

MIT 6.02 DRAFT Lecture Notes
Fall 2011 (Last update: March 5, 2012)
Comments, questions or bug reports?
Please contact {hari,verghese} at mit.edu

CHAPTER 9

Noise

Liars, d—d liars, and experts.

—possibly Judge George Bramwell (quoted in 1885), expressing his opinion of witnesses

There are three kinds of lies: lies, damned lies, and statistics.

—definitely Mark Twain (writing in 1904), in a likely misattribution to Benjamin Disraeli

God does not play dice with the universe.

—Albert Einstein, with probability near 1

In general, many independent factors affect a signal received over a channel. Those that have a repeatable, deterministic effect from one transmission to another are generally referred to as *distortion*. We shall examine a very important class of distortions—those induced by *linear, time-invariant channels*—in later chapters. Other factors have effects that are better modeled as *random*, and we collectively refer to them as **noise**. Communication systems are no exception to the general rule that any system in the physical world must contend with noise. In fact, noise is a fundamental aspect of all communication systems.

In the simplest binary signaling scheme—which we will invoke for most of our purposes in this course—a communication system transmits one of two voltages, mapping a “0” to the voltage V_0 and mapping a “1” to V_1 . The appropriate voltage is held steady over a fixed-duration time slot that is reserved for transmission of this bit, then moved to the appropriate voltage for the bit associated with the next time slot, and so on. We assume in this chapter that any distortion has been compensated for at the receiver, so that in an ideal noise-free case the receiver ends up measuring V_0 in any time slot corresponding to a “0”, and V_1 in any slot corresponding to a “1”.

In this chapter we focus on the case where $V_1 = V_p > 0$ and $V_0 = -V_p$, where V_p is some fixed positive voltage, typically the *peak* voltage magnitude that the transmitter is capable of imposing on the communication channel. This scheme is sometimes referred to as *bipolar signaling* or **bipolar keying**. Other choices of voltage levels are possible, of course.

In the presence of noise, the receiver measures a sequence of voltage samples $y[k]$ that is unlikely to be exactly V_0 or V_1 . To deal with this variation, we described in the previous chapter a simple and intuitively reasonable decision rule, for the receiver to infer whether the bit transmitted in a particular time slot was a “0” or a “1”. The receiver first chooses a single voltage sample from the sequence of received samples within the appropriate time slot, and then compares this sample to a threshold voltage V_t . Provided “0” and “1” are equally likely to occur in the sender’s binary stream, it seems reasonable that we should pick as our threshold the voltage that “splits the difference”, i.e., use $V_t = (V_0 + V_1)/2$. Then, assuming $V_0 < V_1$, return “0” as the decision if the received voltage sample is smaller than V_t , otherwise return “1”.

The receiver could also do more complicated things; for example, it could form an average or a weighted average of all the voltage samples in the appropriate time slot, and then compare this *average* with the threshold voltage V_t . Though such averaging leads in general to improved performance, we focus on the simpler scheme, where a single well-selected sample in the time slot is compared with V_t . In this chapter we will analyze the performance of this decision rule, in terms of the probability of an incorrect decision at the receiver, an event that would manifest itself as a *bit error* at the receiver.

The key points of this chapter are as follows:

1. A simple model—and often a good model—for the net effect at the receiver of noise in the communication system is to assume **additive, Gaussian** noise. In this model, each received signal *sample* is the sum of two components. The first component is the deterministic function of the transmitted signal that would be obtained in the absence of noise. (Throughout this chapter, we will assume no distortion in the channel, so the deterministic function referred to here will actually produce at the receiver exactly the same sample value transmitted by the sender, under the assumption of no noise.) The second component is the noise term, and is a quantity drawn from a Gaussian probability distribution with mean 0 and some variance, *independent of the transmitted signal*. The Gaussian distribution is described in more detail in this chapter.

If this Gaussian noise variable is also independent from one sample to another, we describe the underlying noise process as *white* Gaussian noise, and refer to the noise as *additive white Gaussian noise (AWGN)*; this is the case we will consider. The origin of the term “white” will become clearer when we examine signals in the frequency domain, later in this course. The variance of the zero-mean Gaussian noise variable at any sample time for this AWGN case reflects the **power** or intensity of the underlying white-noise process. (By analogy with what is done with electrical circuits or mechanical systems, the term “power” is generally used for the *square* of a signal magnitude. In the case of a random signal, the term generally denotes the *expected* or *mean* value of the squared magnitude.)

2. If the sender transmitted a signal corresponding to some bit, b , and the receiver measured its voltage as being on the correct side of the threshold voltage V_t , then the bit would be received correctly. Otherwise, the result is a **bit error**. The probability of a bit error is an important quantity, which we will analyze. This probability, typically called the **bit error rate (BER)**, is related to the probability that a Gaussian ran-

dom variable exceeds some level; we will calculate it using the probability density function (PDF) and cumulative distribution function (CDF) of a Gaussian random variable. We will find that, for the bipolar keying scheme described above, when used with the simple threshold decision rule that was also specified above, the BER is determined by the ratio of two quantities: (i) the power or squared magnitude, V_p^2 , of the received sample voltage in the noise-free case; and (ii) the power of the noise process. This ratio is an instance of a **signal-to-noise ratio (SNR)**, and such ratios are of fundamental importance in understanding the performance of a communication system.

