ECE531 Lecture 5: Composite Hypothesis Testing

D. Richard Brown III

Worcester Polytechnic Institute

19-February-2009

Introduction

Suppose we have the following discrete-time signal detection problem:

$$
\begin{aligned}
\text{state } x_0 \quad &\leftrightarrow \quad \text{transmitter sends } s_{0,i} = 0 \\
\text{state } x_1 \quad &\leftrightarrow \quad \text{transmitter sends } s_{1,i} = a\sin(2\pi\omega k + \phi)
\end{aligned}
$$

for sample index $i = 0, \dots, n-1$. As in Lecture 4, we observe these signals in additive noise

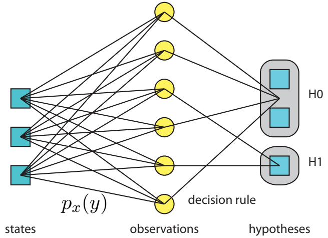
$$
Y_i \quad = \quad s_{j,i} + W_i
$$

for sample index $i = 0, \dots, n-1$ and $j \in \{0, 1\}$ and we wish to optimally decide $\mathcal{H}_0 \leftrightarrow x_0$ versus $\mathcal{H}_1 \leftrightarrow x_1$.

- We know how to do this if a, ω, and ϕ are known: **simple binary hypothesis testing problem** (Bayes, minimax, N-P).
- What if one or more of these parameters is unknown? We will have a **composite binary hypothesis testing problem**.

Introduction

Recall that **simple** hypothesis testing problems have one state per hypothesis. **Composite** hypothesis testing problems have at least one hypothesis containing more than one state.



Example: Y is Gaussian distributed with known variance σ^2 but unknown mean $\mu \in \mathbb{R}$. Given an observation $Y = y$, we want to decide if $\mu > \mu_0$.

- Only two hypotheses here: $\mathcal{H}_0 : \mu \leq \mu_0$ and $\mathcal{H}_1 : \mu > \mu_0$.
- What is the state?
- This is a **binary composite** hypothesis testing problem with an uncountably infinite number of states.

Mathematical Model and Notation

- Let $\mathcal{X}$ be the set of states. $\mathcal{X}$ can now be finite, e.g. $\{0,1\}$, countably infinite, e.g. $\{0,1,\dots\}$, or uncountably infinite, e.g. $\mathbb{R}^2$.
- We still assume a finite number of hypotheses $\mathcal{H}_0,\dots,\mathcal{H}_{M-1}$ with $M<\infty$. As before, hypotheses are defined as a partition on $\mathcal{X}$.
  - Composite HT: At least one hypothesis contains more than one state.
  - If $\mathcal{X}$ is infinite (countably or uncountably), at least one hypothesis must contain an infinite number of states.
- Let $C^\top(x) = [C_{0,x},\dots,C_{M-1,x}]^\top$ be the vector of costs associated with making decision $i \in \mathcal{Z} = \{0,\dots,M-1\}$ when the state is x.
- Let $\mathcal{Y}$ be the set of observations.
- Each state $x \in \mathcal{X}$ has an associated conditional pmf or pdf $p_x(y)$ that specifies how states are mapped to observations. These conditional pmfs/pdfs are, as always, assumed to be known.

The Uniform Cost Assignment in Composite HT Problems

In general, for $i = 0, \dots, M-1$ and $x \in \mathcal{X}$, the UCA is given as

$$C_i(x) = \begin{cases} 0 & \text{if } x \in \mathcal{H}_i \\ 1 & \text{otherwise.} \end{cases}$$

Example: A random variable Y is Gaussian distributed with known variance σ^2 but unknown mean $\mu \in \mathbb{R}$. Given an observation $Y = y$, we want to decide if $\mu > \mu_0$. Hypotheses $\mathcal{H}_0 : \mu \leq \mu_0$ and $\mathcal{H}_1 : \mu > \mu_0$.

Uniform cost assignment:

$$C^\top(\mu) = [C_0(\mu), C_1(\mu)] = \begin{cases} [0,1] & \text{if } \mu \leq \mu_0 \\ [1,0] & \text{if } \mu > \mu_0 \end{cases}$$

Other cost structures, e.g. squared error, are possible in composite HT problems, of course.

Conditional Risk

- Given decision rule ρ, the conditional risk for state $x \in \mathcal{X}$ can be written as

$$
\begin{aligned}
R_x(\rho) &= \int_{\mathcal{Y}} \rho^\top(y) C(x) p_x(y)\, dy \\
&= \mathrm{E}[\rho^\top(Y) C(x) \,|\, \text{state is } x] \\
&= \mathrm{E}_x[\rho^\top(Y) C(x)]
\end{aligned}
$$

  where the last expression is a common shorthand notation used in many textbooks.

- Recall that the decision rule

$$
\rho^\top(y) = [\rho_0(y), \dots, \rho_{M-1}(y)]
$$

  where $0 \le \rho_i(y) \le 1$ specifies the probability of deciding $\mathcal{H}_i$ when you observe $Y = y$. Also recall that $\sum_i \rho_i(y) = 1$ for all $y \in \mathcal{Y}$.

