Lecture 3: Demodulation with Noise

Scribe: 鄧傑方、黃安 Lecture Date: 3/15, 2017 Lecturer: I-Hsiang Wang

Outline

- Random Noise
- Basic Principles of Detection
  - M-ary detection: MAP, MLE
  - Minimum distance (MD) detection for i.i.d. additive Gaussian noise
  - Binary Detection in Multi-Dimensional Space

Recap In the last lecture we introduced the basic architecture of digital modulation/demodulation when there is no noise. The main focus was on how to convert between discrete symbols and waveforms. For most digital communication systems, it boils down to selecting the right pulse shaper. One standard way is to choose the pulse shaper $p(t)$ and the corresponding low pass filter $q(t)$ such that $|\hat{p}(f)|^2$ satisfies the Nyquist Criterion (Bandedge Symmerty) and $\hat{p}(f) = \hat{q}^*(f)$. When $p(t)$ is real, $\hat{q}(f) = \hat{p}^*(f) = \hat{p}(-f)$ and hence $q(t) = p(-t)$ in time domain.

As for the mapping from bits to symbols and the design of constellation set, we simply introduced several standard ways. It turns out that the design is crucial for the performance of detection when noise is present.

Overview of this lecture In this lecture we introduce basic principles of detection, which is making a decision from observations drawn from a probabilistic model (statistical decision theory). In the communication problem, the randomness comes from noise, and the decision making process is called detection. First we introduce the concept of random noise. Then, we derive the so-called Bayesian optimal detector that minimizes the average probability of error, called the maximum a posterior (MAP) detector. Then we specialize it to the case when prior distribution is uniform and obtain the maximum likelihood (ML) detector. Finally under additive Gaussian noise, we show that ML is equivalent to minimum distance (MD) detector.

1 Random Noise

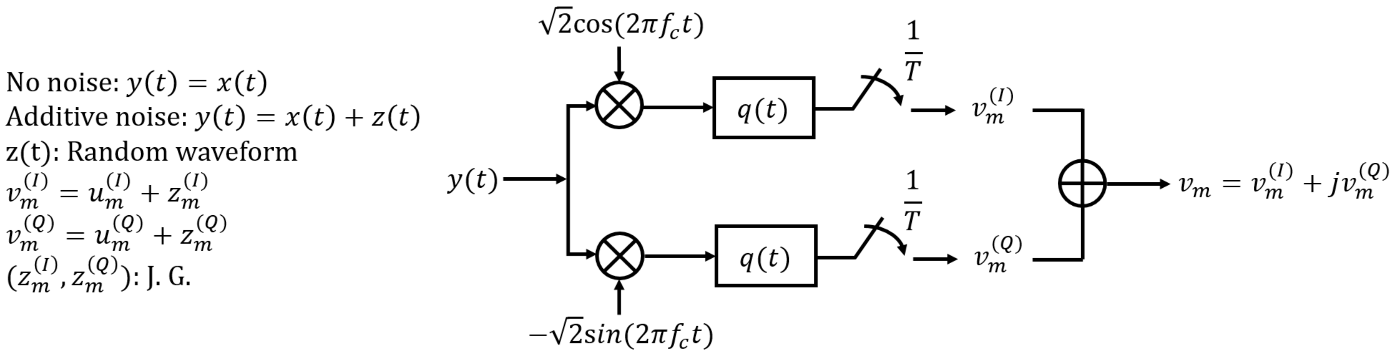

Figure 1: Basic Diagram of Received Signal

The random perturbation comes mainly from the continuous-time domain, and we model it can a *random waveform*, which can be viewed as a *continuous-time random process*.

Definition 1 (Random Process)**.** *A continuous-time random process $\{Z(t) : t \in \mathbb{R}\}$ is an uncountable collection of jointly distributed random variables, indexed by $t \in \mathbb{R}$, whose distribution is characterized by the joint distribution of $\{Z(t) : t \in S\}$ for all finite subset $S \subseteq \mathbb{R}$.*

A discrete-time random process is defined similarly except that the index lies in an infinitely countable set, that is, integer set $\mathbb{Z}$.

