Binary Hypothesis Testing, Multiple Hypothesis Testing

Course: Foundations in Digital Communications

Tobias Oechtering
presented by: **Ragnar Thobaben**

Royal Institute of Technology (KTH),
School of EE and ACCESS Center,
Communication Theory Lab
Stockholm, Sweden

4th lecture

What did we do last lecture?

Outline - Motivation

- Communication means transmitting bits over noisy channels. How should a decoder decide on what has been transmitted? What are the basic principles? What is optimal?
  - Binary Hypothesis Testing (chap 20)
- How do these concepts extend if the message is more than a bit?
  - Multiple Hypothesis Testing (chap 21)

Motivation

- Task in digital communication is to communicate information
  - The receiver has only access to the *received* waveform, which is typically distorted!
  - Need to find strategy to recover information, how to guess intelligently!
- In communication this task is called **decoding**, in statistics it is called **hypothesis testing**, and A. Lapidoth calls it **guessing**.

- In real-world applications the channel output is a continuous-time waveform and many bits should be transmitted.
- To **explain the principles** we first restrict our attention to
  - binary hypothesis testing - two alternatives
  - observations are vectors or scalars

Problem Formulation: Binary Hypothesis Testing

- Two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$: RV H takes 0 or 1.
  - Bayesian setting, i.e, **prior probabilities** are known:

$$\pi_0 = \mathbb{P}[H = 0] \qquad \pi_1 = \mathbb{P}[H = 1]$$

  - Assumption: $\pi_0 = \pi_1 = 1/2$ (maximum information)
- Receiver has **observation** $\boldsymbol{Y}$, a random vector in $\mathbb{R}^d$
  - $f_{\boldsymbol{Y}|H}(\cdot|\cdot)$ denotes the statistical dependency between $\boldsymbol{Y}$ and H

Problem: Find decision rule for guessing H based on $\boldsymbol{Y}$!

$$\Phi_{guess} : \mathbb{R}^d \to \{0, 1\}$$

- **Goal:** Minimize the probability of receiver decoding error!
  - Probability of error: $P_e \triangleq \mathbb{P}\left[\Phi_{guess}(\boldsymbol{Y}) \neq H\right]$
  - **Optimal** decision rule if no other attains smaller P_e (optimal: P_e^*).

Q: What is a good guess on H in the absence of observables?

Guessing after Observing $\boldsymbol{Y}$

- **A posteriori distribution**: $\mathbb{P}\left[H = i | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right]$
  - For mathematical consistency define
    $$\mathbb{P}\left[H = i | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right] = \begin{cases} \frac{\pi_i f_{\boldsymbol{Y}|H}(\boldsymbol{y}_{obs}|i)}{f_{\boldsymbol{Y}}(\boldsymbol{y}_{obs})} & \text{if } f_{\boldsymbol{Y}}(\boldsymbol{y}_{obs}) > 0, \\ 0.5 & \text{otherwise.} \end{cases} \quad i = 0, 1.$$
  - It denotes the probability of hypothesis i after observing $\boldsymbol{y}_{obs}$.

$\Rightarrow$ **Optimal decision rule** (how we resolve ties is arbitrary):

$$\Phi^*_{guess}(\boldsymbol{y}_{obs}) = \begin{cases} 0 & \text{if } \mathbb{P}\left[H = 0 | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right] \geq \mathbb{P}\left[H = 1 | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right] \\ 1 & \text{otherwise} \end{cases}$$
$$= \begin{cases} 0 & \text{if } \pi_0 f_{\boldsymbol{Y}|H}(\boldsymbol{y}_{obs}|0) \geq \pi_1 f_{\boldsymbol{Y}|H}(\boldsymbol{y}_{obs}|1) \\ 1 & \text{otherwise} \end{cases}$$

- **Probability of error** of the optimal decision rule

$$P^*_e(\boldsymbol{y}_{obs}) = \min\left\{\mathbb{P}\left[H = 0 | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right], \mathbb{P}\left[H = 1 | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right]\right\}$$

