ELG5375: Digital Communications

Lecture 6

Dr. Sergey Loyka

EECS, University of Ottawa

March 10, 2026

Statistical Hypothesis Testing

- how to make optimal decisions under uncertainty?
- statistical/Bayesian hypothesis testing (inference)
- many applications
  - communication/information theory
  - signal processing
  - stochastic control
  - big data/AI/ML
  - search engines
- consider the most simple yet non-trivial case first
- extend to more complicated scenarios later

Optimal guessing rule for broken Tx-Rx link

Random but **optimal guessing**

- consider the most simple but non-trivial setup
- Tx sends $m = 0$ or 1 with probabilities P_0 and P_1
- Rx knows P_0, P_1 but receives nothing, e.g. broken Tx-Rx link
- Q: what is the optimal guessing rule? $\widehat{m} =?$
- minimize P_e (BER):

$$P_e \triangleq \Pr\{\hat{m} \neq m\} \; \rightarrow \; \min P_e =? \tag{1}$$

Optimal guessing rule for broken link

- simple but powerful trick: via lower bound (LB)

$$\begin{aligned} P_e &= P_{e|0}P_0 + P_{e|1}P_1 && (2)\\ &= \Pr\{\hat{m}=1\}P_0 + \Pr\{\hat{m}=0\}P_1 && (3)\\ &\geq P_0 \text{ if } P_1 > P_0 && (4) \end{aligned}$$

- LB: "=" if $\Pr\{\hat{m}=1\} = 1 \ \rightarrow \ \hat{m}=1$
- likewise for $P_1 < P_0$, so that

$$\hat{m} = \arg\max_m P_m = \begin{cases} 1, \text{ if } P_1 > P_0 \\ 0, \text{ if } P_1 < P_0 \end{cases} \quad \rightarrow \ \min P_e = \min\{P_0, P_1\} \tag{5}$$

- **no any other rule can do better**

Link is not broken: heuristic approach

Message m is corrupted by noise ξ, Rx measures y:

$$y = m + \xi \tag{6}$$

Heuristic approach: replace $P(m) = P_m$ by $P(m|y)$:

$$\hat{m}(y) = \arg\max_m P(m|y) = \begin{cases} 1, & \text{if } P(1|y) > P(0|y) \\ 0, & \text{if } P(1|y) < P(0|y) \end{cases} \tag{7}$$

where

$$p(y|m) = p_\xi(y-m), \;\; p(y) = \sum_m p(y|m)P(m)$$

$$P(m|y) = \frac{p(y|m)P(m)}{p(y)} = \frac{p_\xi(y-m)P(m)}{\sum_m p_\xi(y-m)P(m)} \tag{8}$$

Link is not broken: heuristic argument

Error probability (from "broken link" case):

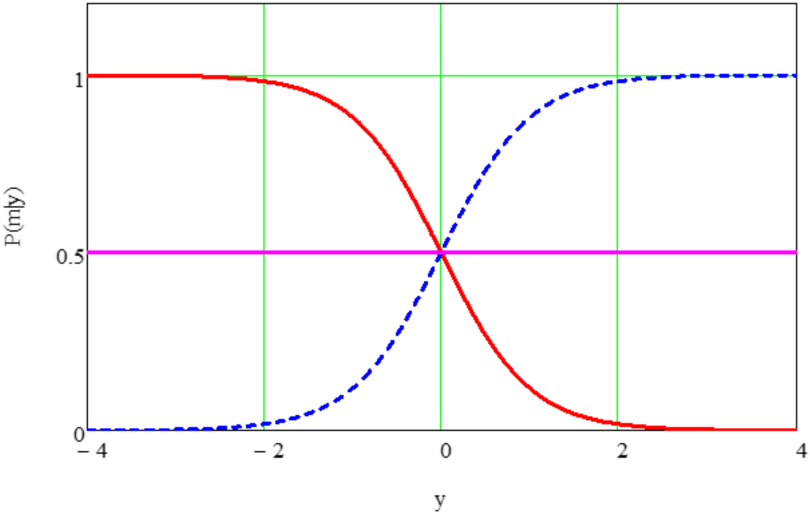
$$P_{e|y} = \min_m P(m|y), \tag{9}$$

$$P_e = \int_{\mathbb{R}} P_{e|y} p(y) dy = \int_{\mathbb{R}} \min_m P(m|y) p(y) dy \tag{10}$$

