



Elements of Communication Theory & Signals

Slot 3 – Probability, Optimal Receiver and BER

Germain PHAM

dpham@telecom-paris.fr

September 30, 2026



What Is a Communication System?



At signal level:

$$r(t) = s(t) + n(t)$$

(1)

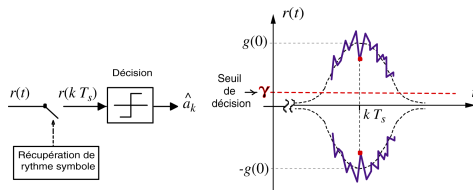
- $s(t)$: transmitted waveform;
- $n(t)$: unwanted disturbance or noise;
- $r(t)$: received waveform.

Today: we concentrate on $r(t)$.

Impairments : Channel Noise

Error probability (for binary symbols)

The receiver observes an effect, **not its cause**.



After Nyquist filtering and correct sampling, the considered model is:

$$r(kT_{\text{symp}}) = a_k g(0) + n_k. \quad (2)$$

a_k is the (true) transmitted symbol (it is *only known* to the transmitter); n_k is the noise at the decision input. Both random variables. The receiver only observes $r(kT_{\text{symp}})$, and has to *guess* what is the (unknown) transmitted symbol $\hat{a}_k$.

$$\hat{a}_k = \text{decision} \{r(kT_{\text{symp}})\} \quad (3)$$

Symbol probability error

$$P_{\text{symp}}(e) \triangleq P(\hat{a}_k \neq a_k) \quad (4)$$

Design question

Which decision minimizes errors?

ω denotes here a sample in the sense of the probability theory, n denotes the noise.

- random sequence of symbols $a_k(\omega)$ *independent and stationary with k* ,
- no ISI (*Nyquist channel*),
- additive noise $n(t, \omega) = n_t$ random stationary centered, independent of the signal, with probability density $f_N(n)$.
- no delay or sampling error (sampling happens at the right time, i.e. at the exact symbol instants)

Experiment: a procedure together with specified observations. Its outcome cannot be predicted with certainty.

Outcome ω : one possible result. The *sample space* Ω contains all distinguishable possible outcomes.

Event E : a set of outcomes. The event occurs exactly when $\omega \in E$.

Set Algebra	Probability
Subset	Event E
Universal set	Sample space Ω
Element	Outcome ω

Example: send and detect one bit.

Record the transmitted bit b and the receiver's binary decision $\hat{b}$:

$$\Omega = \{(0, \hat{0}), (0, \hat{1}), (1, \hat{0}), (1, \hat{1})\}. \quad (5)$$

The error event and the event "0 was sent" are

$$E = \{(0, \hat{1}), (1, \hat{0})\}, \quad (6)$$

$$H_0 = \{(0, \hat{0}), (0, \hat{1})\}, \quad (7)$$

$$H_1 = \{(1, \hat{0}), (1, \hat{1})\}. \quad (8)$$

An outcome such as $(0, \hat{1})$ is not the same object as the singleton event $\{(0, \hat{1})\}$.

Recording a *continuous voltage* instead of a binary decision defines a different observation space. Specify the experiment before assigning probabilities.

Union, intersection and complement

$$E \cup F \triangleq \{\omega : \omega \in E \text{ or } \omega \in F\}, \quad (9)$$

$$E \cap F \triangleq \{\omega : \omega \in E \text{ and } \omega \in F\}, \quad (10)$$

$$E^c \triangleq \Omega \setminus E. \quad (11)$$

Here “or” is inclusive. $\emptyset$ is the impossible event; Ω is the certain event.

Basic identities

$$(E \cup F)^c = E^c \cap F^c, \quad (12)$$

$$(E \cap F)^c = E^c \cup F^c, \quad (13)$$

$$E \cap (F \cup G) = (E \cap F) \cup (E \cap G). \quad (14)$$

Partition

A family $H_0, \dots, H_{M-1}$ is a **partition** of Ω when it is both

$$H_i \cap H_j = \emptyset \quad (i \neq j) \quad (\text{mutually exclusive}), \quad (15)$$

$$\bigcup_{i=0}^{M-1} H_i = \Omega \quad (\text{collectively exhaustive}). \quad (16)$$

For any event E ,

$$E = \bigcup_{i=0}^{M-1} (E \cap H_i), \quad (17)$$

and the events in this union are mutually exclusive.

Receiver example

every error occurs either when 0 was sent or when 1 was sent:

$$E = (E \cap H_0) \cup (E \cap H_1). \quad (18)$$

This decomposition will produce the total error probability.

A probability measure assigns to each event E a number $P(E) \in [0, 1]$, describing how likely it is to occur.

Classical Definition

For a finite sample space Ω with **equally likely outcomes**,

$$P(E) = \frac{|E|}{|\Omega|}, \quad (19)$$

where $|E|$ counts the outcomes belonging to E and $|\Omega|$ counts all possible outcomes.

Example: rolling a fair die

For the event “the result is even,”

$$A = \{2, 4, 6\}, \quad P(A) = \frac{3}{6} = \frac{1}{2}. \quad (20)$$

The probability is obtained from the model, without performing the experiment.

Relative Frequency Definition

Repeat the same experiment n times. Let n_E count the occurrences of E . The probability is the asymptotic relative frequency,

$$P(E) = \lim_{n \rightarrow \infty} \frac{n_E}{n}, \quad (21)$$

provided this limit exists.

Example: repeated die rolls

Suppose 1000 rolls produce 512 even results. The observed relative frequency is

$$\hat{P}(A) = \frac{512}{1000} = 0.512. \quad (22)$$

A finite measurement gives an **estimate**, not the exact probability.

Three probability axioms

1. Nonnegativity:

$$P(E) \geq 0. \quad (23)$$

2. Normalization:

$$P(\Omega) = 1. \quad (24)$$

3. Additivity for mutually exclusive events:

$$P\left(\bigcup_{i=1}^{M \rightarrow \infty} E_i\right) = \sum_{i=1}^{M \rightarrow \infty} P(E_i), \quad (25)$$

provided $E_i \cap E_j = \emptyset$ for $i \neq j$.

Interpretation: when events cannot occur together, their probabilities add.

Consequences (not additional axioms)

$$P(\emptyset) = 0, \quad (26)$$

$$0 \leq P(E) \leq 1, \quad (27)$$

$$P(E^c) = 1 - P(E), \quad (28)$$

$$E \subseteq F \Rightarrow P(E) \leq P(F). \quad (29)$$

For events that may overlap,

$$P(E \cup F) = P(E) + P(F) - P(E \cap F). \quad (30)$$

The intersection is subtracted because it was counted twice.