Probability of Error

- Define **decision region**

$$\mathcal{D} \triangleq \{\boldsymbol{y} \in \mathbb{R}^d : \Phi_{guess}(\boldsymbol{y}) = 0\}$$

then the probability of error is given by

$$\mathbb{P}\left[error|H = 0\right] = \int_{\boldsymbol{y} \neq \mathcal{D}} f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0)\mathrm{d}\boldsymbol{y}$$

$$\mathbb{P}\left[error|H = 1\right] = \int_{\boldsymbol{y} = \mathcal{D}} f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1)\mathrm{d}\boldsymbol{y}$$

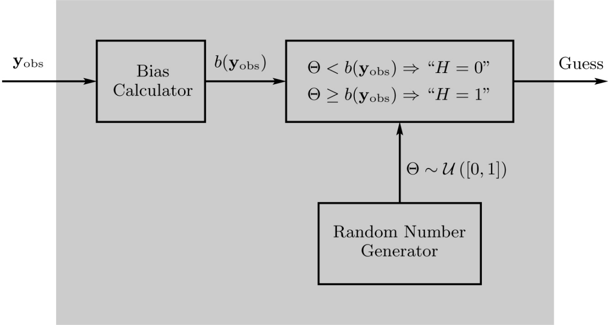
- Unconditional **probability of error**

$$P_e = \int_{\mathbb{R}^d} \Big(\mathrm{I}\left\{\boldsymbol{y} \in \mathcal{D}\right\} f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1)\pi_1 + \mathrm{I}\left\{\boldsymbol{y} \notin \mathcal{D}\right\} f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0)\pi_0\Big)\,\mathrm{d}\boldsymbol{y}$$

$$\geq \int_{\mathbb{R}^d} \min\Big\{f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1)\pi_1, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0)\pi_0\Big\}\,\mathrm{d}\boldsymbol{y} = \mathbb{E}\left[P_e^*(\boldsymbol{Y}_{obs})\right]$$

the RHS is achieved which proves optimality.

Randomized Decision Rule



- Randomized decision rule does not help, since

$$P_e(\boldsymbol{y}_{obs}) = b(\boldsymbol{y}_{obs})\mathbb{P}\left[H = 1|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right] + (1 - b(\boldsymbol{y}_{obs}))\mathbb{P}\left[H = 0|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right]$$
$$\geq \min\left\{\mathbb{P}\left[H = 0|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right], \mathbb{P}\left[H = 1|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right]\right\}$$

Maximum A Posteriori (MAP) Decision Rule

MAP Decision Rule

$$\Phi_{MAP}(\boldsymbol{y}_{obs}) \triangleq \begin{cases} 0 & \text{if } \pi_0 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) > \pi_1 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \\ 1 & \text{if } \pi_0 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) < \pi_1 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \\ \mathcal{U}(\{0,1\}) & \text{if } \pi_0 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) = \pi_1 f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \end{cases}$$

Identical to the previous, expect how it resolves ties.

- MAP decision is often rewritten as threshold test using
  - **likelihood-ratio function** $LR(\boldsymbol{y}) \triangleq \frac{f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0)}{f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1)}$
    - $LR : \mathbb{R}^d \to [0, \infty]$ with convention $\frac{\alpha}{0} = \infty$, $\alpha > 0$, $\frac{0}{0} = 1$
    - threshold $\frac{\pi_1}{\pi_0}$,
  - **log likelihood-ratio function** $LLR(\boldsymbol{y}) \triangleq \ln \frac{f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0)}{f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1)}$
    - $LLR : \mathbb{R}^d \to [-\infty, \infty]$ with conv. $\ln \frac{\alpha}{0} = \infty$, $\ln \frac{0}{\alpha} = -\infty, \alpha > 0$, $\ln \frac{0}{0} = 0$
    - threshold $\ln \frac{\pi_1}{\pi_0}$,

Maximum-Likelihood (ML) Decision Rule

- Different decision rule which ignores the prior:

ML decision rule

$$\Phi_{ML}(\boldsymbol{y}_{obs}) \triangleq \begin{cases} 0 & \text{if } f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) > f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \\ 1 & \text{if } f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) < f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \\ \mathcal{U}(\{0,1\}) & \text{if } f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|0) = f_{\boldsymbol{y}|H}(\boldsymbol{y}_{obs}|1), \end{cases}$$

- ML rule is suboptimal unless H is a priori uniformly distributed.
- Can be also rewritten as threshold tests using $LR(\cdot)$ or $LLR(\cdot)$
  - ML thresholds are 1 and 0