Bayes Risk: Countable States

- When $\mathcal{X}$ is countable, we denote the prior probability of state x as π_x. The Bayes risk of decision rule ρ in this case can be written as

$$\begin{aligned} r(\rho,\pi) &= \sum_{x\in\mathcal{X}} \pi_x R_x(\rho) \\ &= \sum_{x\in\mathcal{X}} \pi_x \int_{\mathcal{Y}} \rho^\top(y) C(x) p_x(y)\, dy \\ &= \int_{\mathcal{Y}} \rho^\top(y) \underbrace{\left(\sum_{x\in\mathcal{X}} C(x)\pi_x p_x(y)\right)}_{g(y,\pi)=[g_0(y,\pi),\dots,g_{M-1}(y,\pi)]^\top} dy \end{aligned}$$

- How should we specify the Bayes decision rule $\rho(y) = \delta^{B\pi}(y)$ in this case?

Bayes Risk: Uncountable States

- When $\mathcal{X}$ is uncountable, we denote the prior density of the states as $\pi(x)$. The Bayes risk of decision rule ρ in this case can be written as

$$\begin{aligned} r(\rho,\pi) &= \int_{\mathcal{X}} \pi(x) R_x(\rho)\, dx \\ &= \int_{\mathcal{X}} \pi(x) \int_{\mathcal{Y}} \rho^\top(y) C(x) p_x(y)\, dy\, dx \\ &= \int_{\mathcal{Y}} \rho^\top(y) \underbrace{\left(\int_{\mathcal{X}} C(x)\pi(x) p_x(y)\, dx \right)}_{g(y,\pi)=[g_0(y,\pi),\dots,g_{M-1}(y,\pi)]^\top} dy \end{aligned}$$

- How should we specify the Bayes decision rule $\rho(y) = \delta^{B\pi}(y)$ in this case?

Bayes Decision Rules for Composite HT Problems

It doesn't matter whether we have a finite, countably infinite, or uncountably infinite number of states, we can always find a deterministic Bayes decision rule as

$$\delta^{B\pi}(y) = \arg \min_{i \in \{0,\dots,M-1\}} g_i(y,\pi)$$

where

$$g_i(y,\pi) = \begin{cases} \sum_{j=0}^{N-1} C_{i,j}\pi_j p_j(y) & \text{when } \mathcal{X} = \{x_0,\dots,x_{N-1}\} \text{ is finite} \\ \sum_{x \in \mathcal{X}} C_i(x)\pi_x p_x(y) & \text{when } \mathcal{X} \text{ is countably infinite} \\ \int_{\mathcal{X}} C_i(x)\pi(x)p_x(y)\,dx & \text{when } \mathcal{X} \text{ is uncountably infinite} \end{cases}$$

When we have only two hypotheses, we just have to compare $g_0(y,\pi)$ with $g_1(y,\pi)$ at each value of y. We decide $\mathcal{H}_1$ when $g_1(y,\pi) < g_0(y,\pi)$, otherwise we decide $\mathcal{H}_0$.

Composite Bayes Hypothesis Testing Example

Suppose $\mathcal{X} = [0,3)$ and we want to decide between three hypotheses

$$\begin{aligned}\mathcal{H}_0 &: 0 \le x < 1\\ \mathcal{H}_1 &: 1 \le x < 2\\ \mathcal{H}_2 &: 2 \le x < 3\end{aligned}$$

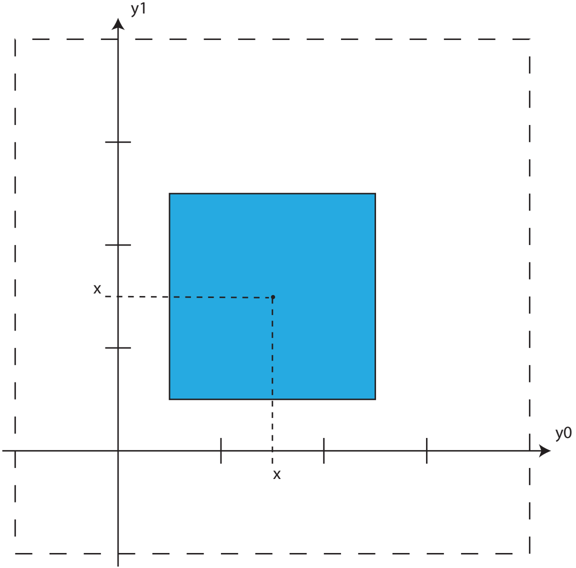
We get two observations $Y_0 = x + \eta_0$ and $Y_1 = x + \eta_1$, where η_0 and η_1 are i.i.d. and uniformly distributed on $[-1,1]$.

- What is the observation space $\mathcal{Y}$?
- What is the conditional distribution of each observation?

$$p_x(y_k) = \begin{cases} \frac{1}{2} & x-1 \le y_k \le x+1 \\ 0 & \text{otherwise} \end{cases}$$

- What is the conditional distribution of the vector observation?

Composite Bayes Hypothesis Testing Example



Composite Bayes Hypothesis Testing Example

Assume the UCA. What does our cost vector $C(x)$ look like?

Let's also assume a uniform prior on the states, i.e.

$$\pi(x) = \begin{cases} \frac{1}{3} & 0 \le x < 3 \\ 0 & \text{otherwise} \end{cases}$$

Do we have everything we need to find the Bayes decision rule? Yes!