Conditional probability definition ♥

$$P(E | F) \triangleq \frac{P(E \cap F)}{P(F)} \quad \text{for } P(F) > 0 \quad (31)$$

Knowing that F occurred restricts the possible outcomes to F and renormalizes their probabilities.

$$P(E \cap F) = P(E | F)P(F) \quad (32)$$

$$= P(F | E)P(E). \quad (33)$$

The last form also requires $P(E) > 0$.

In general,

$$P(E | F) \neq P(F | E). \quad (34)$$

Conditioning is a statement about *information*, not necessarily about causation or temporal order.

A binary receiver's joint probability table

	$\hat{B} = 0$	$\hat{B} = 1$	Total
$B = 0$	0.72	0.08	0.80
$B = 1$	0.04	0.16	0.20
Total	0.76	0.24	1

$$P(\hat{B} = 1 | B = 0) = \frac{0.08}{0.80} = 0.10, \quad (35)$$

$$P(\hat{B} = 0 | B = 1) = \frac{0.04}{0.20} = 0.20, \quad (36)$$

$$P(B = 0 | \hat{B} = 1) = \frac{0.08}{0.24} = 0.33. \quad (37)$$

The first and second describe a channel/receiver error mechanism. The third infers the emitted bit from the observation.

Total probability and Bayes' theorem

Let $\{H_0, \dots, H_{M-1}\}$ partition Ω , with priors $P(H_i) > 0$.

Total probability theorem ♥

$$P(E) = \sum_i P(E \cap H_i) \quad (38)$$

$$= \sum_i P(E | H_i)P(H_i). \quad (39)$$

For a binary receiver,

$$P(E) = P(E | H_0)P(H_0) \quad (40)$$

$$+ P(E | H_1)P(H_1). \quad (41)$$

Example: Using the previous table

$$\begin{aligned} P(E) &= P(\hat{B} = 1 | B = 0)P(B = 0) \\ &+ P(\hat{B} = 0 | B = 1)P(B = 1) \\ &= 0.1 \times 0.8 + 0.2 \times 0.2 = 0.12. \end{aligned} \quad (42)$$

Bayes: infer a cause from an observation ♥

For an observed event F with $P(F) > 0$,

$$P(H_i | F) = \frac{P(F | H_i)P(H_i)}{\sum_j P(F | H_j)P(H_j)} \quad (43)$$

Bayes' theorem in words ♥

posterior = likelihood $\times$ prior / evidence

Example: if the receiver outputs 1

$$\begin{aligned} P(B = 1 | \hat{B} = 1) &= \frac{P(\hat{B}=1|B=1)P(B=1)}{P(\hat{B}=1|B=0)P(B=0)+P(\hat{B}=1|B=1)P(B=1)} \\ &= \frac{0.8 \times 0.2}{0.1 \times 0.8 + 0.8 \times 0.2} = 0.15. \end{aligned} \quad (44)$$

Inference uses both the quality of the observation and the prior probabilities.

Independence is not mutual exclusiveness

Events E and F are **independent** when

$$P(E \cap F) = P(E)P(F). \quad (45)$$

If $P(F) > 0$, this is equivalent to

$$P(E | F) = P(E). \quad (46)$$

Learning that F occurred does not change the probability of E .

By contrast, mutually exclusive events satisfy

$$E \cap F = \emptyset \Rightarrow P(E \cap F) = 0. \quad (47)$$

If both have positive probabilities, mutually exclusive events *cannot* be independent.

More than two events: Mutual independence requires

$$P\left(\bigcap_{i \in J} E_i\right) = \prod_{i \in J} P(E_i) \quad (48)$$

for every nonempty finite subcollection J . Pairwise independence alone is insufficient.

Communication model: Independent, identically distributed (iid) bits have a common distribution and a joint distribution which is the product of the individual distributions:

$$P(B_1 = b_1, \dots, B_n = b_n) = \prod_{k=1}^n P(B_k = b_k). \quad (49)$$

Independence of the *noise and data* is a separate assumption from independence *across time*.

Random variable definition

A real(-valued) **random variable** is a function

$$X : \Omega \longrightarrow \mathbb{R}, \quad \omega \longmapsto X(\omega). \quad (50)$$

It assigns a numerical value to each outcome.

Simple example

Bit and amplitude are different variables:

$$B \in \{0, 1\}, \quad A = a \cdot (2B - 1) \in \{-a, +a\}.$$

An event can be specified by a condition on X :

$$\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}. \quad (51)$$

Thus, $P(X = x)$ is the probability of all outcomes that produce the value x .

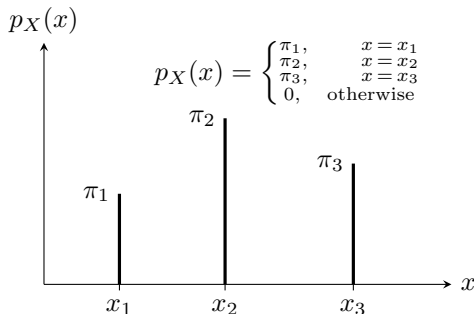
X denotes the random variable; x denotes a possible value or an observed realization.

Discrete case: the possible values form a finite or countably infinite set $\mathcal{X}$.

The **probability mass function** (PMF) is

$$p_X(x) = P(X = x) = \begin{cases} \pi_k, & x = x_k \\ 0, & \text{otherwise} \end{cases} \quad (52)$$

where x_k are the possible values of X and π_k are the corresponding probabilities.



Examples

- An error indicator $J = \mathbb{1}_{\{\hat{B} \neq B\}}$ is:

$$p_J(1) = P_b, \quad p_J(0) = 1 - P_b. \quad (53)$$

It converts an *error* event into a numerical variable (1 for error, 0 for no error). This probability model is also known as a *Bernoulli* distribution.

- even or odd die roll: We roll a six-sided die: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Define the random variable X by

$$X(\omega) = \begin{cases} 1, & \omega \in \{2, 4, 6\} \quad (\text{even}), \\ 0, & \omega \in \{1, 3, 5\} \quad (\text{odd}). \end{cases} \quad (54)$$

The condition $X = 1$ corresponds to the event “the result is even”:

$$\{X = 1\} = \{\omega \in \Omega : X(\omega) = 1\} = \{2, 4, 6\}. \quad (55)$$

Thus, $P(X = 1)$ is the probability of rolling an even number.