$P_{e|y} \Rightarrow$ a measure of confidence for given (measured) y

Conditional vs. unconditional probabilities

$P(m|y)$ and $P(m)$:



Link is not broken: heuristic argument

Error probability:

$$P_{e|y} = \min_m P(m|y), \tag{11}$$

$$P_e = \int_{\mathbb{R}} P_{e|y} p(y) dy = \int_{\mathbb{R}} \min_m P(m|y) p(y) dy \tag{12}$$

But: **can we do better**?

Optimal decision rule: link is not broken

Message m is corrupted by noise ξ, Rx measures y,

$$y = m + \xi \tag{13}$$

- same trick as for broken link (LB), but more work is needed
- Rx decision rule: via decision regions Ω_m

$$\boxed{\text{Rx: } \hat{m}(y) = m \text{ if } y \in \Omega_m} \tag{14}$$

- properties

$$\Omega_1 \cap \Omega_0 = \emptyset,\ \Omega_1 \cup \Omega_0 = \mathbb{R}, \tag{15}$$

$$\Omega_1 = \Omega_0^c = \mathbb{R} - \Omega_0,\ \mathbb{R} = (-\infty, \infty) \tag{16}$$

Statistical hypothesis testing

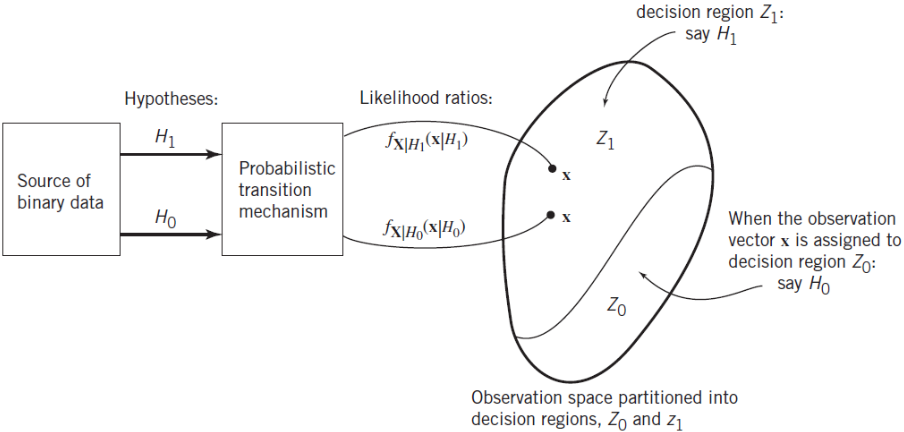


Figure: An illustration of binary hypothesis-testing problem[1]

[1] S. Haykin, Digital Communication Systems, Wiley, 2014.

Optimal Rx as a statistical hypothesis tester

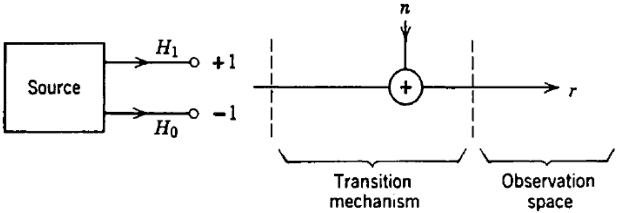


Figure: Hypothesis-testing problem for noisy transmission/measurement[2]

[2] H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, 2001.