3. At the *signal* abstraction, additive white Gaussian noise is often a good noise model. At the *bit* abstraction, this model is inconvenient because we would have to keep going to the signal level to figure out exactly how it affects every bit. Fortunately, the BER allows us to think about the impact of noise in terms of how it affects bits. In particular, a simple, but powerful, model at the bit level is that of a **binary symmetric channel (BSC)**. Here, a transmitted bit b (0 or 1) is interpreted by the receiver as $1 - b$ with probability p_e and interpreted as b with probability $1 - p_e$, where p_e is the probability of a bit error (i.e., the bit error rate). In this model, each bit is corrupted independently of the others, and the probability of corruption is the same for all bits (so the noise process is an example of an “iid” random process: “*independent and identically distributed*”).

■ 9.1 Origins of noise

A common source of noise in radio and acoustic communications arises from interferers who might individually or collectively make it harder to pick out the communication that the receiver is primarily interested in. For example, the quality of WiFi communication is affected by other WiFi communications in the same frequency band (later in the course we will develop methods to mitigate such interference), an example of interference from other users or nodes in the same network. In addition, interference could be caused by sources *external* to the network of interest; WiFi, for example, is affected by cordless phones, microwave ovens, Bluetooth devices, and so on that operate at similar radio frequencies. Microwave ovens are doubly troublesome if you’re streaming music over WiFi, which in the most common mode runs in the 2.4 GHz frequency band today—not only do microwave ovens create audible disturbances that affect your ability to listen to music, but they also radiate power in the 2.4 GHz frequency band. This absorption is good for heating food, but leakage from ovens interferes with WiFi receptions! In addition, wireless communication networks like WiFi, long-range cellular networks, short-range Bluetooth radio links, and cordless phones all suffer from *fading*, because users often move around and signals undergo a variety of reflections that interfere with each other (a phenomenon known as “multipath fading”). All these factors cause the received signal to be different from what was sent.

If the communication channel is a wire on an integrated circuit, the primary source of noise is capacitive coupling between signals on neighboring wires. If the channel is a wire on a printed circuit board, signal coupling is still the primary source of noise, but coupling between wires is largely inductive or carried by unintended electromagnetic radiation.

In both these cases, one might argue that the noise is not truly random, as the signals generating the noise are under the designer's control. However, a signal on a wire in an integrated circuit or on a printed circuit board will frequently be affected by signals on thousands of other wires, so approximating the interference using a random noise model turns out to work very well.

Noise may also arise from truly random physical phenomena. For example, electric current in an integrated circuit is generated by electrons moving through wires and across transistors. The electrons must navigate a sea of obstacles (atomic nuclei), and behave much like marbles traveling through a Pachinko machine. They collide randomly with nuclei and have transit times that vary randomly. The result is that electric currents have random noise. In practice, however, the amplitude of the noise is typically several orders of magnitude smaller than the nominal current. Even in the interior of an integrated circuit, where digital information is transported on micron-wide wires, the impact of electron transit time fluctuations is negligible. By contrast, in optical communication channels, fluctuations in electron transit times in circuits used to convert between optical and electronic signals at the ends of the fiber are the dominant source of noise.

To summarize: there is a wide variety of mechanisms that can be the source of noise; as a result, the bottom line is that *it is physically impossible to construct a noise-free channel*. By understanding noise and analyzing its effects (bit errors), we can develop approaches to reducing the probability of errors caused by noise and to combat the errors that will inevitably occur despite our best efforts. We will also learn in a later chapter about a celebrated and important result of Shannon: provided the information transmission rate over a channel is kept below a limit referred to as the *channel capacity* (determined solely by the distortion and noise characteristics of the channel), we can transmit in a way that makes the probability of error in decoding the sender's message vanish asymptotically as the message size goes to ∞ . This asymptotic performance is attained at the cost of increasing computational burden and increasing delay in deducing the sender's message at the receiver. Much research and commercial development has gone into designing practical methods to come close to this "gold standard".

■ 9.2 Additive White Gaussian Noise: A Simple but Powerful Model

We will posit a simple model for how noise affects the reception of a signal sent over a channel and processed by the receiver. In this model, noise is:

1. **Additive:** Given a received sample value $y[k]$ at the k th sample time, the receiver interprets it as the *sum* of two components: the first is the *noise-free component* $y_0[k]$, i.e., the sample value that would have been received at the k th sample time in the absence of noise, as a result of the input waveform being passed through the channel with only distortion present; and the second is the *noise component* $w[k]$, assumed independent of the input waveform. We can thus write

$$y[k] = y_0[k] + w[k] . \tag{9.1}$$

In the absence of distortion, which is what we are assuming here, $y_0[k]$ will be either V_0 or V_1 .