Cumulative Distribution Function (CDF)

$$F_X(x) \triangleq P(X(\omega) \leq x) = P(X \leq x). \quad (56)$$

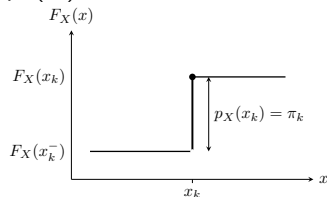
- It is nondecreasing, right-continuous, and satisfies

$$F_X(-\infty) = 0, \quad F_X(+\infty) = 1. \quad (57)$$

- For any $a < b$,

$$P(a < X \leq b) = F_X(b) - F_X(a). \quad (58)$$

For **discrete** X , the CDF has a jump of size $p_X(x_k) = \pi_k$ at x_k :



Probability density function (PDF)

For **continuous** X , f_X is the probability density function (PDF):

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}, \quad (59)$$

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad (60)$$

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x) dx = 1. \quad (61)$$

$$P(X \in D) = \int_D f_X(x) dx, \quad P(X = x) = 0. \quad (62)$$

- A PDF can exceed 1.
- If X is a voltage, f_X has units V^{-1} ; probabilities have no units.
- An event of probability zero need not be empty.
- At a continuity point of f_X , for small $\Delta > 0$,

$$P(x \leq X < x + \Delta) \approx f_X(x)\Delta.$$

Expectation, variance and the measured BER

For an integrable function $g(X)$,

$$\mathbb{E}[g(X)] \triangleq \begin{cases} \sum g(x)p_X(x), & X \text{ discrete,} \\ \int_{-\infty}^{\infty} g(x)f_X(x) dx, & X \text{ has a PDF.} \end{cases} \quad (63)$$

Fundamental cases:

Quantity	Expression
Expectation (mean)	$\mu_X \triangleq \mathbb{E}[X]$
Variance	$\sigma_X^2 = \text{Var}(X)$ $\triangleq \mathbb{E}[(X - \mu_X)^2]$ $= \mathbb{E}[X^2] - \mu_X^2$
Moment of order r	$m_r \triangleq \mathbb{E}[X^r]$

Properties:

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad (64)$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X). \quad (65)$$

For centered noise, its mean-square value equals its variance.

Example

With $J_k = \mathbb{1}_{\{\hat{B}_k \neq B_k\}}$,

$$\mathbb{E}[J_k] = P_b, \quad \text{Var}(J_k) = P_b(1 - P_b). \quad (66)$$

Find more examples in Appendix Slide: III.

Two random variables can describe the *same* experiment: for example, a transmitted symbol and a received measurement.

Discrete pair: joint PMF

$$p_{X,Y}(x, y) = P(X = x, Y = y), \quad (67)$$

$$\sum_x \sum_y p_{X,Y}(x, y) = 1. \quad (68)$$

Marginalization removes an unobserved variable:

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad (69)$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y). \quad (70)$$

The joint table on slide 9 therefore contains both the input and output distributions.

Continuous pair: joint PDF

$$P((X, Y) \in D) = \iint_D f_{X,Y}(x, y) dx dy. \quad (71)$$

For example,

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) \\ = \int_a^b \int_c^d f_{X,Y}(x, y) dy dx. \end{aligned} \quad (72)$$

The marginal PDFs are

$$f_X(x) \triangleq \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy, \quad (73)$$

$$f_Y(y) \triangleq \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx. \quad (74)$$

In an I/Q receiver, a decision region will be a subset of the *two-dimensional observation plane*.

Conditional density ♥

For jointly continuous X, Y and $f_Y(y) > 0$,

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \quad (75)$$

We do *not* divide $P(X = x, Y = y)$ by $P(Y = y)$: both can be zero.

Mixed random variable

For a discrete hypothesis H_i and continuous observation R , consider

$$f_{R|H_i}(r | \text{hypothesis } i) = f_{R|H_i}(r), \quad \pi_i = P(H_i). \quad (76)$$

For a measurable region D ,

$$P(H_i, R \in D) = \pi_i \int_D f_{R|H_i}(r) \mathrm{d}r. \quad (77)$$

This is a mixed discrete/continuous model, not a two-dimensional continuous PDF.

For fixed H_i , $f_{R|H_i}(r)$ is a density in r . For an observed r , the values $f_{R|H_i}(r)$ are **likelihoods** used to compare hypotheses. They are *not* posterior probabilities and need not sum to one over i .

Total probability and Bayes' rule in the mixed discrete/continuous case

Total probability gives the **mixture density**

$$f_R(r) = \sum_i \pi_i f_{R|H_i}(r). \quad (78)$$

Bayes' rule becomes, wherever $f_R(r) > 0$,

$$P(H_i | R = r) = \frac{\pi_i f_{R|H_i}(r)}{\sum_j \pi_j f_{R|H_j}(r)}. \quad (79)$$

See Appendix Slide: IV for a note on the derivation of (79).

Gaussian random variable

A Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\sigma > 0$, has density

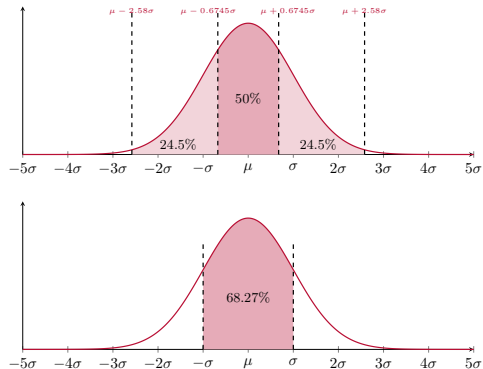
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]. \quad (80)$$

The center is μ ; σ controls the width.

Notation convention: the second parameter of $\mathcal{N}(\mu, \sigma^2)$ is the *variance*. Other sources may use the standard deviation instead, like in Yates–Goodman's book (Gaussian (μ, σ)).

Common values

$\mathbb{P}(\mu - \sigma \leq Y \leq \mu + \sigma)$	0.6827
$\mathbb{P}(\mu - 2\sigma \leq Y \leq \mu + 2\sigma)$	0.9545
$\mathbb{P}(\mu - 2.58\sigma \leq Y \leq \mu + 2.58\sigma)$	0.99
$\mathbb{P}(\mu - 3\sigma \leq Y \leq \mu + 3\sigma)$	0.9973



Standardization of Gaussian random variables

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\sigma > 0$.

Standard normal distribution

Standardization gives

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

The *standard* normal CDF is

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du.$$

General gaussian RV:

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (83)$$

Interval probabilities:

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (84)$$

Affine transformation:

$$cX + d \sim \mathcal{N}(c\mu + d, c^2\sigma^2). \quad (85)$$

For $c = 0$, the result is the deterministic value d .

The erf funct

The (Gauss) erf function

(81) The error function is defined by

$$\operatorname{erf}(z) \triangleq \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \quad (86)$$

(82)

It is related to the *standard* normal CDF by

$$\Phi(z) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \right]. \quad (87)$$

General gaussian RV:

$$F_X(x) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x - \mu}{\sqrt{2}\sigma}\right) \right]. \quad (88)$$

The Q-function: Gaussian tail probabilities

The Q-function ♥

For $Z \sim \mathcal{N}(0, 1)$, define

$$Q(z) = P(Z > z) \triangleq \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-u^2/2} du. \quad (89)$$

By symmetry,

$$Q(-z) = 1 - Q(z), \quad Q(0) = \frac{1}{2}. \quad (90)$$

For $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\sigma > 0$,

$$P(X > \gamma) = Q\left(\frac{\gamma - \mu}{\sigma}\right), \quad (91)$$

$$P(X < \gamma) = Q\left(\frac{\mu - \gamma}{\sigma}\right). \quad (92)$$

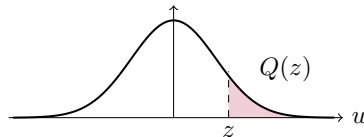
Interpretation: the argument is a signed distance between the mean and the decision boundary, measured in terms of *standard deviations*.