Composite Bayes Hypothesis Testing Example

We want to find the index corresponding to the minimum of three discriminant functions:

$$\begin{aligned} g_0(y,\pi) &= \frac{1}{3}\int_1^3 p_x(y)\,dx \\ g_1(y,\pi) &= \frac{1}{3}\int_0^1 p_x(y)\,dx + \frac{1}{3}\int_2^3 p_x(y)\,dx \\ g_2(y,\pi) &= \frac{1}{3}\int_0^2 p_x(y)\,dx \end{aligned}$$

Note that

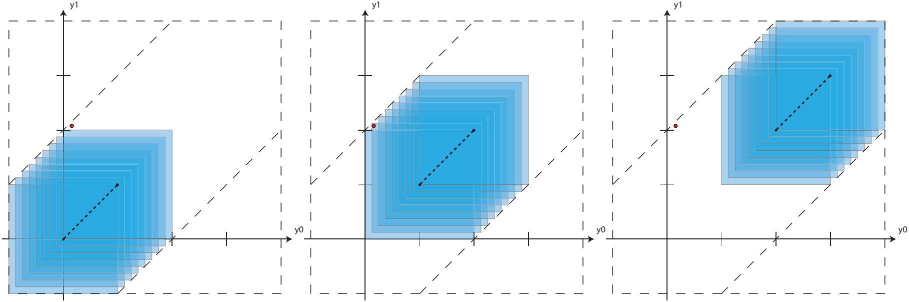
$$\arg \min_{i\in\{0,1,2\}} g_i(y,\pi) \quad \Leftrightarrow \quad \arg \max_{i\in\{0,1,2\}} h_i(y,\pi)$$

where

$$h_0(y,\pi) = \int_0^1 p_x(y)\,dx \quad h_1(y,\pi) = \int_1^2 p_x(y)\,dx \quad h_2(y,\pi) = \int_2^3 p_x(y)\,dx.$$

Composite Bayes Hypothesis Testing Example

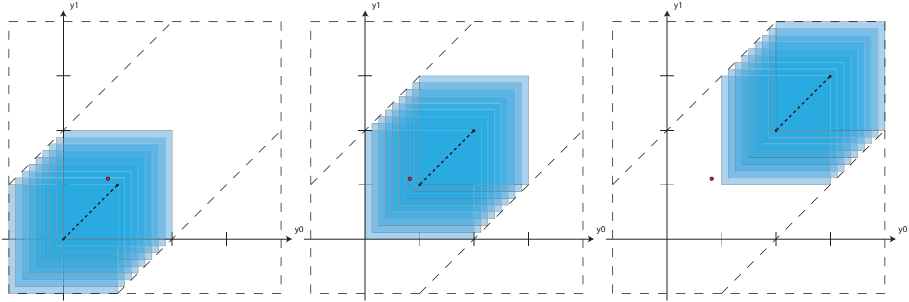
Note that $y \in \mathbb{R}^2$ is fixed and we are integrating with respect to $x \in \mathbb{R}$ here. Suppose $y = [0.2, 2.1]^\top$ (the red dot in the following 3 plots).



Which hypothesis must be true?

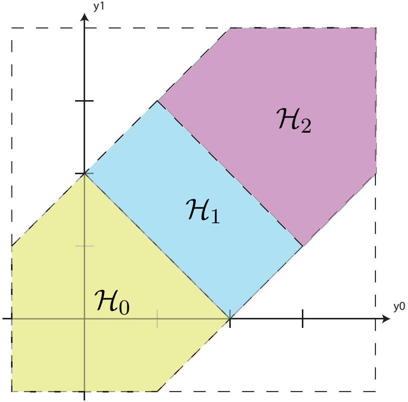
Composite Bayes Hypothesis Testing Example

Suppose $y = [0.8, 1.1]^\top$ (the red dot in the following 3 plots).



Which hypothesis would you choose? Hopefully not $\mathcal{H}_2$!

Composite Bayes Hypothesis Testing Example



$$\delta^{B\pi} = \begin{cases} 0 & \text{if } y_0 + y_1 \le 2 \\ 1 & \text{if } 2 < y_0 + y_1 \le 4 \\ 2 & \text{otherwise.} \end{cases}$$

Composite Neyman-Pearson Hypothesis Testing: Basics

When the problem is simple and binary, recall that

$$\begin{aligned} P_{\mathsf{fp}}(\rho) &:= \mathrm{Prob}(\rho \text{ decides } 1 \,|\, \text{state is } x_0) \\ P_{\mathsf{fn}}(\rho) &:= \mathrm{Prob}(\rho \text{ decides } 0 \,|\, \text{state is } x_1). \end{aligned}$$

The N-P decision rule is then given as

$$\begin{aligned} \rho^{\mathsf{NP}} &= \max_{\rho} P_D(\rho) \\ \text{subject to } P_{\mathsf{fp}}(\rho) &\le \alpha \end{aligned}$$

where $P_D := 1 - P_{\mathsf{fn}} = \mathrm{Prob}(\rho \text{ decides } 1 \,|\, \text{state is } x_1)$. The N-P criterion seeks a decision rule that **maximizes the probability of detection** subject to the constraint that the false positive probability $\le \alpha$.

- Now suppose the problem is binary but not simple. The states can be finite (more than 2), countably infinite, or uncountably infinite.
- How can we extend our ideas of N-P HT to composite hypotheses?

Composite Neyman-Pearson Hypothesis Testing

- Hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are associated with the subsets of states $\mathcal{X}_0$ and $\mathcal{X}_1$, where $\mathcal{X}_0$ and $\mathcal{X}_1$ form a partition on $\mathcal{X}$.
- Given a state $x \in \mathcal{X}_0$, we can write the **conditional false-positive probability** as

$$P_{\mathsf{fp},x}(\rho) \quad := \quad \mathrm{Prob}(\rho \text{ decides } 1 \,|\, \text{state is } x).$$

- Given a state $x \in \mathcal{X}_1$, we can write the **conditional false-negative probability** as

$$P_{\mathsf{fn},x}(\rho) \quad := \quad \mathrm{Prob}(\rho \text{ decides } 0 \,|\, \text{state is } x).$$

  or the **conditional probability of detection** as

$$P_{D,x}(\rho) := \mathrm{Prob}(\rho \text{ decides } 1 \,|\, \text{state is } x) = 1 - P_{\mathsf{fn},x}(\rho).$$

- The binary composite N-P HT problem is then

$$\rho^{\mathsf{NP}} \quad = \quad \arg\max_{\rho} P_{D,x}(\rho) \text{ for all } x \in \mathcal{X}_1$$

  subject to the constraint $P_{\mathsf{fp},x}(\rho) \leq \alpha$ for all $x \in \mathcal{X}_0$.