2. **Gaussian:** The noise component $w[k]$ is random, but we assume it is drawn at each sample time from a fixed Gaussian distribution; for concreteness, we take this to be the distribution of a Gaussian random variable W , so that each $w[k]$ is distributed exactly as W is. The reason why a Gaussian makes sense is because noise is often the result of summing a large number of different and independent factors, which allows us to apply an important result from probability and statistics, called the *central limit theorem*. This states that the sum of independent random variables is well approximated (under rather mild conditions) by a Gaussian random variable, with the approximation improving as more variables are summed in.

The Gaussian distribution is beautiful from several viewpoints, not least because it is characterized by just two numbers: its *mean* μ , and its *variance* σ^2 or *standard deviation* σ . In our noise model, we will assume that the mean of the noise distribution is 0. This assumption is not a huge concession: any consistent non-zero perturbation is easy to compensate for. For zero-mean Gaussian noise, the variance, or equivalently the standard deviation, completely characterizes the noise. The standard deviation σ may be thought of as a measure of the expected “amplitude” of the noise; its square captures the expected power.

For noise not to corrupt the digitization of a bit detection sample, the distance between the noise-free value of the sample and the digitizing threshold should be sufficiently larger than the expected amplitude—or standard deviation—of the noise.

3. **White:** This property concerns the temporal variation in the individual noise samples that affect the signal. If these Gaussian noise samples are independent from one sample to another, the underlying noise process is referred to as white Gaussian noise. “White” refers to the frequency decomposition of the sequence of noise samples, and essentially says that the noise signal contains components of equal expected power at all frequencies. This statement will become clearer later in the course when we talk about the frequency content of signals.

This noise model is generally given the term **AWGN**, for *additive white Gaussian noise*. We will use this term.

■ 9.2.1 Estimating the Noise Parameters

It is often of interest to estimate the noise parameters from measurements; in our Gaussian model, these are the parameters μ and σ^2 . If we simply transmit a sequence of “0” bits, i.e., hold the voltage V_0 at the transmitter, and observe the received samples $y[k]$ for $k = 0, 1, \dots, K-1$, we can process these samples to obtain the statistics of the noise process for additive noise. Under the assumption of no distortion, and constant (or “stationary”) noise statistics, and noise samples $w[k] = y[k] - V_0$ that are independent from one sampling instant to another, we can use the *sample mean* m to estimate μ , where

$$m = \frac{1}{K} \sum_{k=0}^{K-1} w[k]. \quad (9.2)$$

The *law of large numbers* from probability and statistics ensures that as K tends to ∞ , the sample mean m converges to μ , which we have assumed is 0.

With $\mu = 0$, the quantity that is more indicative of the power of the noise is the variance σ^2 , which can be estimated by the *sample variance* s^2 , given by

$$s^2 = \frac{1}{K} \sum_{k=0}^{K-1} (w[k] - m)^2 . \quad (9.3)$$

Again, this converges to σ^2 as K tends to ∞ .

■ 9.2.2 The Gaussian Distribution

Let us now understand the Gaussian distribution in the context of our physical communication channel and signaling process. In our context, the receiver and sender both deal with voltage samples. The sample $y[k]$ at the receiver has a noise term, $w[k]$, contributing to it additively, where $w[k]$ is obtained from the following *probability density function* (PDF), which specifies a Gaussian distribution:

$$f_W(w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w-\mu)^2}{2\sigma^2}} . \quad (9.4)$$

For zero-mean noise, $\mu = 0$.

The PDF $f_W(w)$, which is assumed to govern the distribution of all the noise samples $w[k]$, specifies the probability that W , or equivalently $w[k]$, takes values in the vicinity of w . Specifically,

$$\mathbb{P}(w \leq w[k] \leq w + dw) \approx f_W(w) dw .$$

More generally, the probability that $w[k]$ is between two values w_1 and w_2 is given by

$$\mathbb{P}(w_1 < w[k] \leq w_2) = \int_{w_1}^{w_2} f_W(w) dw .$$

The reason we use the PDF rather than a discrete histogram is that our noise model is inherently “analog”, taking on any real value in $(-\infty, \infty)$. For a noise sample that can take on any value in a continuous range, the natural mathematical tool to use is a continuous-domain random variable, described via its PDF, or via the integral of the PDF, which is called the cumulative distribution function (CDF).

It will be helpful to review the basic definitions and properties of continuous-domain random variables, especially if you aren’t comfortable with these tools. We have provided a brief recap and tutorial in the appendix near the end of this chapter (§9.6).

■ 9.3 Bit Errors

Noise disrupts the quality of communication between sender and receiver because the received noisy voltage samples can cause the receiver to incorrectly identify the transmitted bit, thereby generating a **bit error**. If we transmit a long stream of known bits and count the fraction of received bits that are in error, we obtain a quantity that—by the law of large numbers—asymptotically approaches the **bit error rate (BER)**, which is the *probability that*

any given bit is in error, $\mathbb{P}(\text{error})$. This is the probability that noise causes a transmitted “1” to be reported as “0” or vice versa.