In other words, whenever we sample a random process $\{Z(t)\}$ at any finite number of time instances, say $t_1, t_2, \dots, t_n$, we obtain a collection of jointly distributed random variables $Z(t_1), Z(t_2), \dots, Z(t_n)$. In order to specify the "probability distribution" of this random process, one needs to specify $P_{Z(t_1)}\ \forall t_1 \in \mathbb{R}$, $P_{Z(t_1)Z(t_2)}\ \forall t_1, t_2 \in \mathbb{R}$, ..., etc..

Random process is quite general, while for the application in communication system design, the noise waveform usually has some additional properties. Below we introduce two such properties: *stationarity* and *Gaussianity*.

1.1 Gaussian Process

A particular useful random process that we use to model the random noise effect in communication systems is the *Gaussian process*, defined below.

Definition 2 (Gaussian Process)**.** *A random process $\{Z(t) : t \in \mathbb{R}\}$ is Gaussian if for all $n \in \mathbb{N} = \{1, 2, 3, \dots\}$ and any $\{t_1, t_2, \dots, t_n\} \subseteq \mathbb{R}$, $(Z(t_1), Z(t_2), \dots, Z(t_n))$ is a multivariate Gaussian random vector (jointly Gaussian).*

You can look up the probability density function of a multivariate Gaussian distribution, while the following definition of joint Gaussianity turns out to be more useful later.

Definition 3 (Joint Gaussianity)**.** $(Z_1, Z_2, \dots, Z_n)$ *are* jointly Gaussian *if and only if there exists $m \le n$ i.i.d. standard normal $W_1, W_2, ..., W_m$, constant vector $\underline{B} \in \mathbb{R}^n$, and matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ such that*

$$\underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = \mathbf{A}\underline{W} + \underline{B}, \quad \textit{where } \underline{W} \triangleq \begin{bmatrix} W_1 \\ \vdots \\ W_m \end{bmatrix}. \tag{1}$$

Remark 我們使用 Gaussian distribution 表示雜訊分布：干擾、熱擾動等各種獨立的雜訊疊加在一起，根據中央極限定理，總和的效應大略呈 Gaussian distribution。因此，Gaussian process 是通訊系統中最常用的噪音機率模型。

A jointly Gaussian random vector $\underline{Z}$ follows a multivariate Gaussian distribution, and the probability distribution is completely determined by the first and the second moment:

$$\begin{cases} \text{First Moment (Mean)} & \underline{m} \triangleq \mathbb{E}[\underline{Z}] = \begin{bmatrix} \mathbb{E}[Z_1] \\ \vdots \\ \mathbb{E}[Z_n] \end{bmatrix} \\ \text{Second Moment (Covariance Matrix)} & \mathbf{K} \triangleq \mathbb{E}[(\underline{Z}-\underline{m})(\underline{Z}-\underline{m})^\intercal] = \begin{bmatrix} \mathrm{Var}[Z_1] & \mathrm{Cov}(Z_1, Z_2) & \cdots & \mathrm{Cov}(Z_1, Z_n) \\ \mathrm{Cov}(Z_2, Z_1) & \mathrm{Var}[Z_1] & \cdots & \mathrm{Cov}(Z_2, Z_n) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(Z_n, Z_1) & \cdots & & \mathrm{Var}[Z_n] \end{bmatrix} \end{cases}$$

Hence, for a Gaussian process, its probability distribution is completely characterized by the following two functions.

- **Mean function**: $m_Z(t) \triangleq \mathbb{E}[Z(t)]$
- **Autocovariance function**: $K_Z(t_1, t_2) \triangleq \mathrm{Cov}(Z(t_1), Z(t_2))$.
  Or, **Autocorrelation function**: $R_Z(t_1, t_2) \triangleq \mathbb{E}[Z(t_1)Z(t_2)] = K_Z(t_1, t_2) + m_Z(t_1)m_Z(t_2)$.

1.2 Stationary Random Process

Stationarity refers to the time-invariance of probability distribution of a random process. Roughly speaking, no matter when you observe the random waveform, its probabilistic characteristics remain the same. The following definition makes this argument rigorous.