Composite Neyman-Pearson Hypothesis Testing

Remarks:

- Recall that the decision rule does not know the state. It only knows the observation : $\rho : \mathcal{Y} \mapsto \mathcal{P}_M$.
- We are trying to find one decision rule ρ that maximizes the probability of detection for every state $x \in \mathcal{X}_1$ subject to the constraints $P_{\mathsf{fp},x}(\rho) \leq \alpha$ for every $x \in \mathcal{X}_0$.
- The composite N-P HT problem is a **multi-objective** optimization problem.
- Uniformly most powerful (UMP) decision rule: When a solution exists for the binary composite N-P HT problem, the optimum decision rule ρ^{NP} is UMP (one decision rule gives the best $P_{D,x}$ for all $x \in \mathcal{X}_1$ subject to the constraints).
- Although UMP decision rules are very desirable, they only exist in some special cases.

Some Intuition about UMP Decision Rules

- Suppose we have a binary composite hypothesis testing problem where the null hypothesis $\mathcal{H}_0$ is simple and the alternative hypothesis is composite, i.e. i.e. $\mathcal{X}_0 = x_0 \in \mathcal{X}$ and $\mathcal{X}_1 = \mathcal{X} \backslash x_0$.
- Now pick a state $x_1 \in \mathcal{X}_1$ with conditional density $p_1(y)$ and associate a new hypothesis $\mathcal{H}_1{}'$ with just this one state.
- The problem of deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}'$ is a simple binary HT problem. The N-P lemma tells us that we can find a decision rule

$$\rho'(y) = \begin{cases} 1 & p_1(y) > v' p_0(y) \\ \gamma' & p_1(y) = v' p_0(y) \\ 0 & p_1(y) < v' p_0(y) \end{cases}$$

where v' and γ' are selected so that $P_{\mathsf{fp}} = \alpha$. No other α-level decision rule can be more powerful for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}'$.

Some Intuition about UMP Decision Rules (continued)

- Now pick another state $x_2 \in \mathcal{X}_1$ with conditional density $p_2(y)$ and associate a new hypothesis $\mathcal{H}_1{}''$ with just this one state.
- The problem of deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}''$ is also a simple binary HT problem. The N-P lemma tells us that we can find a decision rule

$$\rho''(y) = \begin{cases} 1 & p_2(y) > v'' p_0(y) \\ \gamma'' & p_2(y) = v'' p_0(y) \\ 0 & p_2(y) < v'' p_0(y) \end{cases}$$

  where v'' and γ'' are selected so that $P_{\text{fp}} = \alpha$. No other α-level decision rule can be more powerful for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}''$.
- The decision rule ρ' cannot be more powerful than ρ'' for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}''$. Likewise, the decision rule ρ'' cannot be more powerful than ρ' for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1{}'$.

Some Intuition about UMP Decision Rules (the payoff)

- When might ρ' and ρ'' have the same power for deciding between $\mathcal{H}_0/\mathcal{H}_1{}'$ and $\mathcal{H}_0/\mathcal{H}_1{}''$?
- Only when the "critical region" $\Gamma_1 = \{y \in \mathcal{Y} : \rho(y) \text{ decides } \mathcal{H}_1\}$ is the same (except possibly on a set of probability zero).
- In our example, we need

$$\{y \in \mathcal{Y} : p_1(y) > v' p_0(y)\} = \{y \in \mathcal{Y} : p_2(y) > v'' p_0(y)\}$$

- Actually, we need this to be true for all of the states $x \in \mathcal{X}_1$.

Theorem

A UMP decision rule exists for the α-level binary composite HT problem $\mathcal{H}_0 \leftrightarrow x_0$ versus $\mathcal{H}_1 \leftrightarrow \mathcal{X} \backslash x_0 = \mathcal{X}_1$ if and only if the critical region $\{y \in \mathcal{Y} : \rho(y) \text{ decides } \mathcal{H}_1'\}$ of the simple binary HT problem $\mathcal{H}_0 \leftrightarrow x_0$ versus $\mathcal{H}_1' \leftrightarrow x_1$ is the same for all $x_1 \in \mathcal{X}_1$.

Example

Suppose we have the composite binary HT problem

$$\begin{aligned}\mathcal{H}_0 &: x = \mu_0 \\ \mathcal{H}_1 &: x > \mu_0\end{aligned}$$

for $\mathcal{X} = [\mu_0, \infty)$ and $Y \sim \mathcal{N}(x, \sigma^2)$.
Pick $\mu_1 > \mu_0$. We already know how to find the α-level N-P decision rule for the simple binary problem

$$\begin{aligned}\mathcal{H}_0 &: x = \mu_0 \\ \mathcal{H}_1' &: x = \mu_1\end{aligned}$$

The N-P decision rule for $\mathcal{H}_0$ versus $\mathcal{H}_1'$ is simply

$$\rho^{\mathsf{NP}}(y) = \begin{cases} 1 & y > \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 \\ \gamma & y = \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 \\ 0 & y < \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 \end{cases} \tag{1}$$

where γ is arbitrary since $P_{\mathsf{fp}} = \mathrm{Prob}(y > \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 \,|\, x = \mu_0) = \alpha$.