Communication links exhibit a wide range of bit error rates. At one end, high-speed (multiple gigabits per second) fiber-optic links implement various mechanisms that reduce the bit error rates to be as low as 1 in 10^{12} . This error rate looks exceptionally low, but a link that can send data at 10 gigabits per second with such an error rate will encounter a bit error every 100 seconds of continuous activity, so it does need ways of masking errors that occur. Wireless communication links usually have errors anywhere between 1 in 10^3 for relatively noisy environments, down to 1 in 10^7 , and in fact allow the communication to occur at different bit rates; higher bit rates are usually also associated with higher bit error rates. In some applications, very noisy links can still be useful even if they have bit error rates as high as 1 in 10^3 or 10^2 .

We now analyze the BER of the simple binary signaling scheme. Recall the receiver thresholding rule, assuming that the sender sends V_0 volts for “0” and $V_1 > V_0$ volts for “1” and that there is no channel distortion (so in the absence of noise, the receiver would see exactly what the sender transmits):

If the received voltage sample $y < V_t = (V_0 + V_1)/2$ then the received bit is reported as “0”; otherwise, it is reported as “1”.

For simplicity, we will assume that the *prior* probability of a transmitted bit being a “0” is the same as it being a “1”, i.e., both probabilities are 0.5. We will find later that when these two prior probabilities are equal, the choice of threshold V_t specified above is the one that minimizes the overall probability of bit error for the decision rule that the receiver is using. When the two priors are unequal, one can either stick to the same threshold rule and calculate the bit error probability, or one could calculate the threshold that minimizes the error probability and then calculate the resulting bit error probability. We will deal with that case in the next section.

The noise resilience of the binary scheme turns out to depend only on the difference $V_1 - V_0$, because the noise is additive. It follows that if the transmitter is constrained to a peak voltage magnitude of V_p , then the best choice of voltage levels is $V_1 = V_p > 0$ and $V_0 = -V_p$, which corresponds to binary keying. The associated threshold is $V_t = 0$. This is the case that we analyze now.

As noted earlier, it is conventional to refer to the square of a magnitude as the *power*, so V_p^2 is the power associated with each voltage sample at the receiver, under the assumption of no distortion, and in the ideal case of no noise. Summing the power of these samples over all T samples in the time slot associated with a particular bit sent over the link yields the **energy per transmitted bit**, $T \cdot V_p^2$. It is thus reasonable to also think of V_p^2 as the **sample energy**, which we shall denote by E_s . With this notation, the voltage levels in bipolar keying can be written as $V_1 = +\sqrt{E_s}$ and $V_0 = -\sqrt{E_s}$.

Now consider in what cases a bit is incorrectly decided at the receiver. There are two mutually exclusive possibilities:

1. The sender sends $b = 0$ at voltage $-\sqrt{E_s}$ and the value received is > 0 ; or
2. The sender sends $b = 1$ at voltage $+\sqrt{E_s}$ and the value received is < 0 .

For a source that is equally likely to send 0’s and 1’s, and given the symmetry of a zero-mean Gaussian about the value 0, *the two events mentioned above have exactly the same proba-*

bilities. Each one of the events has a probability that is half the probability of a zero-mean Gaussian noise variable W taking values larger than $\sqrt{E_s}$ (the “half” is because the probability of $b = 0$ is 0.5, and similarly for $b = 1$). Hence the probability of one or the other of these mutually exclusive events occurring, i.e., the probability of a bit error, is simply the sum of these two probabilities, i.e., the BER is given by the probability of a zero-mean Gaussian noise variable W taking values larger than $\sqrt{E_s}$. The BER is therefore

$$\text{BER} = \mathbb{P}(\text{error}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\sqrt{E_s}}^{\infty} e^{-w^2/(2\sigma^2)} dw. \quad (9.5)$$

We will denote $2\sigma^2$ by N_0 . It has already been mentioned that σ^2 is a measure of the expected power in the underlying AWGN process. However, the quantity N_0 is also often referred to as the **noise power**, and we shall use this term for N_0 too.¹

After a bit of algebra, Equation (9.5) simplifies to

$$\text{BER} = \mathbb{P}(\text{error}) = \frac{1}{\sqrt{\pi}} \cdot \int_{\sqrt{E_s/N_0}}^{\infty} e^{-v^2} dv. \quad (9.6)$$

This equation specifies the *tail probability* of a Gaussian distribution, which turns out to be important in many scientific and engineering applications. It’s important enough to be tabulated using two special functions called the *error function* and the *complementary error function*, denoted $\text{erf}(z)$ and $\text{erfc}(z) = 1 - \text{erf}(z)$ respectively, and defined thus:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \cdot \int_0^z e^{-v^2} dv, \quad (9.7)$$

and

$$\text{erfc}(z) = 1 - \text{erf}(z) = \frac{2}{\sqrt{\pi}} \cdot \int_z^{\infty} e^{-v^2} dv. \quad (9.8)$$