Definition 4 (Stationary Process)**.** *A random process $\{Z(t)\}$ is stationary if and only if for any finite sampling $\{t_1, t_2, \dots, t_n\}$ and time shift $\tau \in \mathbb{R}$,*

$$P_{Z(t_1), Z(t_2), \dots, Z(t_n)} \equiv P_{Z(t_1+\tau), Z(t_2+\tau), \dots, Z(t_n+\tau)}.$$

The above is also called *strict-sense stationary* in the literature. On the other hand, for a Gaussian process, since the distribution is completely characterized by the first and the second moments, for it to be stationary it suffices to satisfy the following *wide-sense stationary* property:

Definition 5 (Wide-Sense Stationarity). *A random process $\{Z(t)\}$ is* wide-sense stationary *(WSS) if and only if the following two hold:*

- *Mean function is fixed: $m_Z(t) = m$ for all t.*
- *Autocorrelation function is only a function of time difference: $R_Z(t, t+\tau) = R_Z(\tau)$ for all t and τ.*

Hence, a Gaussian process is stationary if and only if it is WSS.

1.3 Filtering and Sampling a Random Process

As illustrated in Figure 1, the noise process will be filtered and sampled prior to reaching the detector. Hence, the random waveform should also be converted into two streams of noisy symbols $\{Z_m^{(I)}\}$ and $\{Z_m^{(Q)}\}$, which correspond to two *discrete-time* random processes.

We will introduce a useful noise model for $\{Z(t)\}$, called *White Gaussian Noise* (WGN) process. It turns out that when the noise random waveform $\{Z(t)\}$ is a WGN process, the two streams of discrete-time random processes$\{Z_m^{(I)}\}$ and $\{Z_m^{(Q)}\}$ are i.i.d. Gaussian random variables. This gives the basic probabilistic framework of detecting transmitted symbols from noisy observations.

We first discuss about the spectral property of a random process, before and after filtering. To talk about the "spectral" (frequency-domain) properties of a random process, the following definition is useful.

Definition 6 (Power Spectral Density). *For a WSS random process $\{Z(t)\}$ with autocorrelation function $R_Z(\tau)$, its* power spectral density (PSD) *is defined as the Fourier transform of $R_Z(\tau)$:*

$$S_Z(f) \triangleq \int_{-\infty}^{\infty} R_Z(\tau) e^{-j2\pi f\tau}\, d\tau. \tag{2}$$

The following useful property about the PSD of a LTI-filtered random process is particularly useful (proof can be found in standard textbooks on statistical signal processing/digital communications).

Claim 7. *For a WSS random process $\{X(t)\}$ with PSD $S_X(f)$ and a linear time invariant (LTI) filter with impulse response $h(\tau)$ and frequency response $\hat{h}(f)$, let $Y(t) = (X * h)(t)$. Then, $\{Y(t)\}$ is also WSS and its PSD is*

$$S_Y(f) = |\hat{h}(f)|^2 S_X(f). \tag{3}$$

The white noise model is a useful model because it can simulate WSS processes with *any* PSD by passing a WGN through a LTI filter to create the target PSD.

Definition 8 (White Gaussian Noise). *A WSS Gaussian process is called* white *if and only if it is zero mean with autocorrelation function $R_Z(\tau) = \frac{N_0}{2}\delta(\tau)$, that is, its PSD $S_Z(f) = \frac{N_0}{2}$.*

In words, in the PSD of a WGN, all frequency components have the same amount of contribution in terms of energy, and hence it is called "white".

The following fact is given without proof (can be found in standard digital communication textbook).

Claim 9. *If $\{Z(t)\}$ is a WGN with PSD $N_0/2$, then both $\{Z_m^{(I)}\}$ and $\{Z_m^{(Q)}\}$ are i.i.d. Gaussian with zero mean and variance $N_0/2$. Furthermore, $\{Z_m^{(I)}\}$ and $\{Z_m^{(Q)}\}$ are independent.*

2 Basic Principles of Detection

As shown in Figure 1, the goal of detection is to determine u_m from $V_m = u_m + Z_m$ for all m, where each u_m takes values in a constellation set $\mathcal{A}$ and $\{Z_m\}$ are i.i.d Gaussian.

Remark u, V, Z above can either be thought of as complex scalars or real 2-dimensional vectors. Below we will begin with detection in real scalars, but eventually this will be extended to detection in multi-dimensional space.