Optimal decision rule: link is not broken

- our strategy:
  1. find good LB for P_e (BER) that holds for **any** Rx
  2. show achievability $\rightarrow$ optimal Rx (decision rule)

Optimal decision rule: link is not broken

- some definitions

$$p(y,m) = p(y|m)P_m,\ \ f(y) \triangleq \min_{m=0,1} p(y,m) \le p(y,1), p(y,0) \quad (17)$$

- error probability P_e:

$$\begin{aligned} P_e &= P_{e|0}P_0 + P_{e|1}P_1 && (18)\\ &= \Pr\{y \in \Omega_1 | m=0\}P_0 + \Pr\{y \in \Omega_0 | m=1\}P_1 && (19)\\ &= \int_{\Omega_1} p(y,0)dy + \int_{\Omega_0} p(y,1)dy && (20)\\ &\ge \int_{\Omega_1} f(y)dy + \int_{\Omega_0} f(y)dy = \int_{\mathbb{R}} f(y)dy = \mathrm{LB} && (21) \end{aligned}$$

- LB: independent of Ω_0, Ω_1, i.e.
- holds for **any** decision rule, including optimal one

Optimal decision rule: how to achieve the LB?

- how to achieve the LB?

$$P_{e0} = \int_{\Omega_1} p(y,0)dy \geq \int_{\Omega_1} f(y)dy \tag{22}$$

- "=" if $p(y,0) = f(y)\ \forall y \in \Omega_1$, i.e

$$\boxed{p(y,0) < p(y,1)\ \forall y \in \Omega_1} \tag{23}$$

- likewise for P_{e1}:

$$P_{e1} = \int_{\Omega_0} p(y,1)dy \tag{24}$$

$$= \int_{\Omega_0} f(y)dy \ \boxed{\text{if } p(y,0) \geq p(y,1)\ \forall y \in \Omega_0} \tag{25}$$

Optimal decision rule: link is not broken

- optimal decision regions

$$\Omega_1 = \{y : p(y,1) > p(y,0)\} \quad (26)$$
$$\Omega_0 = \{y : p(y,1) \leq p(y,0)\} \quad (27)$$

- optimal decision rule:

$$\boxed{\widehat{m}(y) = \begin{cases} 1, & \text{if } y \in \Omega_1 \\ 0, & \text{if } y \in \Omega_0 \end{cases}} \quad (28)$$

- equivalently

$$\widehat{m}(y) = \arg\max_m p(y,m) \quad (29)$$

- required properties of Ω_m: OK

Optimal decision = MAP decision

- maximum a posteriori probability (MAP)
- note the following:

$$p(y,m) = p(y|m)P_m = P(m|y)p(y) \tag{30}$$

$$\Rightarrow \arg\max_m p(y,m) = \arg\max_m P(m|y) \tag{31}$$

- so that optimal decision = MAP decision:

$$\boxed{\widehat{m}(y) = \arg\max_m p(y,m) = \arg\max_m P(m|y) \triangleq \mathrm{MAP}} \tag{32}$$

- **no any other decision rule can do better**
- therefore, the heuristic approach is **optimal**
- note: $P(m|y) \neq P(m)$ and this is very important

Optimal decision via likelihood ratio (LR)

- the optimal decision rule

$$y \in \Omega_1 \triangleq \{y : p(y,1) > p(y,0)\} \ \rightarrow \ \widehat{m} = 1 \tag{33}$$

$$y \in \Omega_0 \triangleq \{y : p(y,0) \geq p(y,1)\} \ \rightarrow \ \widehat{m} = 0 \tag{34}$$

- same via likelihood ratio $\Lambda(y)$:

$$\Lambda(y) = \frac{p(y|1)}{p(y|0)} \underset{0}{\overset{1}{\gtrless}} \frac{P_0}{P_1} \tag{35}$$