Example

Note that the critical region $\{y \in \mathcal{Y} : y > \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0\}$ does not depend on μ_1. Thus, by the theorem, (1) is a UMP decision rule for this binary composite HT problem.

Also note that the probability of detection for this problem can be expressed as

$$
\begin{aligned}
P_D(\rho^{\mathsf{NP}}) &= \mathrm{Prob}\left(y > \sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 \,|\, x = \mu_1\right) \\
&= Q\left(\frac{\sqrt{2}\sigma \mathsf{erfc}^{-1}(2\alpha) + \mu_0 - \mu_1}{\sigma}\right)
\end{aligned}
$$

does depend on μ_1, as you might expect. But there is no α-level decision rule that gives a better probability of detection (power) than ρ^{NP} for any state $x \in \mathcal{X}_1$.

Example

What if we had this composite HT problem instead?

$$\begin{aligned}\mathcal{H}_0 &: x = \mu_0 \\ \mathcal{H}_1 &: x \neq \mu_0\end{aligned}$$

for $\mathcal{X} = \mathbb{R}$ and $Y \sim \mathcal{N}(x, \sigma^2)$.

- For $\mu_1 < \mu_0$, the most powerful critical region for an α-level decision rule is $\{y \in \mathcal{Y} : y < \mu_0 - \sqrt{2}\sigma \mathrm{erfc}^{-1}(2\alpha)\}$ Does not depend on μ_1. Excellent!
- Hold on. This is not the same critical region as when $\mu_1 > \mu_0$. So, in fact, the critical region does depend on μ_1.
- We can conclude that no UMP decision rule exists for this binary composite HT problem.

The theorem is handy for both finding UMP decision rules and also showing when UMP decision rules can't be found.

Monotone Likelihood Ratio

Despite UMP decision rules only being available in "special cases", lots of real problems fall into these special cases. One such class of problems is the class of binary composite HT problems with monotone likelihood ratios.

Definition (Lehmann, Testing Statistical Hypotheses, p.78)

The real-parameter family of densities $p_x(y)$ for $x \in \mathcal{X} \subset \mathbb{R}$ is said to have **monotone likelihood ratio** if there exists a real-valued function $T(y)$ that only depends on y such that, for any $x_1 \in \mathcal{X}$ and $x_0 \in \mathcal{X}$ with $x_0 < x_1$, the distributions $p_{x_0}(y)$ and $p_{x_1}(y)$ are distinct and the ratio

$$L_{x_1/x_0}(y) := \frac{p_{x_1}(y)}{p_{x_0}(y)} = f(x_0, x_1, T(y))$$

is a non-decreasing function of $T(y)$.

Monotone Likelihood Ratio: Example

Consider the Laplacian density $p_x(y) = \frac{b}{2}e^{-b|y-x|}$ with $b > 0$ and $\mathcal{X} = [0, \infty)$.
Let's see if this family of densities has monotone likelihood ratio...

$$L_{x_1/x_0}(y) := \frac{p_{x_1}(y)}{p_{x_0}(y)} = e^{b(|y-x_0|-|y-x_1|)}$$

Since $b > 0$, $L_{x_1/x_0}(y)$ will be non-decreasing in $T(y) = y$ provided that $|y - x_0| - |y - x_1|$ is non-decreasing in y for all $0 \le x_0 < x_1$. We can write

$$|y - x_0| - |y - x_1| = \begin{cases} x_0 - x_1 & y < x_0 \\ 2y - x_1 - x_0 & x_0 \le y \le x_1 \\ x_1 - x_0 & y > x_1. \end{cases}$$

Note that $x_0 - x_1 < 0$ is a constant and $x_1 - x_0 > 0$ is also a constant with respect to y. Hence we can see that the Laplacian family of densities $p_x(y) = \frac{b}{2}e^{-b|y-x|}$ with $b > 0$ and $\mathcal{X} = [0, \infty)$ is monotone in $T(y) = y$.

Existence of a UMP Decision Rule for Monotone LR

Theorem (Lehmann, Testing Statistical Hypotheses, p.78)

Let $\mathcal{X}$ be a subinterval of the real line. Fix $\lambda \in \mathcal{X}$ and define the hypotheses $\mathcal{H}_0 : x \in \mathcal{X}_0 = \{x \leq \lambda\}$ versus $\mathcal{H}_1 : x \in \mathcal{X}_1 = \{x > \lambda\}$. If the family of densities $p_x(y)$ are distinct for all $x \in \mathcal{X}$ and has monotone likelihood ratio in $T(y)$, then the decision rule

$$\rho(y) = \begin{cases} 1 & T(y) > \tau \\ \gamma & T(y) = \tau \\ 0 & T(y) < \tau \end{cases}$$

where τ and γ are selected so that $P_{\mathsf{fp},x=\lambda} = \alpha$ is UMP for testing $\mathcal{H}_0 : x \in \mathcal{X}_0$ versus $\mathcal{H}_1 : x \in \mathcal{X}_1$.