Figure 2: Block diagram of BPSK/Binary PAM

2.1 Motivating Example: Binar Detection in Scalar

Problem Setup

- Model (BPSK/Binary PAM): a bit $c \in \{0,1\}$ is mapped to $x \in \{\pm A\}$ (say $0 \mapsto -A$, $1 \mapsto +A$).
  The received signal
  $$Y = x + Z, \quad x \in \{\pm A\},\ Z \sim \mathcal{N}(0, \sigma^2) \tag{4}$$
- Thinking of the bit c being random, there is a *prior* distribution $P_C(\cdot)$ for the random C, where $P_C(0) = \Pr\{C = 0\}$ and $P_C(1) = \Pr\{C = 1\}$.
- The statistical model relating the observation Y and the hidden bit C is determined by the conditional distribution (density) $f_{Y|C}(y|c)$.
- Goal: Given the prior distribution P_C and the statistical model $f_{Y|C}$, we would like to find a rule, $\hat{c} : \mathbb{R} \to \{0,1\}$, such that $P_e \triangleq \Pr\{C \neq \hat{c}(Y)\}$ is minimized.

Analysis of Probability of Error P_e Let's take a deeper look into P_e and figure out how to find $\hat{c}$ that minimizes it.

1. First, note that $\hat{c}$ is a mapping from $\mathbb{R}$ to $\{0,1\}$ and is equivalent to *decision regions* D_0 and D_1 :
   - $D_0 \triangleq \{y \in \mathbb{R} : \hat{c}(y) = 0\}$
   - $D_1 \triangleq \{y \in \mathbb{R} : \hat{c}(y) = 1\}$

   Finding the best $\hat{c}$ is equivalent to finding the best partition D_0, D_1 of $\mathbb{R}$.
2. Probability of success:
   $$\begin{aligned}\Pr\{C = \hat{c}(Y)\} &= \sum_{i=0,1} \Pr\{c = i \text{ and } \hat{c}(Y) = i\} \\ &= \sum_{i=0,1} \Pr\{c = i | Y \in D_i\} \Pr\{Y \in D_i\} = \sum_{i=0,1} \int_{D_i} \boxed{P_{C|Y}(i|y)} f_Y(y)\, dy.\end{aligned} \tag{5}$$
3. Question: Given y, in order to maximize (5), should it belong to D_0 or D_1?
   Answer: If $P_{C|Y}(0|y) \geq P_{C|Y}(1|y)$, it should belong to D_0. Otherwise, it should belong to D_1.
   Key quantity: $P_{C|Y}(i|y)$. This is called *posterior probability* (事後機率)

Optimal Detection Rule From the above discussion, we come up with an optimal decision rule that minimizes P_e:

$$\hat{c}_{\text{MAP}}(y) = \arg\max_{i=0,1} P_{C|Y}(i|y). \tag{6}$$

This is called the maximum a posterior (MAP) detector.

2.2 Optimal Detection

The above principle easily extends to M-ary detection in n-dimensional space.

Theorem 10 (Optimal Bayesian Detection Rule: MAP). *For a M-ary detection problem where*

- *Hidden $C \in \mathcal{M}$ follows prior distribution P_C, where $|\mathcal{M}| = M = 2^\ell$*
- *Input $\underline{X} \in \mathcal{A}$ is determined by C in a one-to-one manner, $\mathcal{A} = \{\underline{a}_1, \ldots, \underline{a}_M\}$.*

- *Statistical model $f_{\underline{Y}|\underline{X}}$ which directly gives $f_{\underline{Y}|C}$.*

The optimal rule $\hat{c}: \mathbb{R}^n \to \mathcal{M}$ that minimizes probability of error $P_e = \Pr\{C \neq \hat{c}(\underline{Y})\}$ is given by the following maximum a posterior (MAP) detection rule

$$\hat{c}_{\text{MAP}}(\underline{y}) = \arg\max_{i\in\mathcal{M}} P_{C|\underline{Y}}(i|\underline{y}). \tag{7}$$

事前 vs. 事後機率 The posterior probability $P_{C|\underline{Y}}(c|\underline{y})$ can be viewed as the probability of C *after* observing $\underline{Y} = \underline{y}$. In contrast, the prior probability $P_C(c)$ is the probability of C *before* observing $\underline{Y} = \underline{y}$.