- log-likelihood ratio (LLR) $\log \Lambda(y)$ is more convenient in many cases

$$\log \Lambda(y) \underset{0}{\overset{1}{\gtrless}} \log \frac{P_0}{P_1} \tag{36}$$

Optimal decision via likelihood ratio (LR)

- optimal decision regions via LR

$$y \in \ \Omega_1 : \Lambda(y) > \alpha \ \rightarrow \ \widehat{m} = 1 \qquad (37)$$

$$y \in \ \Omega_0 : \Lambda(y) \leq \alpha \ \rightarrow \ \widehat{m} = 0 \qquad (38)$$

- where $\alpha = \frac{P_0}{P_1}$ (decision threshold)
- only $\Lambda(y)$ is needed for optimal decision $\rightarrow$ "sufficient statistics"
- Rx signal y can be multi-dimensional... **but**
- $\Lambda(y)$ is always one-dimensional (simple)
- sufficient statistics: significant reduction in dimensionality/complexity

Optimal decision via likelihood ratio (LR)

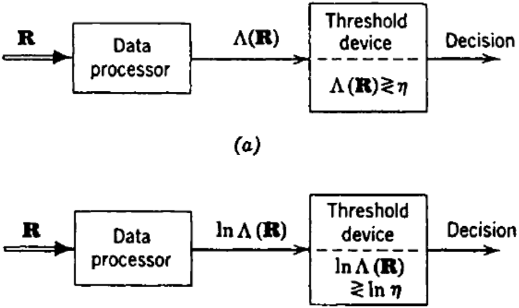


Figure: Likelihood ratio Rx/decision device[3]

[3]H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, 2001.

Maximum likelihood (ML) decision rule

- very often $P_1 = P_0 \rightarrow \alpha = 1$ (worst case)
- can also be used for unknown P_1, P_0
- maximum likelihood (ML) decision rule:

$$y \in \Omega_1 : \Lambda(y) > 1 \rightarrow \widehat{m} = 1 \tag{39}$$

$$y \in \Omega_0 : \Lambda(y) \leq 1 \rightarrow \widehat{m} = 0 \tag{40}$$

An example: binary transmission in Gaussian noise

- system model: noisy measurement

$$y = m + \xi \tag{41}$$

- $m = 1$ or 0 is binary Tx signal (message)
- $\xi \sim \mathcal{N}(0, \sigma_0^2)$ is Gaussian noise
- also: inaccurate (noisy) measurement, radar
- statistical hypothesis testing:

$$\begin{aligned} H_0 &: \; m = 0 \\ H_1 &: \; m = 1 \end{aligned} \tag{42}$$

An example: binary transmission in Gaussian noise

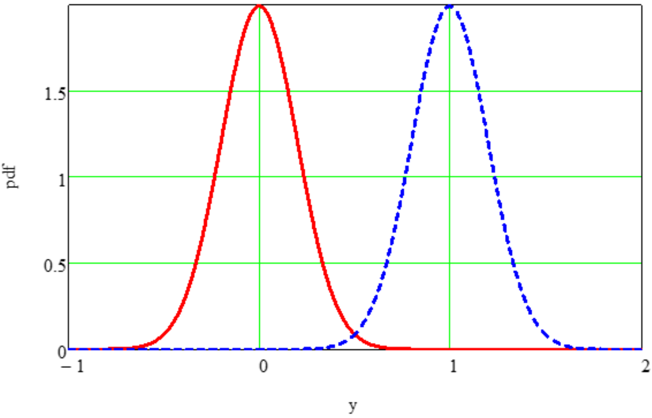


Figure: Conditional Gaussian pdf $p(y|m)$, $\sigma_0 = 0.2$.