Bhattacharyya Bound

Using $\min\{a,b\} \leq \sqrt{ab}$ and $\sqrt{ab} \leq \frac{a+b}{2}$, $a,b \geq 0$ we have

$$\begin{aligned} P_e^* &= \int_{\mathbb{R}^d} \min\left\{ f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1)\pi_1, f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0)\pi_0 \right\} \mathrm{d}\boldsymbol{y} \\ &\leq \int_{\mathbb{R}^d} \sqrt{f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1)\pi_1 f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0)\pi_0 \mathrm{d}\boldsymbol{y}} \\ &= \sqrt{\pi_0\pi_1} \int_{\mathbb{R}^d} \sqrt{f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1) f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0) \mathrm{d}\boldsymbol{y}} \\ &\leq \frac{1}{2} \int_{\mathbb{R}^d} \sqrt{f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1) f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0) \mathrm{d}\boldsymbol{y}} \end{aligned}$$

Bhattacharyya Bound

$$P_e^* \leq \frac{1}{2} \int_{\mathbb{R}^d} \sqrt{f_{\boldsymbol{Y}|H}(\boldsymbol{y}|1) f_{\boldsymbol{Y}|H}(\boldsymbol{y}|0) \mathrm{d}\boldsymbol{y}}$$

Conditional Independence

- RVs X and Y are said to be **independent** if we have

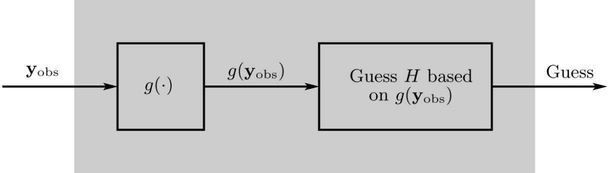
$$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

- RVs X and Y are said to be **conditional independent** given RV Z if we have

$$P_{X,Y,Z}(x,y,z) = P_{X|Z}(x|z)P_{Y|Z}(y|z)P_Z(z), \quad P_Z(z) > 0$$

  - Notation: $X - Z - Y$ known as *Markov chain*
  - Equivalently:
    - $P_{X|YZ}(x|y,z) = P_{X|Z}(x|z)$, $P_{YZ}(y,z) > 0$.
    - $P_{Y|XZ}(y|x,z) = P_{Y|Z}(y|z)$, $P_{XZ}(x,z) > 0$.

Processing



- **Processing:** Z is the result of processing Y with respect to H if H and Z are conditionally independent given Y.
- **Processing is Futile:** If Z is the result of processing Y with respect to H, then no decision rule based on Z can outperform an optimal guessing rule base on Y.

- The concept of **sufficient statistic** denotes the outcome of processing (mappings) $T : \mathbb{R}^d \to \mathbb{R}^{d'}$ where the optimal performance is still achievable for every $\boldsymbol{y}_{obs} \in \mathbb{R}^d$.

Guessing in the Presence of a Random Parameter

- The extension is conceptually straightforward - one has to distinguish between the following two cases:

1. **Random parameter Θ not observed:**
   - conditional density $f_{Y|H}(\boldsymbol{y}_{obs}|0) = \int_\theta f_{Y\Theta|H}(\boldsymbol{y}_{obs}, \theta|0)\mathrm{d}\theta$
   - Likelihood ratio: $LR(\boldsymbol{y}_{obs}) = \frac{\int_\theta f_{Y\Theta|H}(\boldsymbol{y}_{obs},\theta|0)\mathrm{d}\theta}{\int_\theta f_{Y\Theta|H}(\boldsymbol{y}_{obs},\theta|1)\mathrm{d}\theta}$
2. **Random parameter Θ observed:**
   - Treat observed random parameter $\Theta = \theta_{obs}$ as input
   - Likelihood ratio: $LR(\boldsymbol{y}_{obs}, \theta_{obs}) = \frac{f_{Y\Theta|H}(\boldsymbol{y}_{obs},\theta_{obs}|0)}{f_{Y\Theta|H}(\boldsymbol{y}_{obs},\theta_{obs}|1)}$

Let's take a break!