The Q-function is implemented as `qfunc` in Matlab/Octave.

Link with the error function

Since $Q(z) = 1 - \Phi(z)$,

$$Q(z) = \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \right]. \quad (93)$$



z	$Q(z)$
1	1.587×10^{-1}
2	2.275×10^{-2}
3	1.350×10^{-3}
4	3.167×10^{-5}

Tail probabilities change rapidly.

Example: noise-induced clipping at a receiver

At a given sampling instant, the received voltage is

$$R = A + N, \quad N \sim \mathcal{N}(0, \sigma^2), \quad (94)$$

where $A = 1 \text{ V}$ and $\sigma = 0.2 \text{ V}$.

The receiver clips input voltages above 1.4 V .

What is the probability of clipping?

$$P(R > 1.4 \text{ V}) = Q\left(\frac{1.4 - 1}{0.2}\right) \quad (95)$$

$$= Q(2) \approx 0.0228 \quad (96)$$

$$\approx 2.28\% \quad (97)$$

Hypothesis testing and optimal decisions

Waveform reception, matched filter and BER

A hypothesis test partitions the observation space

A **hypothesis** specifies a probability model for the observation. For binary transmission,

$$H_0 : \text{bit 0 sent}, \quad H_1 : \text{bit 1 sent}. \quad (98)$$

A deterministic decision rule maps each observation to a decoded bit:

$$\hat{B}(r) = \begin{cases} \hat{0}, & r \in D_0, \\ \hat{1}, & r \in D_1. \end{cases} \quad (99)$$

H_i describes the true transmitted state. D_i is a region of possible observations. They are not the same object.

Four possible outcomes (radar terminology)

	True H_0	True H_1
Decide H_0 $\hat{B}(r) = \hat{0}$	correct	miss (type II)
Decide H_1 $\hat{B}(r) = \hat{1}$	false alarm (type I)	correct

Recall the mixed RV slide (slide 17):

$$P_{\text{FA}}(D_1) = P(R \in D_1 \mid H_0) = \int_{D_1} f_0(r) dr, \quad (100)$$

$$P_{\text{MISS}}(D_0) = P(R \in D_0 \mid H_1) = \int_{D_0} f_1(r) dr \quad (101)$$

The probability of error is:

$$P_e = P(\text{error} \cap H_0) + P(\text{error} \cap H_1) \quad (102)$$

$$= \pi_0 P_{\text{FA}}(D_1) + \pi_1 P_{\text{MISS}}(D_0). \quad (103)$$

Design question: how should we choose D_0 and D_1 ? The answer depends on the criterion used to compare decision rules.

A decision rule is optimal for a chosen criterion

Our goal: choose a **decision rule** that assigns each observation r to D_0 or D_1 . What makes a rule *best* depends on the **criterion**.

Out of the scope of the lesson but, for more details, please refer to [1, 2].

The priors may be unknown

With known f_0, f_1 , we can compute a rule's conditional error probabilities P_{FA} and P_{MISS} .

$$P_e = \pi_0 P_{\text{FA}} + \pi_1 P_{\text{MISS}}.$$

Without $\pi_i = P(H_i)$, this quantity is **generally unknown**.

Example: Neyman–Pearson. Minimize P_{MISS} at a prescribed $P_{\text{FA}} = \alpha$.

Errors may have different costs

P_e counts all errors equally, not their consequences.

With **known priors and specified costs**, minimize the **expected cost** (Bayes risk):

$$\mathbb{E}[C] = \pi_0 C_{10} P_{\text{FA}} + \pi_1 C_{01} P_{\text{MISS}}.$$

C_{10} : cost of a false alarm.

C_{01} : cost of a miss.

Correct decisions cost zero.

Our focus in the next slide

Known priors; all errors count equally. Minimize $P_e \rightarrow$ **MAP rule**.

Minimum-error decision rule for a binary transmission

The value sampled at times multiple of T_s is doubly random:

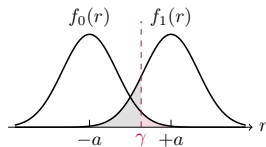
$$r(kT_s, \omega) = R_k = a_k(\omega) + n(kT_s, \omega) \quad (104)$$

■ random signal term : $a_k(\omega)$, with
 $a_k(\omega) \in \{-a, +a\}$

■ random noise term : $n(kT_s, \omega) = N_k$

$$\text{Given } H_0, R_k = -a + N_k \sim f_{N_k}(r + a) = f_0(r) \quad (105)$$

$$\text{Given } H_1, R_k = +a + N_k \sim f_{N_k}(r - a) = f_1(r) \quad (106)$$



Shaded areas: *conditional* error probability.

The minimum error probability criterion is also known as the MAP (maximum a posteriori) decision rule. See [1, 2] for more details.

The error likelihoods are:

$$P(e | H_0) = P(N \geq \gamma + a) \quad (107)$$

$$P(e | H_1) = P(N < \gamma - a). \quad (108)$$

Therefore, for any threshold γ ,

$$P_e(\gamma) = \pi_0 P(N \geq \gamma + a) + \pi_1 P(N < \gamma - a) \quad (109)$$

The minimum error probability is obtained when:

$$\frac{dP_e(\gamma)}{d\gamma} = 0 \text{ and } \frac{d^2P_e(\gamma)}{d\gamma^2} > 0. \quad (110)$$

We can easily show that:

$$\frac{dP_e(\gamma)}{d\gamma} = -\pi_0 f_N(\gamma + a) + \pi_1 f_N(a - \gamma) \quad (111)$$

Therefore, the minimum error probability occurs when:

$$\frac{f_N(a - \gamma^*)}{f_N(a + \gamma^*)} = \frac{\pi_0}{\pi_1} \quad (112)$$

Minimum-error decision rule for a binary transmission: Gaussian channel

When the noise is Gaussian $\mathcal{N}(0, \sigma^2)$, we have

$$P_e(\gamma) = \pi_0 Q\left(\frac{\gamma + a}{\sigma}\right) + \pi_1 Q\left(\frac{a - \gamma}{\sigma}\right) \quad (113)$$

and the optimal threshold γ^* satisfies

$$\frac{f_N(a - \gamma^*)}{f_N(a + \gamma^*)} = \exp\left(\frac{2\gamma^* a}{\sigma^2}\right) = \frac{\pi_0}{\pi_1} \quad (114)$$