Existence of a UMP Decision Rule: Finite Real States

Corollary (Finite Real States)

Let $\mathcal{X} = \{x_0 < x_1 < \cdots < x_{N-1}\} \subset \mathbb{R}$ and partition $\mathcal{X}$ into two order preserving sets $\mathcal{X}_0 = \{x_0, \ldots, x_j\}$ and $\mathcal{X}_1 = \{x_{j+1}, \ldots, x_{N-1}\}$. Then the α-level N-P decision rule for deciding between the **simple** *hypotheses $\mathcal{H}_0 : x = x_j$ versus $\mathcal{H}_1 : x = x_{j+1}$ is a UMP α-level test for deciding between the composite hypotheses $\mathcal{H}_0 : x \in \mathcal{X}_0$ versus $\mathcal{H}_1 : x \in \mathcal{X}_1$.*

Intuition: Suppose $\mathcal{X}_0 = \{x_0, x_1\}$ and $\mathcal{X}_1 = \{x_2, x_3\}$. We can build a table of α-level N-P decision rules for the four simple binary HT problems:

	$\mathcal{H}_1 : x = x_2$	$\mathcal{H}_1 : x = x_3$
$\mathcal{H}_0 : x = x_0$	ρ_{02}^{NP}	ρ_{03}^{NP}
$\mathcal{H}_0 : x = x_1$	ρ_{12}^{NP}	ρ_{13}^{NP}

The theorem says that ρ_{12}^{NP} a UMP α-level decision rule for the composite HT problem. Why?

Coin Flipping Example

Suppose we have a coin with $\mathrm{Prob}(\text{heads}) = x$, where $0 \le x \le 1$ is the unknown state. We flip the coin n times and observe the number of heads $Y = y \in \{0, \dots, n\}$. We know that, conditioned on the state x, the observation Y is distributed as

$$p_x(y) \quad = \quad \binom{n}{y} x^y (1-x)^{n-y}.$$

Fix $0 < \lambda < 1$ and define our composite binary hypotheses as

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad Y \sim p_x(y) \text{ for } x \in [0, \lambda] = \mathcal{X}_0 \\ \mathcal{H}_1 \quad &: \quad Y \sim p_x(y) \text{ for } x \in (\lambda, 1] = \mathcal{X}_1 \end{aligned}$$

Does there exist a UMP decision rule with significance level α?

- Are $p_x(y)$ distinct for all $x \in \mathcal{X}$? Yes.
- Does this family of densities have monotone likelihood ratio in some test statistic $T(y)$? Let's check...

Coin Flipping Example

$$L_{x_1/x_0}(y) = \frac{\binom{n}{y} x_1^y (1-x_1)^{n-y}}{\binom{n}{y} x_0^y (1-x_0)^{n-y}} = \phi^y \left(\frac{1-x_1}{1-x_0}\right)^n$$

where $\phi := \frac{x_1(1-x_0)}{x_0(1-x_1)} > 1$ when $x_1 > x_0$. Hence this family of densities has monotone likelihood ratio in the test statistic $T(y) = y$.

We don't have a finite number of states. How can we find the UMP decision rule? According the the theorem, a UMP decision rule will be

$$\rho(y) = \begin{cases} 1 & y > \tau \\ \gamma & y = \tau \\ 0 & y < \tau \end{cases}$$

with τ and γ selected so that $P_{\mathsf{fp},x=\lambda} = \alpha$. We just have to find τ and γ.

Coin Flipping Example

To find τ, we can use our procedure from **simple** binary N-P hypothesis testing. We want to find the smallest value of τ such that

$$P_{\mathsf{fp},x=\lambda}(\delta^\tau) = \sum_{j>\tau} \underbrace{\mathrm{Prob}[y = j \,|\, x = \lambda]}_{p_\lambda(y)} = \sum_{j>\tau}^{n} \binom{n}{j} \lambda^j (1-\lambda)^{n-j} \le \alpha.$$

Once we find the appropriate value of τ, if $P_{\mathsf{fp},x=\lambda}(\delta^\tau) = \alpha$, then we are done. We can set γ to any arbitrary value in $[0,1]$ in this case.
If $P_{\mathsf{fp},x=\lambda}(\delta^\tau) < \alpha$, then we have to find γ such that $P_{\mathsf{fp},x=\lambda}(\rho) = \alpha$. The UMP decision rule is just the usual randomization

$$\rho \quad = \quad (1-\gamma)\delta^\tau + \gamma\delta^{\tau-\epsilon}$$

and, in this case,

$$\gamma = \frac{\alpha - \sum_{j=\tau+1}^{n} \binom{n}{j} \lambda^j (1-\lambda)^{n-j}}{\binom{n}{\tau} \lambda^\tau (1-\lambda)^{n-\tau}}.$$

Power Function

For binary composite hypothesis testing problems with $\mathcal{X}$ a subinterval of the real line and a particular decision rule ρ, we can define the power function of ρ as

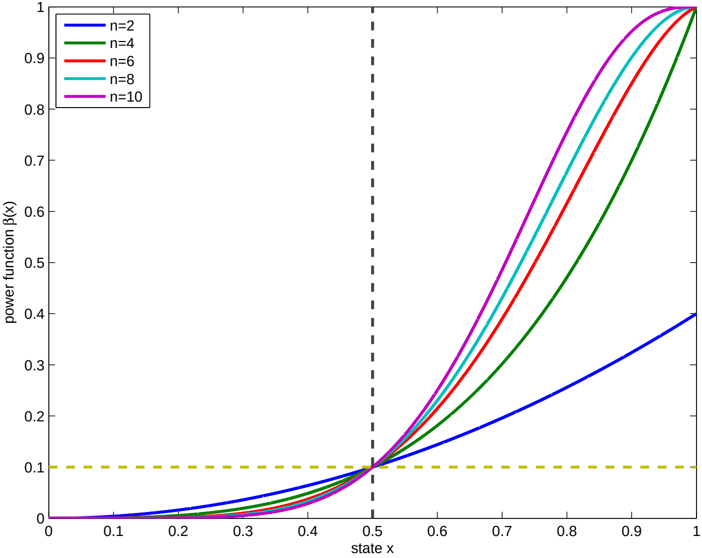
$$\beta(x) \quad := \quad \mathrm{Prob}(\rho \text{ decides } 1 \,|\, \text{state is } x)$$

