{-sa} f_w(a)\, da, && (11) \\ F_w^{'}(0) &= \int_{-\infty}^{\infty} af_w(a)\, da && (12) \\ &\stackrel{\text{def}}{=} \mathbb{E}\left[w\right]. && (13) \end{aligned}$$

The quantity $\mathbb{E}[w]$ is the *mean* of the noise w and is a readily measured quantity: it is just the *average* of all the noise measurements. In the sequence above, we blithely interchanged the differentiation and integration signs. mathematically speaking this step has to be justified more carefully. This will take us somewhat far from our main goal and we will not pursue this too much here.

Now for the third term:

$$\begin{aligned} \frac{d^2}{ds^2}F_w(s) &= \int_{-\infty}^{\infty} a^2e^{-sa} f_w(a)\, da, && (14) \\ F_w^{''}(0) &= \int_{-\infty}^{\infty} a^2 f_w(a)\, da && (15) \\ &= \mathbb{E}\left[w^2\right]. && (16) \end{aligned}$$

Here the quantity $\mathbb{E}[w^2]$ is the *second moment* of the noise w and is a readily measured quantity: it is just the *average* of the square of the noise measurements. Again, we have interchanged the differentiation and integration signs in the calculation above.

In conclusion, the first few terms of the Taylor series expansion of the Laplace transform of the density of the additive noise w involves easily measured quantities: mean and second moment. Sometimes the second moment is also calculated via the *variance*:

$$\mathrm{Var}(w) \stackrel{\text{def}}{=} \mathbb{E}\left[w^2\right] - \left(E\left[w\right]\right)^2. \tag{17}$$

These two quantities, the mean and variance, are also referred to simply as *first* and *second order statistics* of the measurements and are fairly easily calculated. Let us denote these two quantities by μ and σ^2_{th}, respectively henceforth. While we may not have access to the densities of the individual sub-noises, we can calculate their first and second order statistics by using the assumption that each of the sub-noises contributes the same level to the total noise. This means that, since

$$\mathbb{E}\left[w\right] = \sum_{k=1}^{m} \mathbb{E}\left[n_k\right], \tag{18}$$

we can say that

$$E\left[n_k\right] = \frac{\mu}{m}, \quad \forall k = 1 \ldots m. \tag{19}$$

Similarly for statistically independent sub-noises $n_1, \ldots, n_m$ we have

$$\mathrm{Var}\left(w\right) = \sum_{k=1}^{m} \mathrm{Var}\left(n_k\right), \tag{20}$$

we can say that

$$\mathrm{Var}\left(n_k\right) = \frac{\sigma^2}{m}, \quad \forall k = 1 \ldots m, \tag{21}$$

$$\mathbb{E}\left[n_k^2\right] = \frac{\sigma^2}{m} + \frac{\mu^2}{m^2}. \tag{22}$$

Here we used Equation (17) in arriving at the second step.

Now we can use an approximation as in Equation (8), by ignoring the higher order terms, to write

$$F_{n_k}(s) \approx 1 - \frac{\mu s}{m} + \frac{\sigma^2 s}{2m} + \frac{\mu^2 s^2}{2m^2}, \quad \forall k = 1 \ldots m. \tag{23}$$

Substituting this into Equation (7), we get

$$F_w(s) \approx \left(1 - \frac{\mu s}{m} + \frac{\sigma^2 s^2}{2m} + \frac{\mu^2 s^2}{2m^2}\right)^m, \quad \forall s \in \mathbb{C}. \tag{24}$$

We are interested in the density function of the noise w for large number of sub-noises, i.e., when m is large. From elementary calculus techniques, we know the limiting formula:

$$\lim_{m \to \infty} F_w(s) = e^{-\mu s + \frac{s^2 \sigma^2}{2}}, \quad \forall s \in \mathbb{C}. \tag{25}$$

Remarkably, we have arrived at a *universal* formula for the density function that is parameterized by only two simply measured physical quantities: the first order and second order statistics (mean μ and variance σ^2, respectively). This calculation is known, esoterically, as the *central limit theorem*.

It turns out that the density function whose Laplace transform corresponds to the one in Equation (25) is

$$f_w(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}, \quad \forall a \in \mathbb{R}. \tag{26}$$

This density function is called *Gaussian*, in honor of the first person who discovered it. It is also called the *normal* density since it also shows up in many real world situations which are entirely unrelated to additive noises (all the way from temperature measurements to weights of people to the eventual grades of the students in this course (hopefully!), are all "normally" behaved).

There are some important modern day data that are famously *not* normal: size of packets on the internet and the number of goods bought in an online store. I recommend the recent books

> C. Anderson, *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion, 2006;

and

> Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, 2007,

that make for quite interesting reading (unrelated to the scope of this course). You can also get a broader feel for how such measurements are harnessed in making engineering and economic decisions.

Looking Forward

In the next lecture we will see how to use this particular structure of the density function in choosing our communication transmit and receive strategies.