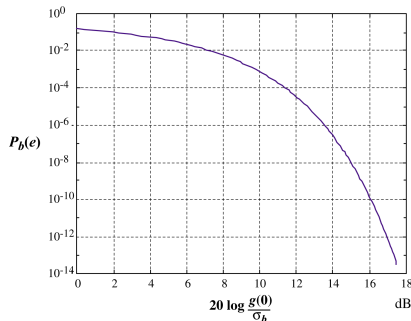
So the optimal threshold γ^* is:

$$\gamma^* = \frac{\sigma^2}{2a} \ln \frac{\pi_0}{\pi_1} \quad (115)$$

For equiprobable symbols $\gamma^* = 0$ and,

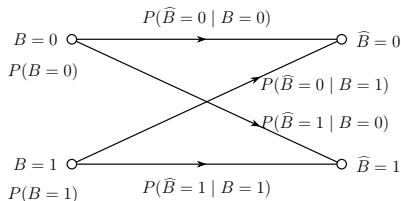
$$P_e \stackrel{\pi_0=\pi_1}{=} Q\left(\frac{a}{\sigma}\right). \quad (116)$$

Recall that a is the symbol amplitude. The performance depends on the *symbol-amplitude-to-noise ratio* a/σ .



Exercise 1 – Binary transmission through a noisy channel I

A binary communication system transmits a bit $B \in \{0, 1\}$. The receiver delivers a decision $\hat{B} \in \{0, 1\}$. Because of noise, this decision may differ from the transmitted bit.



The probabilities are

$$P(B = 0) = 0.5, \quad (117)$$

$$P(\hat{B} = 1 \mid B = 0) = 0.1, \quad (118)$$

$$P(\hat{B} = 0 \mid B = 1) = 0.2. \quad (119)$$

Figure: Binary communication system

- (a) Find $P(\hat{B} = 0)$ and $P(\hat{B} = 1)$.
- (b) Given that $\hat{B} = 0$, find the probability that $B = 0$ was sent.
- (c) Given that $\hat{B} = 1$, find the probability that $B = 1$ was sent.
- (d) Calculate the probability of error $P_e = P(\hat{B} \neq B)$.
- (e) Calculate the probability of a correct decision $P_c = P(\hat{B} = B)$.

Exercise 2 – A MAP decision from a binary observation I

Consider the same type of system, now with

$$P(\hat{B} = 0 \mid B = 0) = 0.9, \quad P(\hat{B} = 1 \mid B = 1) = 0.6.$$

The prior $\pi_0 = P(B = 0)$ is variable: $0 \leq \pi_0 \leq 1$ and $P(B = 1) = 1 - \pi_0$.

Treat $\hat{B}$ as a *preliminary* decision: a final MAP inference about B may revise it.

Maximum a posteriori (MAP) criterion. For each observed value $\hat{B} = \hat{b}$,

- accept $B = 0$ if

$$P(B = 0 \mid \hat{B} = \hat{b}) > P(B = 1 \mid \hat{B} = \hat{b}) \quad (120)$$

- accept $B = 1$ if

$$P(B = 1 \mid \hat{B} = \hat{b}) > P(B = 0 \mid \hat{B} = \hat{b}) \quad (121)$$

(At equality, either choice is optimal.)

Using Bayes' rule, find the range of π_0 for which MAP strictly favors:

- (a) accepting $B = 0$ after observing $\hat{B} = 0$;
- (b) accepting $B = 0$ for *both* possible values of $\hat{B}$.

Give exact bounds and discuss equality at the boundaries.

Complete the probability model. Since the bits and the receiver decisions are binary,

$$P(B = 1) = 1 - 0.5 = 0.5,$$

$$P(\hat{B} = 0 \mid B = 0) = 1 - 0.1 = 0.9,$$

$$P(\hat{B} = 1 \mid B = 1) = 1 - 0.2 = 0.8.$$

(a) Probabilities of the receiver decisions. Conditioning on the transmitted bit gives

$$P(\hat{B} = j) = \sum_{i=0}^1 P(\hat{B} = j \mid B = i)P(B = i).$$

Therefore,

$$P(\hat{B} = 0) = 0.9(0.5) + 0.2(0.5) = \boxed{0.55},$$

$$P(\hat{B} = 1) = 0.1(0.5) + 0.8(0.5) = \boxed{0.45}.$$

As a check, these probabilities sum to 1.

Solution 1 – Binary transmission through a noisy channel II

(b) **Probability that 0 was sent, given that 0 was decided.** Bayes' rule gives

$$\begin{aligned} P(B = 0 \mid \hat{B} = 0) &= \frac{P(\hat{B} = 0 \mid B = 0)P(B = 0)}{P(\hat{B} = 0)} \\ &= \frac{0.9(0.5)}{0.55} \approx \boxed{0.8182}. \end{aligned}$$

(c) **Probability that 1 was sent, given that 1 was decided.** Similarly,

$$\begin{aligned} P(B = 1 \mid \hat{B} = 1) &= \frac{P(\hat{B} = 1 \mid B = 1)P(B = 1)}{P(\hat{B} = 1)} \\ &= \frac{0.8(0.5)}{0.45} \approx \boxed{0.8889}. \end{aligned}$$

Do not reverse the conditioning: $P(B = 0 \mid \hat{B} = 0)$ is not $P(\hat{B} = 0 \mid B = 0)$.

Solution 1 – Binary transmission through a noisy channel

III

(d) Probability of error. The error event consists of two disjoint cases:

$$\{\hat{B} \neq B\} = \{B = 0, \hat{B} = 1\} \cup \{B = 1, \hat{B} = 0\}.$$

Hence

$$\begin{aligned} P_e &= P(\hat{B} = 1 \mid B = 0)P(B = 0) + P(\hat{B} = 0 \mid B = 1)P(B = 1) \\ &= 0.1(0.5) + 0.2(0.5) = \boxed{0.15}. \end{aligned}$$

(e) Probability of a correct decision. A decision is either correct or incorrect, so

$$\boxed{P_c = 1 - P_e = 0.85.}$$

A direct calculation gives the same result:

$$\begin{aligned} P_c &= P(\hat{B} = 0 \mid B = 0)P(B = 0) + P(\hat{B} = 1 \mid B = 1)P(B = 1) \\ &= 0.9(0.5) + 0.8(0.5) = 0.85. \end{aligned}$$

Complete the new probability model. The missing conditional probabilities are

$$P(\hat{B} = 1 \mid B = 0) = 0.1, \quad P(\hat{B} = 0 \mid B = 1) = 0.4.$$

With $\pi_0 = P(B = 0)$ and $\pi_1 = 1 - \pi_0$, total probability gives

$$P(\hat{B} = 0) = 0.9\pi_0 + 0.4(1 - \pi_0) = 0.4 + 0.5\pi_0,$$

$$P(\hat{B} = 1) = 0.1\pi_0 + 0.6(1 - \pi_0) = 0.6 - 0.5\pi_0.$$

These output probabilities are positive for every $\pi_0 \in [0, 1]$.