Multiple Hypothesis Testing

- Instead of (two) hypothesis let's have **RV** $M \in \mathcal{M}$ **(messages)**.
  - Prior $\pi_m = \mathbb{P}[M = m]$, $m \in \mathcal{M}$.
  - Non-degenerate prior if $\pi_m > 0$ for $m \in \mathcal{M}$.
- **Observation:** RV $\boldsymbol{Y} \in \mathbb{R}^d$ with $f_{\boldsymbol{Y}|M}(\cdot|m)$
- **Decision rule:** $\Phi_{guess} : \mathbb{R}^d \to \mathcal{M}$
- **Error probability:** $P_e = \mathbb{P}\left[\phi_{guess}(\boldsymbol{Y}) \neq M\right]$
- A guessing (decision) rule is **optimal** if no other rule achieves a lower P_e. Denote the optimal error probability P_e^*.
- For mathematical consistency define **a posteriori distribution**:

$$\mathbb{P}\left[M = m|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right] = \begin{cases} \frac{\pi_m f_{\boldsymbol{Y}|M}(\boldsymbol{y}_{obs}|m)}{f_{\boldsymbol{Y}}(\boldsymbol{y}_{obs})} & \text{if } f_{\boldsymbol{Y}}(\boldsymbol{y}_{obs}) > 0, \\ 1/|\mathcal{M}| & \text{otherwise.} \end{cases}$$

Guessing after Observing $\boldsymbol{Y}$

- Guess $\tilde{m}$ which leads to the highest *a posteriori* probability.
- Success prob.: $\mathbb{P}\left[correct|\boldsymbol{Y}=\boldsymbol{y}_{obs}\right]=\max_{m'\in\mathcal{M}}\left\{\mathbb{P}\left[M=m'|\boldsymbol{Y}=\boldsymbol{y}_{obs}\right]\right\}$
- Error probability: $P_e^*(\boldsymbol{y}_{obs})=1-\max_{m'\in\mathcal{M}}\left\{\mathbb{P}\left[M=m'|\boldsymbol{Y}=\boldsymbol{y}_{obs}\right]\right\}$
- Define outcomes of maximal a posteriori probability

$$\mathcal{M}(\boldsymbol{y}_{obs})=\left\{\tilde{m}\in\mathcal{M}:\pi_{\tilde{m}}f_{\boldsymbol{Y}|M}(\boldsymbol{y}_{obs})=\max_{m'\in\mathcal{M}}\left\{\pi_{m'}f_{\boldsymbol{Y}|M}(\boldsymbol{y}_{obs})\right\}\right\}$$

Optimal Multi-hypothesis Testing

Any guessing rule that satisfies the following is optimal

$$\Phi^*_{guess}(\boldsymbol{y}_{obs})\in\mathcal{M}(\boldsymbol{y}_{obs}),\qquad \boldsymbol{y}_{obs}\in\mathbb{R}^d.$$

Proof

- Every deterministic decision rule results in a partition $\{\mathcal{D}_m\}$ of $\mathbb{R}^d$
$$\bigcup_{m\in\mathcal{M}} \mathcal{D}_m = \mathbb{R}^d \qquad \mathcal{D}_m \cap \mathcal{D}_{m'} = \emptyset$$
where $\mathcal{D}_m$ covers observations $\boldsymbol{y}_{obs}$ leading to guess m (and vice versa).
- Searching for an optimal decision rule is equivalent to searching for an optimal way to partition $\mathbb{R}^d$

$$\mathbb{P}\left[correct\right] = \sum_{m\in\mathcal{M}} \pi_m \mathbb{P}\left[correct|M = m\right] = \sum_{m\in\mathcal{M}} \pi_m \int_{\mathcal{D}_m} f_{\boldsymbol{Y}|M}(\boldsymbol{y}|m)\mathrm{d}\boldsymbol{y}$$
$$= \int_{\mathbb{R}^d} \left(\sum_{m\in\mathcal{M}} \pi_m f_{\boldsymbol{Y}|M}(\boldsymbol{y}|m) \mathrm{I}\left\{\boldsymbol{y} \in \mathcal{D}_m\right\} \right) \mathrm{d}\boldsymbol{y}$$

- The integral will be maximized if we assign $\boldsymbol{y}$ to the set $\mathcal{D}_{\tilde{m}}$ with $\tilde{m} \in \mathcal{M}(\boldsymbol{y})$. □