Equivalent Forms of MAP Posterior probability is not that easy to compute. However, notice that

$$P_{C|Y}(i|y) = \frac{f_{Y|C}(y|i)P_C(i)}{f_Y(y)} \propto f_{Y|C}(y|i)P_C(i) \quad \text{for fixed y.}$$

Therefore,

$$\hat{c}_{\text{MAP}}(\underline{y}) = \arg\max_{i\in\mathcal{M}} f_{\underline{Y}|C}(\underline{y}|i)P_C(i). \tag{8}$$

Moreover, if the prior is uniform, that is, $P_C(c) = 1/M$ for all $c \in \mathcal{M}$, then

$$\hat{c}_{\text{MAP}}(\underline{y}) \equiv \hat{c}_{\text{ML}}(\underline{y}) = \arg\max_{i\in\mathcal{M}} f_{\underline{Y}|C}(\underline{y}|i). \tag{9}$$

Short Summary

- MAP detection: $\hat{c}_{\text{MAP}}(\underline{y}) = \arg\max_{i\in\mathcal{M}} f_{\underline{Y}|C}(\underline{y}|i)P_C(i)$.
- ML detection: $\hat{c}_{\text{ML}}(\underline{y}) = \arg\max_{i\in\mathcal{M}} f_{\underline{Y}|C}(\underline{y}|i)$.
- MAP is the optimal detector
- ML $\equiv$ MAP when prior is uniform, i.e., $P_C(c) = \frac{1}{M} \quad \forall i \in \mathcal{M},\ M = |\mathcal{M}|$.

2.3 Specialization to the Additive Gaussian Noise Channel

The above principles can be applied to many problems other than communication. For our purpose, let's specialize the statistical model $f_{Y|X}$ to the additive Gaussian noise model $Y = X + Z$ where $Z \sim \mathcal{N}(0, \sigma^2)$, as follows:

$$C \in \mathcal{M} \longleftrightarrow \underline{X} \in \mathcal{A} = \{\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_M\}, \quad \mathcal{A} \subseteq \mathbb{R}^n : \text{constellation set}, \quad M = 2^l$$

$$\underline{Y} = \underline{X} + \underline{Z} \quad \underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \quad Z_1, Z_2, \ldots, Z_n\text{: iid } \mathcal{N}(0, \sigma^2)$$

$$\Longrightarrow \underline{Y}|C = i \ \sim \mathcal{N}(\underline{a}_i, \sigma^2 \mathbf{I}_n)$$

$$\Longrightarrow f_{\underline{Y}|C}(\underline{y}|i) = f_{\underline{Y}|i}(y|i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{k=1}^n \exp\left(\frac{-(y_k - a_{ik})^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\frac{-1}{2\sigma^2}\sum_{k=1}^n |y_k - a_{ik}|^2\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\frac{-1}{2\sigma^2}\|\underline{y} - \underline{a}_i\|^2\right)$$

$$\text{(If prior is uniform)} \quad \hat{c}_{\text{MAP}}(\underline{y}) = \hat{c}_{\text{ML}}(\underline{y}) = \arg\min_{i\in\mathcal{M}} \|\underline{y} - \underline{a}_i\| = \text{Minimum Distance (MD) Detection}$$

Hence, when the additive noises are i.i.d Gaussian, ML $\equiv$ MD. Based on this principle, the decision regions of constellation sets can be drawn, as shown in Figure 6.

2.4 Performance Analysis: Binary PAM

Let's begin with the simplest example: binary PAM and derive the performance of ML.