Apply Bayes' rule to each possible observed output.

$$P(B = i \mid \hat{B} = j) = \frac{P(\hat{B} = j \mid B = i)\pi_i}{P(\hat{B} = j)}, \quad i, j \in \{0, 1\}.$$

For each fixed j , MAP selects the larger of the two posteriors. This rule minimizes the probability of an incorrect final inference.

(a) **Decide** $B = 0$ **after observing** $\hat{B} = 0$. The two posterior probabilities are

$$P(B = 0 \mid \hat{B} = 0) = \frac{0.9\pi_0}{0.4 + 0.5\pi_0},$$

$$P(B = 1 \mid \hat{B} = 0) = \frac{0.4(1 - \pi_0)}{0.4 + 0.5\pi_0}.$$

MAP strictly favors $B = 0$ when

$$\begin{aligned}\frac{0.9\pi_0}{0.4 + 0.5\pi_0} &> \frac{0.4(1 - \pi_0)}{0.4 + 0.5\pi_0} \\ \iff 0.9\pi_0 &> 0.4(1 - \pi_0) \\ \iff 1.3\pi_0 &> 0.4 \\ \iff \pi_0 &> \frac{4}{13} \approx 0.3077.\end{aligned}$$

At $\pi_0 = 0.3077$, both posteriors are $1/2$: either final decision is optimal for the observation $\hat{B} = 0$.

(b) **Decide $B = 0$ regardless of the observed output.** Which can be read as:

$$\begin{cases} \text{Decide } B = 0 \text{ if } \hat{B} = 0, \text{ and} \\ \text{Decide } B = 0 \text{ if } \hat{B} = 1 \end{cases} \quad (122)$$

For $\hat{B} = 1$, Bayes' rule gives

$$P(B = 0 \mid \hat{B} = 1) = \frac{0.1\pi_0}{0.6 - 0.5\pi_0}, \quad P(B = 1 \mid \hat{B} = 1) = \frac{0.6(1 - \pi_0)}{0.6 - 0.5\pi_0}.$$

$$\begin{aligned} \frac{0.1\pi_0}{0.6 - 0.5\pi_0} &> \frac{0.6(1 - \pi_0)}{0.6 - 0.5\pi_0} \\ \iff 0.1\pi_0 &> 0.6(1 - \pi_0) \\ \iff \pi_0 &> \frac{6}{7} \approx 0.8571. \end{aligned}$$

We need both $\pi_0 > 4/13$ and $\pi_0 > 6/7$. Therefore,

$$\boxed{0.8571 < \pi_0 \leq 1.}$$

At $\pi_0 = 6/7$, the observation $\hat{B} = 1$ gives an equal posterior probability for both $B = 0$ and $B = 1$.

Summary – How the prior changes the final MAP inference.

Prior $\pi_0 = P(B = 0)$	Observe $\hat{B} = 0$	Observe $\hat{B} = 1$
$0 \leq \pi_0 < 0.3077$	Decide $B = 1$	Decide $B = 1$
$\pi_0 = 0.3077$	Either bit	Decide $B = 1$
$0.3077 < \pi_0 < 0.8571$	Decide $B = 0$	Decide $B = 1$
$\pi_0 = 0.8571$	Decide $B = 0$	Either bit
$0.8571 < \pi_0 \leq 1$	Decide $B = 0$	Decide $B = 0$

For intermediate priors, MAP keeps the preliminary decision. For sufficiently unequal priors, MAP always selects the more probable transmitted bit, even when the preliminary decision is the opposite bit.

Boundary convention. The strict ranges above identify a unique decision. If we choose $B = 0$ when the two posterior probabilities are equal, the lower bounds in answers (a) and (b) become inclusive: $\pi_0 \geq 4/13$ and $\pi_0 \geq 6/7$.

Hypothesis testing and optimal decisions

Waveform reception, matched filter and BER

A random process $n(t, \omega)$ describes random waveforms: a realization fixes ω ; varying t traces one waveform.

For the real additive white Gaussian noise (AWGN) model, the two-sided noise power spectral density is

$$S_n(f) = \frac{N_0}{2}, \quad -\infty < f < \infty. \quad (123)$$

Its ideal autocorrelation is written formally as

$$R_n(\tau) = \mathbb{E}[n(t)n(t+\tau)] = \frac{N_0}{2}\delta(\tau). \quad (124)$$

White describes the spectrum; **Gaussian** describes the probability law of linear observations. They are distinct properties.

After a real receive filter $h(t)$ and sampling at T_0 ,

$$N_h = \int_{-\infty}^{\infty} n(\tau)h(T_0 - \tau) d\tau. \quad (125)$$

For a finite-energy filter, this Gaussian variable has

$$\mathbb{E}[N_h] = 0, \quad (126)$$

$$\sigma_h^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} |h(t)|^2 dt \quad (127)$$

$$= \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2 df. \quad (128)$$

Ideal continuous-time white noise has infinite total power; it is an idealized generalized process. We assign a finite variance to its *filtered observation*.

σ_h^2 corresponds to the noise power at the filter output.

Matched filter: maximize the decision-point SNR I

Recall the notation of the previous lesson: a_k is the transmitted symbol and $g(t)$ is the overall pulse observed after (receiver) filtering. For one isolated symbol, the input of the receive filter is

$$r(t) = a_k p(t - kT_{\text{symp}}) + n(t), \quad (129)$$

where $p(t)$ combines the transmit pulse and the channel response.