Example: Multi-Hypothesis Testing for 2D Signals

- M is uniformly distributed over $\mathcal{M}$
- 2D-observations $\boldsymbol{Y}$

$$f_{Y^{(1)}Y^{(2)}|M}(y^{(1)}, y^{(2)}|m) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y^{(1)} - a_m)^2 + (y^{(2)} - b_m)^2}{2\sigma^2}\right)$$

- ML rule: “Nearest-Neighbor” decoding rule

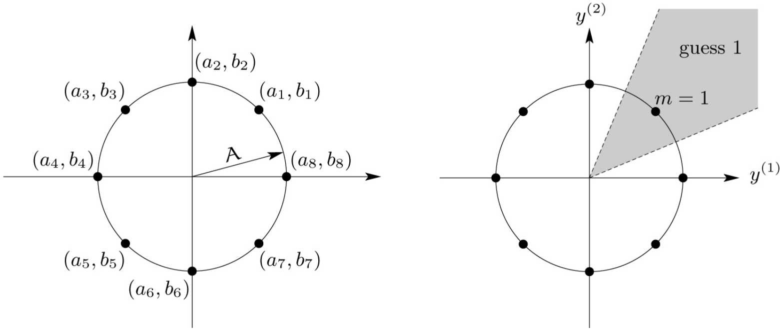
$$f_{Y^{(1)}Y^{(2)}|M}(y^{(1)}, y^{(2)}|\tilde{m}) = \max_{m' \in \mathcal{M}} \left\{ f_{Y^{(1)}Y^{(2)}|M}(y^{(1)}, y^{(2)}|m') \right\}$$

$$\Leftrightarrow \quad \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y^{(1)}-a_{\tilde{m}})^2+(y^{(2)}-b_{\tilde{m}})^2}{2\sigma^2}\right) = \max_{m' \in \mathcal{M}} \left\{ \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y^{(1)}-a_{m'})^2+(y^{(2)}-b_{m'})^2}{2\sigma^2}\right) \right\}$$

$$\Leftrightarrow \quad (y^{(1)} - a_{\tilde{m}})^2 + (y^{(2)} - b_{\tilde{m}})^2 = \min_{m' \in \mathcal{M}} (y^{(1)} - a_{m'})^2 + (y^{(2)} - b_{m'})^2$$

- The last equation denotes the *nearest neighbor* decoding rule.

Example: Decision region 8-PSK



- Shaded region denotes ML decision region $\mathcal{D}_1$

Union-of-Events Bound

- Given two not necessarily disjoint events $\mathcal{V}$ and $\mathcal{W}$

$$\mathbb{P}[\mathcal{V} \cup \mathcal{W}] = \mathbb{P}[\mathcal{V}] + \mathbb{P}[\mathcal{W}] - \mathbb{P}[\mathcal{V} \cap \mathcal{W}] \leq \mathbb{P}[\mathcal{V}] + \mathbb{P}[\mathcal{W}]$$

$\Rightarrow$ **Union-of-events bound:** $\mathbb{P}\left[\bigcup_j \mathcal{V}_j\right] \leq \sum_j \mathbb{P}\left[\mathcal{V}_j\right]$

- The Union bound can be used to derive upper bounds on the error analysis
  - $P_{MAP}(error|M = m) \leq \mathbb{P}\left[\boldsymbol{Y} \in \bigcup_{m' \neq m} \mathcal{B}_{m,m'} | M = m\right]$
    - $\mathcal{B}_{m,m'} = \{\boldsymbol{y} \in \mathbb{R}^d : \pi_{m'} f_{Y|M}(\boldsymbol{y}|m') \geq \pi_m f_{Y|M}(\boldsymbol{y}|m)\}$
  - to obtain a Union-Bhattacharyya bound

$$p_e^* \leq \frac{1}{2|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{m \neq m'} \int \sqrt{f_{Y|M}(\boldsymbol{y}|m) f_{Y|M}(\boldsymbol{y}|m')} \mathrm{d}\boldsymbol{y}$$

  - etc.

Outlook - Assignment

- Binary Hypothesis Testing
- Multiple Hypothesis Testing

Next lecture

Sufficient statistics, multivariate Gaussian distribution

- Reading Assignment: Chap 22-23
- Homework:
  - Problems in textbook: Exercise 20.1, 20.2, 20.3, 20.4, 20.13, 21.2, 21.4, and 21.8
  - Deadline: Dec 2