An example: binary transmission in Gaussian noise

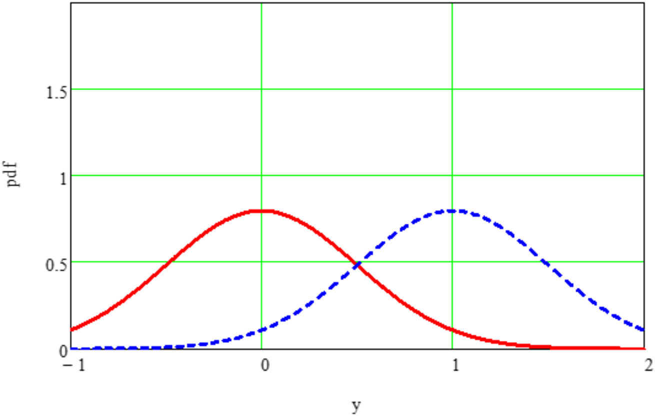


Figure: Conditional Gaussian pdf $p(y|m)$, $\sigma_0 = 0.5$.

An example: binary transmission in Gaussian noise

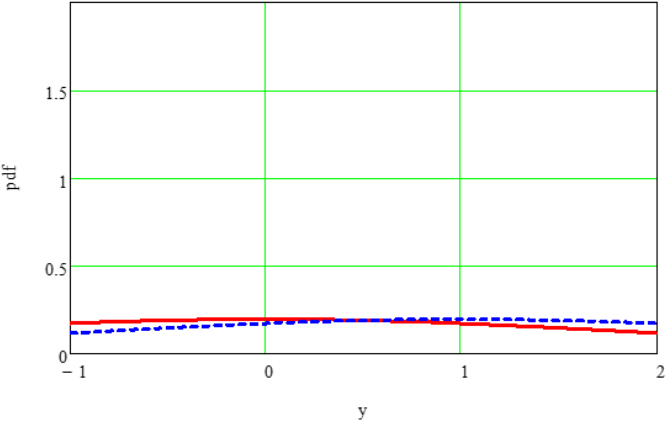
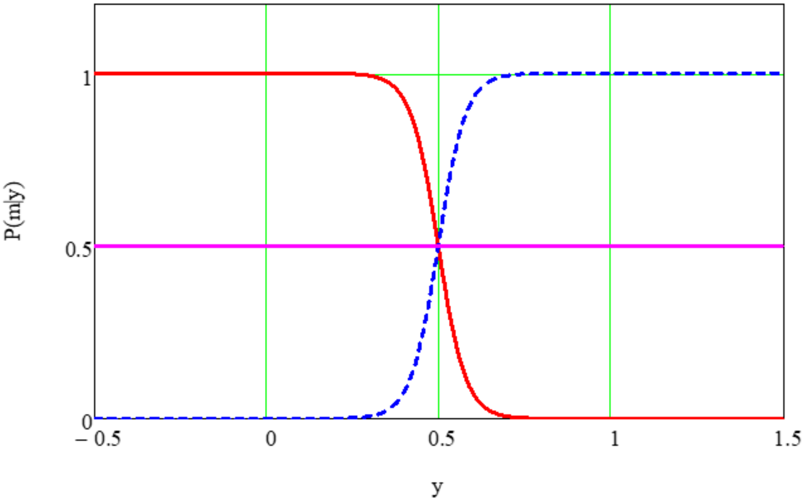


Figure: Conditional Gaussian pdf $p(y|m)$, $\sigma_0 = 2$.

An example: binary transmission in Gaussian noise

$P(m|y)$ and $P(m)$:



An example: binary transmission in Gaussian noise

- system model: noisy measurement

$$y = m + \xi \tag{43}$$

- LLR & ML decision rule:

$$\ln \Lambda(y) = \frac{y - \frac{1}{2}}{\sigma_0^2} \underset{0}{\overset{1}{\gtrless}} 0 \iff \boxed{y \underset{0}{\overset{1}{\gtrless}} \frac{1}{2}} \tag{44}$$