One can now easily write the following important (and pleasingly simple) equation for the BER of our simple binary signaling scheme over an AWGN channel:

$$\text{BER} = \mathbb{P}(\text{error}) = \frac{1}{2} \text{erfc}\left(\sqrt{\frac{E_s}{N_0}}\right). \quad (9.9)$$

Equation (9.9) is worth appreciating and perhaps even committing to memory (at least for the duration of the course!). But it is more important to understand how we arrived at it and why it makes sense. The BER for our bipolar keying scheme with the specified decision rule at the receiver is determined entirely by the ratio $\frac{E_s}{N_0}$. The numerator of this ratio is the power of the signal used to send a bit, or equivalently the power or energy E_s of the voltage *sample* selected from the corresponding time slot at the receiver in the noise-free case, assuming no distortion (as we are doing throughout this chapter). The denominator of the ratio is the noise power N_0 encountered during the reception of the signal. This ratio is also commonly referred to as the **signal-to-noise ratio (SNR)** of the

¹The factor of 2 between the two uses of the term arises from the fact that under one notational convention the distribution of expected noise power over frequency is examined over both negative and positive frequencies, while under the other convention it is examined over just positive frequencies—but this difference is immaterial for us.

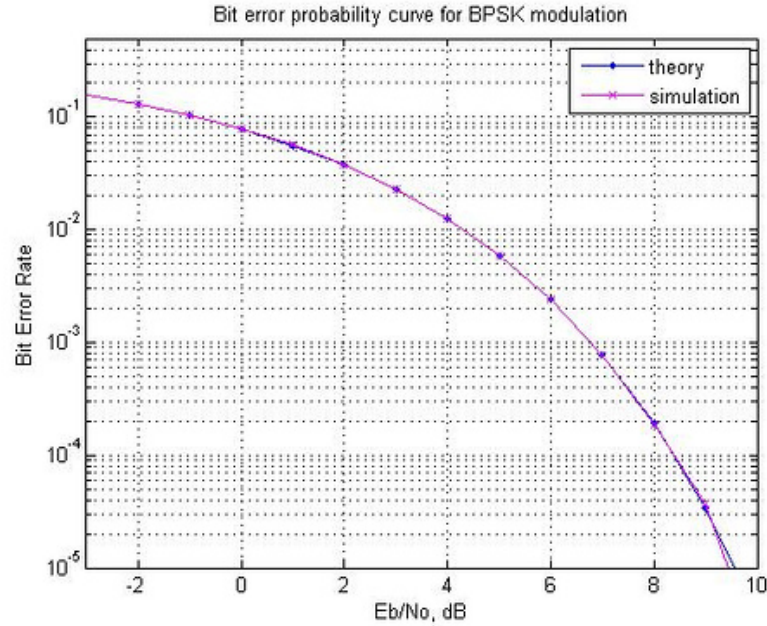


Figure 9-1: The BER of the simple binary signaling scheme in terms of the erfc function. The chart shows the theoretical prediction and simulation results. Picture courtesy: www.dsplog.com

communication scheme.

The greater the SNR, the lower the BER, and vice versa. Equation (9.9) tells us how the two quantities relate to one another for our case, and is plotted in Figure 9-1. The shape of this curve is characteristic of the BER v. SNR curves for many signaling and channel coding schemes, as we will see in the next few chapters. More complicated signaling schemes will have different BER-SNR relationships, but the BER will almost always be a function of the SNR.

■ 9.4 BER: The Case of Unequal Priors

When the prior probability of the sender transmitting a “0” is the same as a “1”, the optimal digitizing threshold is indeed 0 volts, by symmetry, if a “0” is sent at $-\sqrt{E_s}$ and a “1” at $+\sqrt{E_s}$ volts. But what happens when a “0” is more likely than a “1”, or vice versa?

If the threshold remains at 0 volts, then the probability of a bit error is the *same* as Equation (9.9). To see why, suppose the prior probability of a “0” is p_0 and a “1” is $p_1 = 1 - p_0$. Then, the probability of bit error can be simplified using a calculation similar to the previous section to give us

$$\mathbb{P}(\text{error}) = \frac{p_0}{2} \operatorname{erfc}(\sqrt{E_s/N_0}) + \frac{p_1}{2} \operatorname{erfc}(\sqrt{E_s/N_0}) = \frac{1}{2} \operatorname{erfc}(\sqrt{E_s/N_0}). \quad (9.10)$$

This equation is the same as Equation (9.9). It should make intuitive sense: when the threshold is 0 volts, the channel has the property that the probability of a “0” becoming a “1” is the same as the opposite flip. The probability of a “0” flipping depends *only* on the threshold used and the signal-to-noise ratio, and not on p_0 in this case.

Note, however, that when $p_0 \neq p_1 \neq 1/2$, the optimal digitizing threshold is not 0 (or, in general, not half-way between the voltage levels used for a “0” and a “1”). Intuitively, if zeroes are more likely than ones, the threshold should actually be *greater* than 0 volts, because the odds of any given bit being 0 are higher, so one might want to “guess” that a bit is a “0” even if the voltage level were a little larger than 0 volts. Similarly, if the prior probability of a “1” were larger, then the optimal threshold will be *lower* than 0 volts.