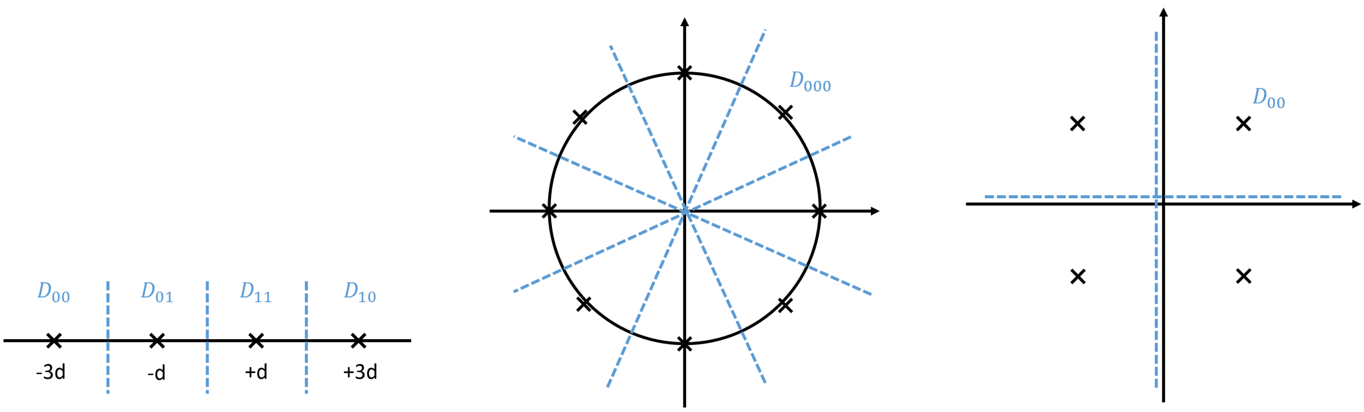

Figure 3: 4-PAM

Figure 4: 8-PSK

Figure 5: 4QAM

Figure 6: Examples of Decision Regions

Probability of Error Recall the model under binary PAM: $Y = X + Z$, $X \in \{\pm d\}$, $Z \sim \mathcal{N}(0, \sigma^2)$, uniform prior.

$$\begin{aligned} P_e &= \Pr\{\hat{c}(Y) \neq C\} \\ &= \Pr\{C = 0\} \Pr\{\hat{c}(Y) \neq 0 | C = 0\} + \Pr\{C = 1\} \Pr\{\hat{c}(Y) \neq 1 | C = 1\} \end{aligned}$$

$$\begin{aligned} \Pr\{\hat{c}(Y) \neq 0 | C = 0\} &= \Pr\{Y \in D_1 | C = 0\} \\ &= \Pr\{Y > 0 | C = 0\} \\ &= \Pr\{\mathcal{N}(-d, \sigma^2) > 0\} \\ &= \Pr\{\mathcal{N}(0, \sigma^2) > d\} \\ &= \Pr\left\{\mathcal{N}(0, 1) > \frac{d}{\sigma}\right\} = Q\left(\frac{d}{\sigma}\right) \end{aligned}$$

$$\Pr\{\hat{c}(Y) \neq 1 | C = 1\} = Q\left(\frac{d}{\sigma}\right) \quad \text{(by symmetry)}$$

Hence, the probability of error for binary-PAM $\{\pm d\}$ under Gaussian noise $\mathcal{N}(0, \sigma^2)$ is

$$P_e = Q\left(\frac{d}{\sigma}\right). \tag{10}$$

Definition 11 (Q Function)**.** *Q function denotes the tail probability (complementary CDF) of the standard normal:*

$$Q(x) \triangleq \Pr\{\mathcal{N}(0, 1) > x\} = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

Claim 12 (Bound on Q Function)**.**

$$Q(x) \leq \frac{1}{2} e^{-\frac{1}{2}x^2}, \ \forall x \geq 0.$$

In fact, $\ln Q(x) \asymp -\frac{1}{2}x^2$*, that is,*

$$\lim_{x \to \infty} \frac{\ln Q(x)}{-\frac{1}{2}x^2} = 1.$$

Signal-to-Noise Ratio

- Noise variance: σ^2
- Signal energy: $E_s = \frac{1}{2}|+d|^2 + \frac{1}{2}|-d|^2 = d^2$
- $\mathsf{SNR} \triangleq \frac{\text{signal energy}}{\text{noise variance}} = \frac{E_s}{\sigma^2} = \frac{d^2}{\sigma^2}$

Hence, we can conclude that $P_e = Q(\sqrt{\mathsf{SNR}})$ and $\ln P_e \asymp -\frac{1}{2}\mathsf{SNR}$. P_e increases when SNR decreases; P_e decreases when SNR increases.