The energy of the pulse arriving at the receiver is:

$$E_p = \int_{-\infty}^{\infty} |p(t)|^2 dt. \quad (130)$$

Let $h(t)$ be the receive filter and define

$$g(t) = (p * h)(t). \quad (131)$$

We sample at

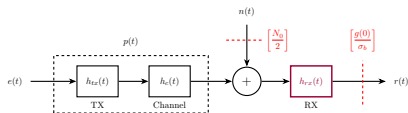
$$t_k = kT_{\text{symp}} + T_0. \quad (132)$$

The useful contribution at the decision instant is therefore

$$a_k g(T_0), \quad (133)$$

with

$$g(T_0) = \int_{-\infty}^{\infty} p(\tau) h(T_0 - \tau) d\tau. \quad (134)$$



Matched filter: maximize the decision-point SNR II

For AWGN with two-sided PSD $N_0/2$,

$$\text{SNR}_{\text{out}} = \frac{|a_k|^2 |g(T_0)|^2}{(N_0/2) \int_{-\infty}^{\infty} |h(t)|^2 dt}. \quad (135)$$

By Cauchy–Schwarz,

$$|g(T_0)|^2 \leq \left(\int |p(t)|^2 dt \right) \left(\int |h(t)|^2 dt \right). \quad (136)$$

Hence

$$\boxed{\text{SNR}_{\text{out}} \leq \frac{2|a_k|^2 E_p}{N_0}}. \quad (137)$$

Equality is obtained with the **matched filter** (♡)

$$\boxed{h(t) = c p^*(T_0 - t)}, \quad c \neq 0. \quad (138)$$

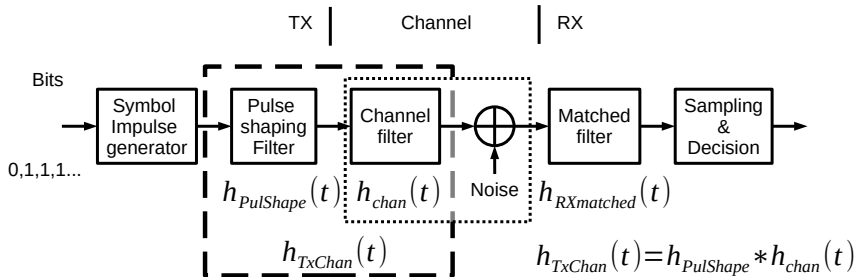
Equivalently, in frequency,

$$H(f) = c P^*(f) e^{-j2\pi f T_0}. \quad (139)$$

Thus the best receive filter is a **time-reversed, conjugated copy of the received pulse**.

Then, the maximum SNR depends on the received pulse energy E_p , not on the particular shape of $p(t)$.

Conclusion : Transceiver model (optimal receiver)



Optimal transceiver ♥

The TX and RX filters $h_{PulShape}(t)$ and $h_{RXmatched}(t)$ are set so that:

- $h_{tot}(t) = h_{PulShape} * h_{chan} * h_{RXmatched}(t)$ is an *ISI Nyquist* filter
- $h_{RXmatched} = h_{TxChan}^*(-t)$ is the *matched filter* of h_{TxChan}

For a constellation $\mathcal{A} = \{a_0, \dots, a_{M-1}\}$, let $H_i : A = a_i$ and $\pi_i = P(A = a_i)$.

The observation regions D_i form a partition.
The **transition probabilities** are

$$P(\hat{A} = a_j | A = a_i) = P(\hat{A} = a_j | H_i) \quad (140)$$

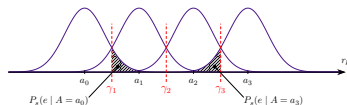
$$= \int_{D_j} f_{R|H_i}(r | a_i) dr. \quad (141)$$

$$P(A \neq \hat{A}) = \sum_i \pi_i P(\hat{A} \neq a_i | A = a_i) \quad (142)$$

$$= 1 - \sum_i \pi_i P(\hat{A} = a_i | A = a_i). \quad (143)$$

For each possible transmitted symbol, define its prior-weighted likelihood score

$$S_i(r) = \pi_i f_{R|H_i}(r | a_i). \quad (144)$$



The MAP receiver chooses the symbol with the largest score:

$$\hat{A}_{\text{MAP}}(r) = a_i \quad \text{if } S_i(r) \geq S_j(r),$$

for every $j \in \{0, \dots, M-1\}$. (145)

If several symbols have the same largest score, any of them is an optimal choice.

For $R = A + N$, with independent $N \sim \mathcal{N}(0, \sigma^2)$,

$$\pi_i f_{R|H_i}(r | a_i) \propto \pi_i \exp \left[-\frac{(r - a_i)^2}{2\sigma^2} \right]. \quad (146)$$

Thus MAP is equivalent to minimizing

$$(r - a_i)^2 - 2\sigma^2 \ln \pi_i. \quad (147)$$

With equal priors, this becomes

nearest-neighbor detection:

$$\hat{a} = \arg \min_{a_i \in \mathcal{A}} |r - a_i|^2. \quad (148)$$

With average received signal power P_{av} [W],

$$E_b = \frac{P_{av}}{R_b}, \quad E_s = \frac{P_{av}}{R_s}. \quad (149)$$

For $k = \log_2 M$ bits per symbol,

$$R_b = kR_s, \quad E_s = kE_b. \quad (150)$$

E_b is measured in joules; N_0 in watts per hertz, also equivalent to joules. Therefore E_b/N_0 is dimensionless.

$$\left(\frac{E_b}{N_0}\right)_{dB} = 10 \log_{10} \left(\frac{E_b}{N_0}\right). \quad (151)$$

Use $10 \log_{10}$ for an energy ratio, not $20 \log_{10}$.

For equal-prior binary antipodal signaling,

$$s_0(t) = -\sqrt{E_b} p(t), \quad s_1(t) = +\sqrt{E_b} p(t),$$

$$\int |p(t)|^2 dt = 1.$$

The matched-filter observation is

$$Y = \pm \sqrt{E_b} + N, \quad N \sim \mathcal{N}(0, N_0/2). \quad (152)$$

Use $\gamma = 0$ and standardize the distance to the boundary:

$$P_b = Q\left(\frac{\sqrt{E_b}}{\sqrt{N_0/2}}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \quad (153)$$

This is also the coherent **BPSK** BER once the signal is placed on an RF carrier.

For $P_b = 10^{-3}$, $Q^{-1}(10^{-3}) \simeq 3.090$ gives $E_b/N_0 \simeq 4.775$, or 6.79 dB.

Real L-PAM: symbol errors, bit errors and Gray labeling

Let $L = 2^m$ equally likely levels be

$$\mathcal{A}_L = \{-(L-1)d, \dots, -d, d, \dots, (L-1)d\}. \quad (154)$$

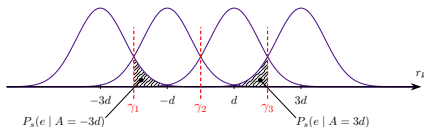
Neighbor spacing is $2d$; best boundaries are at the midpoints. Let $N \sim \mathcal{N}(0, \sigma^2)$.

Each outer symbol has one error tail $Q(d/\sigma)$; each of the $L - 2$ inner symbols has two:

$$P_{s,\text{PAM}} = \frac{2Q(d/\sigma) + 2(L-2)Q(d/\sigma)}{L} \quad (155)$$

$$= 2 \left(1 - \frac{1}{L}\right) Q\left(\frac{d}{\sigma}\right). \quad (156)$$

This symbol-error expression is **exact** under the stated model.