- For each $x \in \mathcal{X}_1$, $\beta(x)$ is the probability of a true positive.
- For each $x \in \mathcal{X}_0$, $\beta(x)$ is the probability of a false positive.
- Hence, a plot of $\beta(x)$ versus x displays both the probability of a false positive and the probability of detection of ρ for all states $x \in \mathcal{X}$.

Our UMP decision rule for the coin flipping problem specifies that we always decide 1 when we observe $\tau + 1$ or more heads, and we decide 1 with probability γ if we observe τ heads. Hence,

$$\beta(x) \quad = \quad \sum_{j=\tau+1}^{n} \binom{n}{j} x^j (1-x)^{n-j} + \gamma \binom{n}{\tau} x^\tau (1-x)^{n-\tau}$$

Coin Flipping Example: Power Function $\lambda = 0.5$, $\alpha = 0.1$



Locally Most Powerful Decision Rules

- Although UMPs exists in many cases of practical interest, we've already seen that they don't always exist.
- If a UMP decision rule does not exist, an alternative is to seek the most powerful decision rule when x is very close to the threshold between $\mathcal{X}_0$ and $\mathcal{X}_1$.
- If such a decision rule exists, it is said to be a **locally most powerful** (LMP) decision rule.

Locally Most Powerful Decision Rules: Setup

- Consider the situation where $\mathcal{X} = [\lambda_0, \lambda_1] \subset \mathbb{R}$ for some $\lambda_1 > \lambda_0$. We wish to test $\mathcal{H}_0 : x = \lambda_0$ versus $\mathcal{H}_1 : \lambda_0 < x \le \lambda_1$ subject to $P_{\mathsf{fp},x=\lambda_0} \le \alpha$.
- For $x = \lambda_0$, the power function $\beta(x)$ corresponds to the probability of false positive; for $x > \lambda_0$, the power function $\beta(x)$ corresponds to the probability of detection.
- We assume that $\beta(x)$ is differentiable at $x = \lambda_0$. Then

$$\beta'(\lambda_0) = \frac{d}{dx}\beta(x)|_{x=\lambda_0}$$

  is the slope of $\beta(x)$ at $x = \lambda_0$.
- Taylor series expansion of $\beta(x)$ around $x = \lambda_0$

$$\beta(x) \quad = \quad \beta(\lambda_0) + \beta'(\lambda_0)(x - \lambda_0) + \dots$$

- Our goal is to find a decision rule ρ that maximizes the slope $\beta'(\lambda_0)$ subject to $\beta(\lambda_0) \le \alpha$. This decision rule is LMP of size α. Why?

Locally Most Powerful Decision Rules: Solution

Definition

A decision rule ρ^* is a locally most powerful decision rule of size α at $x = \lambda_0$ if, for any other size α decision rule ρ, $\beta'_{\rho^*}(\lambda_0) \geq \beta'_{\rho}(\lambda_0)$.

The LMP decision rule ρ takes the form

$$\rho(y) = \begin{cases} 1 & L'_{\lambda_0}(y) > \tau \\ \gamma & L'_{\lambda_0}(y) = \tau \\ 0 & L'_{\lambda_0}(y) < \tau \end{cases}$$

(see your textbook pp. 37-38 for the details) where

$$L'_{\lambda_0}(y) := \frac{d}{dx} L_{x/\lambda_0}(y)|_{x=\lambda_0} = \frac{\frac{d}{dx} p_x(y)|_{x=\lambda_0}}{p_{\lambda_0}(y)}$$

and where τ and γ are selected so that $P_{\mathsf{fp},x=\lambda_0} = \beta(\lambda_0) = \alpha$.

LMP Decision Rule Example

- Consider a Laplacian random variable Y with unknown mean $x \in [0, \infty)$. Given one observation of Y, we want to decide $\mathcal{H}_0 : x = 0$ versus $\mathcal{H}_1 : x > 0$ subject to an upper bound α on the false positive probability.
- The conditional density of Y is $p_x(y) = \frac{b}{2} e^{-b|y-x|}$.
- The likelihood ratio can be written as

$$L_{x/0}(y) = \frac{p_x(y)}{p_0(y)} = e^{b(|y|-|y-x|)}$$

- A UMP decision rule actually does exist here (left as an exercise).
- To find an LMP decision rule, we need to compute

$$L_0'(y) = \frac{d}{dx} L_{x/0}(y)|_{x=0^+}$$

LMP Decision Rule Example

$$\begin{aligned} L_0'(y) &= \frac{d}{dx} L_{x/0}(y)|_{x=0^+} \\ &= \frac{d}{dx} e^{b(|y|-|y-x|)}|_{x=0^+} \\ &= b\,\mathrm{sgn}(y) \end{aligned}$$

- Note that $L_0'(y)$ doesn't exist when $y = 0$, but $\mathrm{Prob}[Y = 0] = 0$ so this doesn't really matter.
- We need to find the decision threshold τ on $L_0'(y)$.
- Note that $L_0'(y)$ can take on only two values: $L_0'(y) \in \{-b, b\}$.
- Hence we only need to consider a finite number of decision thresholds $\tau \in \{-2b, -b, b, 2b\}$.
- $\tau = -2b$ corresponds to what decision rule? Always decide $\mathcal{H}_1$.
- $\tau = 2b$ corresponds to what decision rule? Always decide $\mathcal{H}_0$.