So what *is* the optimal digitizing threshold, assuming the receiver uses a single threshold to decide on a bit? Let’s assume it is V_t , then write an equation for the error probability (BER) in terms of V_t , differentiate it with respect to V_t , set the derivative to 0, and determine the optimal value. One can then also verify the sign of the second derivative to establish that the optimal value is indeed a minimum.

Fortunately, this calculation is not that difficult or cumbersome, because V_t will show up in the limit of the integration, so differentiation is straightforward. We will use the property that

$$\frac{d}{dz} \operatorname{erfc}(z) = \frac{d}{dz} \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-v^2} dv = -\frac{2}{\sqrt{\pi}} e^{-z^2}. \quad (9.11)$$

The equation for the BER is a direct extension of what we wrote earlier in Equation (9.10) to the case where we use a threshold V_t instead of 0 volts:

$$\mathbb{P}(\text{error}) = \frac{p_0}{2} \operatorname{erfc}\left(\frac{V_t + \sqrt{E_s}}{\sqrt{N_0}}\right) + \frac{p_1}{2} \operatorname{erfc}\left(\frac{\sqrt{E_s} - V_t}{\sqrt{N_0}}\right). \quad (9.12)$$

Using Equation (9.11) to differentiate the RHS of Equation (9.12) and setting it to 0, we get the following equation for V_t :

$$-p_0 e^{-(V_t + \sqrt{E_s})^2/N_0} + p_1 e^{-(V_t - \sqrt{E_s})^2/N_0} = 0. \quad (9.13)$$

Solving Equation (9.13) gives us

$$V_t = \frac{N_0}{4\sqrt{E_s}} \cdot \log_e \frac{p_0}{p_1}. \quad (9.14)$$

It is straightforward to verify by taking the second derivative that this value of V_t does indeed minimize the BER.

One can sanity check a few cases of Equation (9.14). When $p_0 = p_1$, we know the answer is 0, and we get that from this equation. When p_0 increases, we know that the threshold should shift toward the positive side, and vice versa, both of which follow from the equation. Also, when the noise power N_0 increases, we expect the receiver to pay less attention to the received measurement and more attention to the prior (because there is more uncertainty in any received sample), and the expression for the threshold does indeed accomplish that, by moving the threshold further away from the origin and towards the side associated with the less likely bit.

Note that Equation (9.14) is for the case when a “0” and “1” are sent at voltages symmetric about 0. If one had a system where different voltages were used, say V_0 and V_1 , then the threshold calculation would have to be done in analogous fashion. In this case, the optimal value would be offset from the mid-point, $(V_0 + V_1)/2$.

$10 \log_{10} \alpha$	α
100	10000000000
90	1000000000
80	100000000
70	10000000
60	1000000
50	100000
40	10000
30	1000
20	100
10	10
0	1
-10	0.1
-20	0.01
-30	0.001
-40	0.0001
-50	0.000001
-60	0.0000001
-70	0.00000001
-80	0.000000001
-90	0.0000000001
-100	0.00000000001

Figure 9-2: The dB scale is a convenient log scale; α is the absolute ratio between two energy or power quantities in this table.

■ 9.5 Understanding SNR

The SNR of a communication link is important because it determines the bit error rate; later, we will find that an appropriate SNR also determines the capacity of the channel (the maximum possible rate at which communication can occur reliably). Because of the wide range of energy and power values observed over any communication channel (and also in other domains), it is convenient to represent such quantities on a *log scale*. When measured as the ratio of two energy or power quantities, the SNR is defined on a **decibel scale** according to the following formula.

Let α denote the ratio of two energy or power quantities, such as the energy per sample, E_s , and the noise power, $N_0 = 2\sigma^2$. Then, we say that the decibel separation corresponding to this ratio α is

$$\text{SNR}_{\text{dB}} = 10 \cdot \log_{10} \alpha. \quad (9.15)$$

Figure 9-2 shows some values on the dB scale. A convenient rule to remember is that **3 dB is equivalent to a ratio of about 2**, because $\log_{10} 2 = 0.301$.

The decibel (or dB) scale is widely used in many applications. It goes between $-\infty$ and ∞ , and succinctly captures ratios that may be multiple powers of ten apart. The online

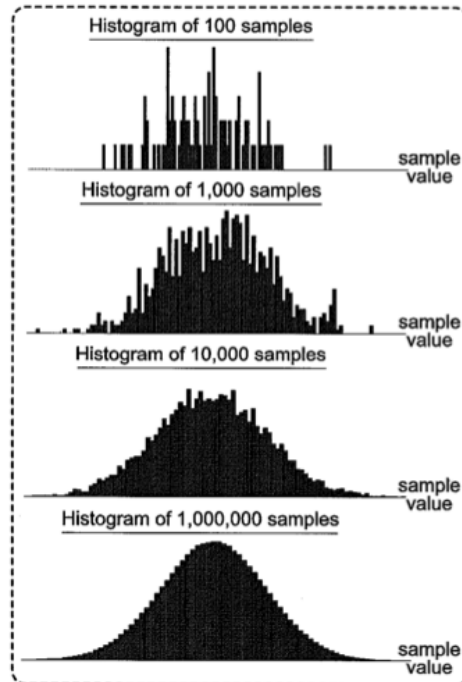


Figure 9-3: Histograms become smoother and more continuous when they are made from an increasing number of samples. In the limit when the number of samples is infinite, the resulting curve is a probability density function.

problem set has some simple calculations to help you get comfortable with the dB scale.