- equivalently

$$y \in \ \Omega_1 : y > 1/2 \ \rightarrow \ \widehat{m} = 1 \tag{45}$$

$$y \in \ \Omega_0 : y \leq 1/2 \ \rightarrow \ \widehat{m} = 0 \tag{46}$$

- **Q**: find optimal decision rule if $m = \pm 1$ or, more generally, $\pm A$

An example: binary transmission in Gaussian noise

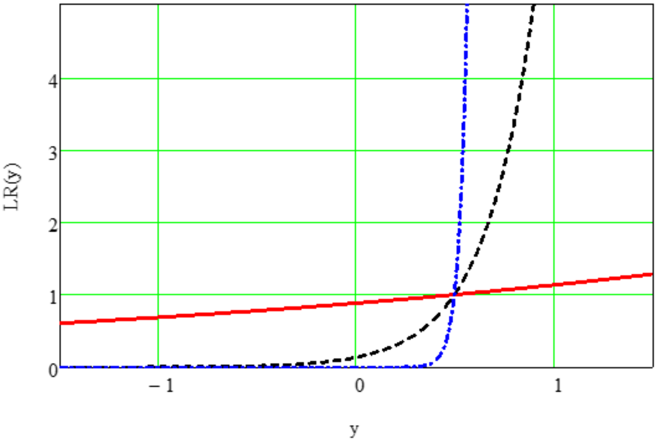


Figure: Likelihood ratio (LR) in AWGN with 2-PAM; $\sigma_0 = 0.2, 0.5, 2$.

An example: binary transmission in Gaussian noise

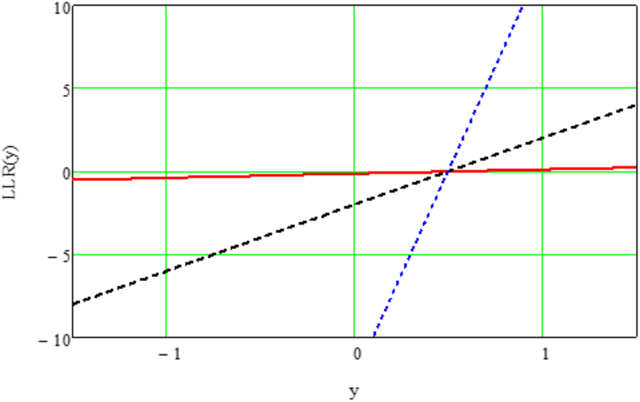


Figure: Log-likelihood ratio (LLR) in AWGN with 2-PAM; $\sigma_0 = 0.2, 0.5, 2.$

An example: binary transmission in Gaussian noise

- Probability of error $P_e = \Pr\{\widehat{m} \neq m\}$:

$$P_{e|1} = P_{e|0} = P_e = Q(1/(2\sigma_0)) \tag{47}$$

- where $Q(\cdot)$ is the CCDF of the standard Gaussian RV $\mathcal{N}(0,1)$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt \tag{48}$$

- Q: prove this; additionally
- show that
  - $P_e = 0$ if $\sigma_0 = 0$ (explain, in plain language, why)
  - $P_e = 1/2$ if $\sigma_0 = \infty$ (explain why)
- what is the maximum possible value of P_e?