For a unit-energy pulse,

$$E_{s,\text{PAM}} = \mathbb{E}[A^2] = \frac{L^2 - 1}{3} d^2. \quad (157)$$

An incorrect symbol contains between 1 and m incorrect bits, hence

$$\frac{P_s}{m} \leq P_b \leq P_s. \quad (158)$$

Gray labeling assigns one-bit-different labels to adjacent levels, for example

level/ d	-3	-1	1	3
label	00	01	11	10

(159)

When nearest-neighbor errors dominate,

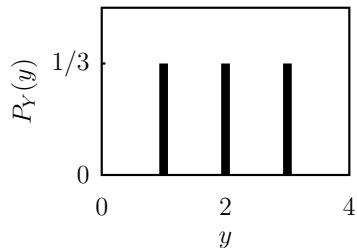
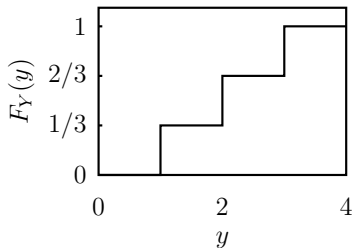
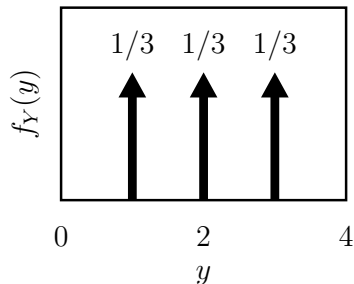
$$P_{b,\text{PAM}} \approx \frac{P_{s,\text{PAM}}}{\log_2 L}. \quad (160)$$

Gray labeling does not make this an exact identity at every SNR.

-  H. Hsu, *Schaum's Outline of Probability, Random Variables, and Random Processes, Fourth Edition*.
McGraw Hill LLC, 2019.
-  R. Yates and D. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*.
Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers, Wiley, 2014.

Appendix slides

The PMF, CDF, and PDF of a discrete random variable Y .



Over n received bits,

$$K = \sum_{k=1}^n J_k, \quad \hat{P}_b = \frac{K}{n}. \quad (161)$$

If the J_k are iid Bernoulli, then

$$P(K = m) = \binom{n}{m} P_b^m (1 - P_b)^{n-m}, \quad (162)$$

$$\mathbb{E}[K] = nP_b, \quad \text{Var}(\hat{P}_b) = \frac{P_b(1 - P_b)}{n}. \quad (163)$$

$\hat{P}_b$ converges to P_b as the experiment grows. A finite measured BER is *an estimate*, not the theoretical probability itself.

$\binom{n}{m} = \frac{n!}{m!(n-m)!}$, usually read as " n choose m " (the number of ways to choose m elements from n).

We would be tempted to write the basic Bayes' rule as:

$$P(H_i | R = r) = \frac{P(H_i \cap \{R = r\})}{P(R = r)} \quad (164)$$

but both the numerator and the denominator are zero for a continuous random variable, making this expression undefined. So to develop the proof of (79), we need to consider conditioning on a small interval around r and then taking the limit as the interval width goes to zero.

For variables with finite second moments,

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \quad (165)$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (166)$$

Independence means that the joint law factorizes; when a joint PDF exists,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y). \quad (167)$$

Independence implies zero covariance. The converse is false in general.

The useful Gaussian exception

For a *jointly Gaussian* pair, zero covariance is equivalent to independence. Gaussian marginal distributions alone do not guarantee joint Gaussianity.

Two independent noise coordinates If N_I, N_Q are iid $\mathcal{N}(0, \sigma^2)$,

$$f_{N_I, N_Q}(n_I, n_Q) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{n_I^2 + n_Q^2}{2\sigma^2}\right]. \quad (168)$$

Contours of equal density are circles: only the distance from the origin matters.

Linear combinations of jointly Gaussian variables remain Gaussian, with

$$\mathbb{E}[cX + dY] = c\mu_X + d\mu_Y, \quad (169)$$

$$\text{Var}(cX + dY) = c^2\sigma_X^2 + d^2\sigma_Y^2 \quad (170)$$

$$+ 2cd \text{Cov}(X, Y). \quad (171)$$

These facts will justify independent noise components after orthogonal I/Q projections.

Hsu §8.3 “Decision Tests”: one-page overview

For binary tests, $\Lambda(r) = f_1(r)/f_0(r) \geq \eta \Rightarrow H_1$. Different criteria give different thresholds.

Test in Hsu §8.3	Design criterion	Threshold / interpretation
A. Maximum likelihood	Largest likelihood	$\eta = 1$; no prior weighting
B. MAP	Largest posterior probability	$\eta = \pi_0/\pi_1$
C. Neyman–Pearson	Maximize P_D for $P_{FA} \leq \alpha_0$	Set η by the false-alarm constraint
D. Bayes	Minimum average decision cost	$\eta = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}$
E. Minimum error	Error cost 1; correct-decision cost 0	Identical to MAP
F. Minimax	Minimum worst-case risk over priors	Bayes rule for a least-favorable prior

C_{ij} is the cost of deciding H_i when H_j is true; the risk is $\mathcal{R} = \sum_{i,j} C_{ij}\pi_j P(R \in D_i | H_j)$. The Bayes row assumes $C_{10} > C_{00}$ and $C_{01} > C_{11}$. With discrete observations, randomization at a boundary may be needed to satisfy an exact Neyman–Pearson constraint.

Read Hsu, §8.3, for the derivations and the distinctions between these six tests. Our equiprobable uncoded symbol detector uses $ML = MAP$.

Correlation, sufficient observations and the Nyquist condition

Normalize a real received pulse to unit energy, $\int |p(t)|^2 dt = 1$, and choose a unit-gain matched filter.

At the decision instant, filtering is correlation:

$$Y = \int_{-\infty}^{\infty} r(t)p(t) dt = A + N, \quad N \sim \mathcal{N}(0, N_0/2). \quad (172)$$

For the hypotheses $r(t) = a_i p(t) + n(t)$, all other orthogonal projections contain only noise with the same law under every hypothesis.

Thus Y is a **sufficient observation for the decision**: those additional noise-only coordinates do not change the likelihood comparison.

With several signal-space dimensions, retain one correlation output per basis function. I/Q transmission will require two.

For a symbol stream,

$$r(t) = \sum_{\ell} A_{\ell} p(t - \ell T_s) + n(t). \quad (173)$$

Zero ISI after matched filtering requires

$$\int p(t - kT_s)p(t - \ell T_s) dt = \delta_{k\ell}. \quad (174)$$

Here $\delta_{k\ell} = 1$ if $k = \ell$, and 0 otherwise. Then

$$Y_k = \int r(t)p(t - kT_s) dt = A_k + N_k, \quad (175)$$

$$\mathbb{E}[N_k N_{\ell}] = \frac{N_0}{2} \delta_{k\ell}. \quad (176)$$

The jointly Gaussian noise samples are independent. A matched filter alone does *not* ensure zero ISI: the received pulse must also satisfy this root-Nyquist condition.