■ 9.6 Appendix: A Brief Recap of Continuous Random Variables

To understand what a PDF is, let us imagine that we generate 100 or 1000 independent noise samples and plot each one on a histogram. We might see pictures that look like the ones shown in Figure 9-3 (the top two pictures), where the horizontal axis is the value of the noise sample (binned) and the vertical axis is the frequency with which values showed up in each noise bin. As we increase the number of noise samples, we might see pictures as in the middle and bottom of Figure 9-3. The histogram is increasingly well approximated by a continuous curve. Considering the asymptotic behavior as the number of noise samples becomes very large leads to the notion of a **probability density function (PDF)**.

Formally, let X be the random variable of interest, and suppose X can take on any value in $(-\infty, \infty)$. Then, if the PDF of the underlying random variable is the non-negative function $f_X(x) \geq 0$, it means that the probability the random variable X takes on a value between x and $x + dx$, where dx is a small increment around x , is $f_X(x) dx$. More generally, the probability that a random variable X lies in the range $(x_1, x_2]$ is given by

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx . \quad (9.16)$$

An example of a PDF $f_X(x)$ is shown in Figure 9-5.

The PDF is by itself *not* a probability; the *area* under any portion of it is a probability. Though $f_X(x)$ itself may exceed 1, the area under any part of it is a probability, and can never exceed 1. Also, the PDF is *normalized* to reflect the fact that the probability X takes *some* value is always 1, so

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 .$$

Mean The mean μ_X of a random variable X can be computed from its PDF as follows:

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx . \quad (9.17)$$

If you think of the PDF as representing a “probability mass” distribution on the real axis, then the mean is the location of its center of mass; pivoting the real axis at this point will allow the mass distribution to remain in balance, without tipping to one side or the other.

The *law of large numbers* states that if $x[k]$ is an iid random process with the underlying PDF at each time being $f_X(x)$, then the sample mean converges to the the mean μ_X as the number of samples approaches ∞ :

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=0}^{K-1} x[k] = \mu_X . \quad (9.18)$$

Variance The variance is a measure of spread around the mean, and is defined by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx . \quad (9.19)$$

(To continue the mass analogy, the variance is analogous to the moment of inertia of the probability mass. Probability mass that is further away from the center of mass on either side, i.e., further away from the mean, contributes significantly more to the variance than mass close to the mean.) Again, under appropriate conditions, the sample variance for an iid process $x[k]$ converges to the variance. The *standard deviation* is defined as the square root of the variance, namely σ_X .

Cumulative distribution function The integral of the PDF from $-\infty$ to x ,

$$F_X(x) = \int_{-\infty}^x f_X(\alpha) d\alpha ,$$

is called the **cumulative distribution function (CDF)**, because it represents the cumulative probability that the random variable takes on a value $\leq x$. The CDF increases monotonically (or, more precisely, is monotonically non-decreasing) from 0 when $x = -\infty$ to 1 when $x = \infty$.

Example: Uniform distribution This simple example may help illustrate the idea of a PDF better, especially for those who haven’t see this notion before. Suppose that a random

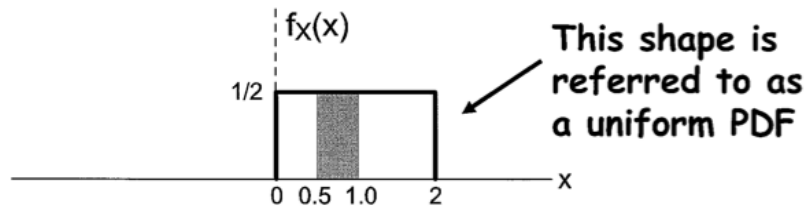


Figure 9-4: PDF of a uniform distribution.

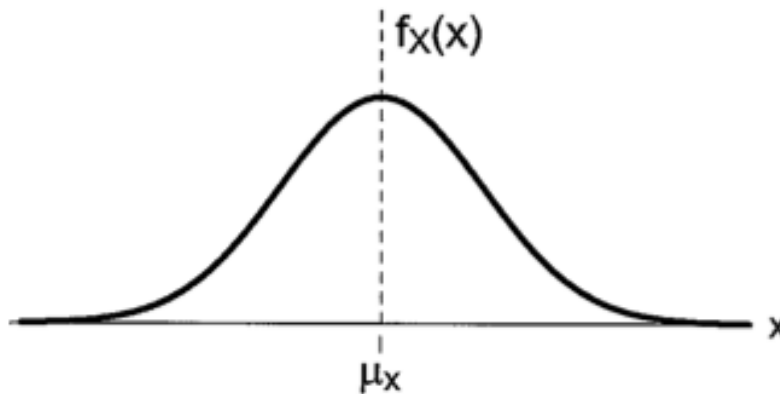


Figure 9-5: PDF of a Gaussian distribution, aka a “bell curve”.

variable X can take on any value between 0 and 2 with equal probability, and always lies in that range. What is the corresponding PDF?