LMP Decision Rule Example

- When $x = 0$, $p_x(y) = \frac{b}{2}e^{-b|y|}$.
- When $x = 0$ we get a false positive with probability 1 if $b\text{sgn}(y) > \tau$ and with probability γ if $b\text{sgn}(y) = \tau$.
- False positive probability as a function of the threshold τ on $L'_0(y)$.

$\tau =$	$-2b$	$-b$	b	$2b$
$P_{\text{fp}} =$	1			0

- From this, it is easy to determine the required values for τ and γ as a function of α.

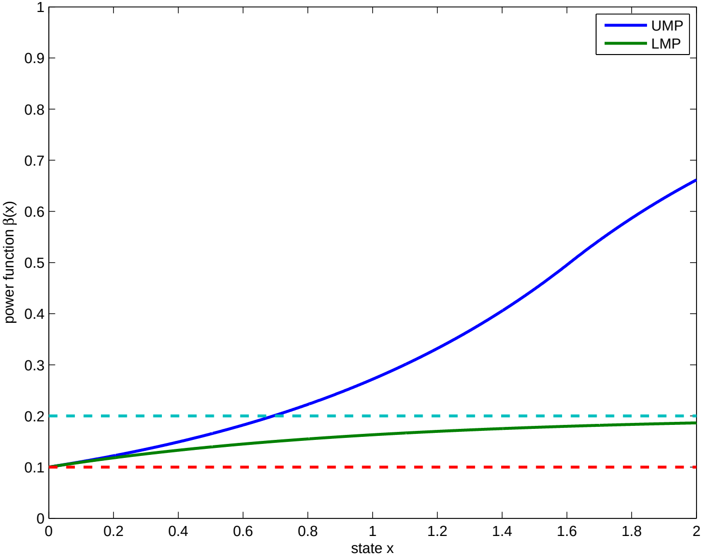
$\alpha = 0$	$\tau = 2b$	$\gamma =$ arbitrary
$0 < \alpha \leq \frac{1}{2}$	$\tau = b$	$\gamma = 2\alpha$
$\frac{1}{2} < \alpha < 1$	$\tau = -b$	$\gamma = 2\alpha - 1$
$\alpha = 1$	$\tau = -2b$	$\gamma =$ arbitrary

LMP Decision Rule Example

The power function of the LMP decision rule for $0 < \alpha < \frac{1}{2}$ is

$$\begin{aligned}
\beta(x) &= \mathrm{Prob}(\mathsf{sgn}(y) > 1 \,|\, \text{state is } x) + \gamma \mathrm{Prob}(\mathsf{sgn}(y) = 1 \,|\, \text{state is } x) \\
&= \gamma \int_0^\infty \frac{b}{2} e^{-b|y-x|}\, dy \\
&= 2\alpha \left(\int_0^x \frac{b}{2} e^{-b|y-x|}\, dy + \int_x^\infty \frac{b}{2} e^{-b|y-x|}\, dy \right) \\
&= 2\alpha \left(\frac{1 - e^{-bx}}{2} + \frac{1}{2} \right) \\
&= 2\alpha - \alpha e^{-bx}
\end{aligned}$$

LMP Decision Rule Example: $\alpha = 0.1$, $b = 1$



LMP Remarks

- Note that your textbook develops the notion of an LMP decision rule for the situation where $\mathcal{H}_0 : x = \lambda_0$ versus $\mathcal{H}_1 : \lambda_0 < x \leq \lambda_1$ for $\mathcal{X} = [\lambda_0, \lambda_1] \subset \mathbb{R}$ for some $\lambda_1 > \lambda_0$.
- The usefulness of the LMP decision rule in this case should be clear:
  - The slope of the power function $\beta(x)$ is maximized for states in the alternative hypothesis close to $x = \lambda_0$.
  - The false-positive probability is guaranteed to be less than or equal to α for all states in the null hypothesis (there is only one state in the null hypothesis).
- When we have the one-sided test $\mathcal{H}_0 : x \leq \lambda_0$ versus $\mathcal{H}_1 : x > \lambda_0$ with a composite null hypothesis, the same technique can be applied but you must check that $P_{\mathsf{fp},x} \leq \alpha$ for all $x \leq \lambda_0$.
  - The slope of the power function $\beta(x)$ is still maximized for states in the alternative hypothesis close to $x = \lambda_0$.
  - The false-positive probability of the LMP decision rule is not necessarily less than or equal to α for all states in the null hypothesis. You must check this.

Conclusions

- Simple hypothesis testing is not sufficient for detection problems with unknown parameters. These types of problems require **composite hypothesis testing**.
- Bayesian M-ary composite hypothesis testing.
  - Basically the same as Bayesian M-ary simple hypothesis testing: find the cheapest commodity (minimum discriminant function).
- Neyman Pearson binary composite hypothesis testing.
  - Multi-objective optimization problem.
  - Would like to find uniformly most powerful (UMP) decision rules. This is possible in some cases:
    - ⋆ Monotone likelihood ratio
    - ⋆ Other cases discussed in Lehmann "Testing Statistical Hypotheses"
  - Locally most powerful (LMP) decision rules can often be found in cases when UMP decision rules do not exist.