An example: binary transmission in Gaussian noise

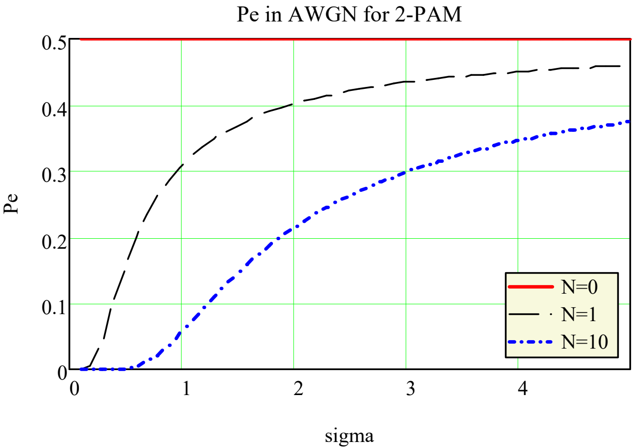


Figure: Error probability in AWGN with 2-PAM

An example: binary transmission in Gaussian noise

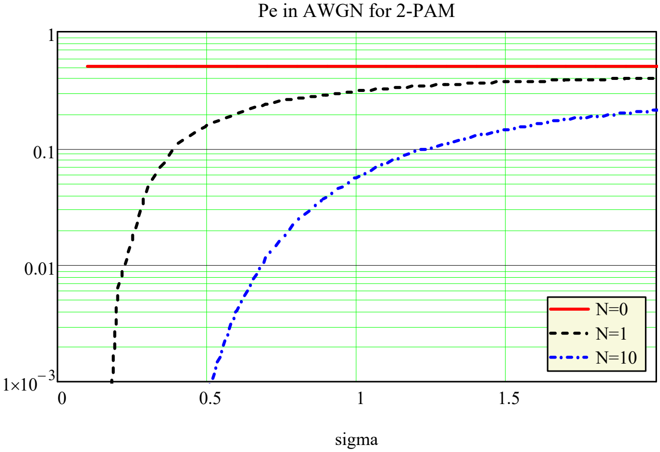


Figure: Error probability in AWGN with 2-PAM

Advanced system: multiple transmissions (measurements)

- transmit (measure) same m multiple (N) times:

$$y_i = m + \xi_i, \;\; i = 1...N \tag{49}$$

- $\xi_i \sim \mathcal{N}(0, \sigma_0^2)$, iid
- Rx vector $\mathbf{y} = [y_1, ..., y_N]$
- LLR & ML decision rule:

$$\ln \Lambda(\mathbf{y}) = \frac{N}{\sigma_0^2}(\overline{y}_N - 1/2) \underset{0}{\overset{1}{\gtrless}} 0 \;\Leftrightarrow\; \boxed{\overline{y}_N \underset{0}{\overset{1}{\gtrless}} 1/2} \tag{50}$$

- $\overline{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the sample average,
- only $\overline{y}_N$ is needed for optimal decision, not the whole vector $\mathbf{y} = [y_1, ..., y_N]$, i.e.
- $\overline{y}_N =$ **sufficient statistics**

Extension to M-ary modulation/message

- sending $n_b > 1$ bits in a message:

$$m \in \{1, 2, ...M\}, \;\; n_b = \log_2 M \tag{51}$$

- much higher data rate is possible
- one message/symbol carriers n_b bits
- **optimal Rx?**
- apply binary message ($M = 2$) techniques

M-ary modulation/message: broken link

- **broken link**: extension of $M = 2$ case, with minor modifications
- bounding trick still works, with $P_c = 1 - P_e$:

$$P_c = \sum_{m=1}^{M} P_{c|m} P_m \leq P_{max} \triangleq \max_m P_m = \mathrm{UB} \tag{52}$$

- so that $P_e \geq 1 - P_{max} \triangleq \mathrm{LB}$
- UB/LB: "=" if

$$\Pr\{\hat{m} = m^*\} = 1 \;\rightarrow\; \hat{m} = m^* \triangleq \arg\max_m P_m \tag{53}$$

- **no any other rule can do better**

M-ary message: link is not broken

- **Heuristic approach**: same as before ($M = 2$)
- replace $P(m) = P_m$ by $P(m|y)$:

$$\hat{m}(y) = \arg\max_m P(m|y) \tag{54}$$

- message error probability :

$$P_{e|y} = 1 - P_{c|y} = 1 - \max_m P(m|y), \tag{55}$$

$$P_e = \int_{\mathbb{R}} P_{e|y} p(y) dy = 1 - \int_{\mathbb{R}} \max_m P(m|y) p(y) dy \tag{56}$$