Because the probability of X being in the range $(x, x + dx)$ is independent of x as long as x is in $[0, 2]$, it must be the case that the PDF $f_X(x)$ is some constant, h , for $x \in [0, 2]$. Moreover, it must be 0 for any x outside this range. We need to determine h . To do so, observe that the PDF must be normalized, so

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^2 h dx = 1, \quad (9.20)$$

which implies that $h = 0.5$. Hence, $f_X(x) = 0.5$ when $0 \leq x \leq 2$ and 0 otherwise. Figure 9-4 shows this *uniform PDF*.

One can easily calculate the probability that an x chosen from this distribution lies in the range $(0.3, 0.7)$. It is equal to $\int_{0.3}^{0.7} (0.5) dx = 0.2$.

A uniform PDF also provides a simple example that shows how the PDF, $f_X(x)$, could easily exceed 1. A uniform distribution whose values are always between 0 and δ , for some $\delta < 1$, has $f_X(x) = 1/\delta$, which is always larger than 1. To reiterate a point made before: the PDF $f_X(x)$ is *not a probability*, it is a probability *density*, and as such, could take on any non-negative value. The only constraint on it is that the total area under its curve (the integral over the possible values it can take) is 1.

As an exercise, you might try to determine the PDF, mean and variance of a random variable that is uniformly distributed in the arbitrary (but finite-length) interval $[a, b]$.

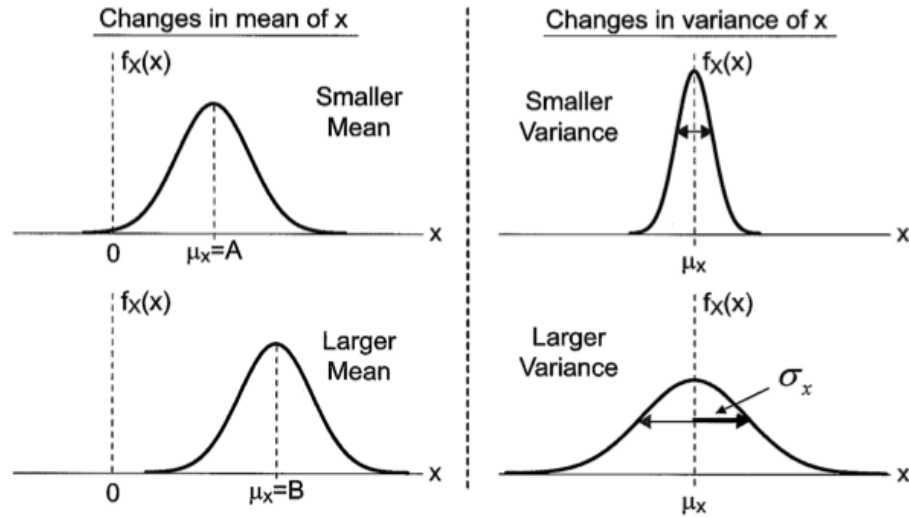


Figure 9-6: Changing the mean of a Gaussian distribution merely shifts the center of mass of the distribution because it just shifts the location of the peak. Changing the variance widens the curve.

Example: Gaussian distribution The PDF for a Gaussian random variable X is given by

$$f_W(w) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/(2\sigma_X^2)}. \quad (9.21)$$

This equation is plotted in Figure 9-5, which makes evident why a Gaussian distribution is colloquially referred to as a “bell curve”. The curve tapers off to 0 rapidly because of the e^{-x^2} dependence. The form of the expression makes clear that the PDF is symmetric about the value μ_X , which suffices to deduce that this parameter is indeed the mean of the distribution. It is an exercise in calculus (which we leave you to carry out, if you are sufficiently interested in the details) to verify that the area under the PDF is indeed 1 (as it has to be, for any PDF), and that the variance is in fact the parameter labeled as σ_X^2 in the above expression. Thus the Gaussian PDF is completely characterized by the mean and the variance of the distribution.

Changing the mean simply shifts the distribution to the left or right on the horizontal axis, as shown in the pictures on the left of Figure 9-6. Increasing the variance is more interesting from a physical standpoint; it widens (or fattens) the distribution and makes it more likely for values further from the mean to be selected, compared to a Gaussian with a smaller variance. A Gaussian random variable with a wider distribution (i.e., a larger variance) has more “power” compared to a narrower one.

■ Acknowledgments

Thanks to Bethany LaPenta and Kerry Xing for spotting various errors.