- **can we do better?**

Optimal Rx for M-ary message

- link (channel) is not broken
- channel: stochastic mapping (cond. prob. dist.) $p(y|m)$

$$m \xrightarrow{p(y|m)} y \tag{57}$$

- e.g., for M-PAM with additive noise,

$$m \xrightarrow{\text{M-PAM}} a_m \xrightarrow{p(y|a_m)} y = a_m + \xi \tag{58}$$

- Rx: via decision regions, as before:

$$\widehat{m}(y) = D\{y\},\ \Omega_m = \{y : D\{y\} = m\} \tag{59}$$

Extension to M-ary modulation/message

- symbol (not bit!) error probabilities:

$$P_{e|m} = \Pr\{y \notin \Omega_m | m\} = \int_{\Omega_m^c} p(y|m) dy \tag{60}$$

$$P_e = \sum_{m=1}^{M} P_{e|m} P_m = \sum_m \int_{\Omega_m^c} p(y, m) dy \tag{61}$$

- **optimal Rx – via lower bound** with $F(y) \triangleq \max_m p(y, m)$

$$P_e = 1 - P_c \geq 1 - \int_{\mathbb{R}} F(y) dy \triangleq \mathrm{LB} \tag{62}$$

with equality if $p(y, m) = F(y)\ \forall y \in \Omega_m$

Extension to M-ary modulation/message

- optimal decision regions

$$\Omega_m = \{y : p(y,m) = F(y)\} \\ = \{y : p(y,m) \geq p(y,m') \; \forall m' \neq m\} \quad (63)$$

(ties are broken arbitrarily)

- and **optimal decisions = MAP rule** (same as before)

$$\widehat{m}(y) = D\{y\} = m' \text{ if } y \in \Omega_{m'} \quad (64)$$

$$= \arg\max_m p(y,m) = \arg\max_m P(m|y) \quad (65)$$

- min. error probability

$$P_{e,min} = \mathrm{LB} = 1 - \int_{\mathbb{R}} F(y)dy = 1 - \sum_m \int_{\Omega_m} F(y)dy \quad (66)$$

- **Q:** prove all of the above, especially UB/LB in (52)(62)

Sufficient statistics

- key idea: through away unnecessary data
- numerous applications
  - communication/information theory
  - signal processing
  - stochastic control
  - optimal decisions under uncertainty
  - ML/AI/big data/search engines

Sufficient statistics

- key idea: through away unnecessary data
- if $p(\mathbf{r}_2|\mathbf{r}_1, m) = p(\mathbf{r}_2|\mathbf{r}_1)$, then

$$\arg\max_m p(\mathbf{r}_1, \mathbf{r}_2|m)p(m) = \arg\max_m p(\mathbf{r}_1|m)p(m) \tag{67}$$

  i.e. $\mathbf{r}_2$ is irrelevant data
- $\mathbf{r}_1$ = sufficient statistics
- optimal Rx is using $\mathbf{r}_1$ only
- Markov chain (MC): $m \rightarrow \mathbf{r}_1 \rightarrow \mathbf{r}_2$
- can significantly simplify Rx!

Summary

- optimal decision rule: minimize P_e
- key: Rx via decision regions
- MAP decision rule
- ML decision rule, LR/LLR
- examples
- sufficient statistics (simple but optimal Rx)

Reading

- S. Haykin, Digital Communication Systems, Wiley, 2014. Ch. 3.11–3.15
- J.M. Wozencraft, I.M. Jacobs, Principles of Communication Engineering, Wiley, 1965. Ch. 2 (especially "A Communication Example"), Ch. 4.1–4.3
- H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, 2001 (original 1968). Ch. 2.1–2.3.
- A. Lapidoth, A Foundation in Digital Communication, Cambridge University Press, 2017. Ch. 20.1-20.10 (the latter section is especially relevant); Ch. 21.1–21.3.