# ECE531 Lecture 2b: Neyman-Pearson Hypothesis Testing (Finite Number of Possible Observations)

D. Richard Brown III

Worcester Polytechnic Institute

26-January-2011

# Examples of real-world hypothesis testing problems

- To approve a new flu test, the FDA requires the test to have a **false positive** rate of no worse than 10% and a detection rate of at least 75%.
- After major bicycling races, many riders are tested for the presence of performance enhancing drugs. The false positive rate of these tests must be less than $x$% and the detection rate must be at least $y$%.
- False positives in radar systems: incoming airplane is detected as an enemy airplane when it is actually friendly. These false positives must occur with rate less than $x$%, and the detection rate must be maximized.

In many hypothesis testing problems, there is a **fundamental asymmetry** between the consequences of

- "false positive" (decide $\mathcal{H}_1$ when the true state is $x_0$) and
- "miss / false negative" (decide $\mathcal{H}_0$ when the true state is $x_1$).

# Neyman-Pearson Terminology

Neyman-Pearson hypothesis testing is always binary (simple or composite).

$$\begin{aligned}\mathcal{H}_0 &: \text{"null" hypothesis or "signal absent"} \\ \mathcal{H}_1 &: \text{"alternative" hypothesis or "signal present"}\end{aligned}$$

Common terminology for simple binary hypothesis testing:

- A "type I error" is when you decide $\mathcal{H}_1$ when the state is $x_0$. Also called a "false alarm" or "**false positive**".

$$R_0(D) = \mathrm{Prob}(\text{decide } \mathcal{H}_1 | \text{state is } x_0) = P_{\mathsf{fp}}(D)$$

- A "type II error" is when you decide $\mathcal{H}_0$ when the state is $x_1$. Also called a "miss" or "**false negative**".

$$R_1(D) = \mathrm{Prob}(\text{decide } \mathcal{H}_0 | \text{state is } x_1) = P_{\mathsf{fn}}(D)$$

- The "power" of a test is the probability of correctly deciding $\mathcal{H}_1$ when the state is $x_1$ or, in other words,

$$\text{power} = \mathrm{Prob}(\text{true positive}) = 1 - \mathrm{Prob}(\text{false negative})$$

The power of the test is also the probability of detecting the signal is present.

# The Neyman-Pearson Criterion

## Definition

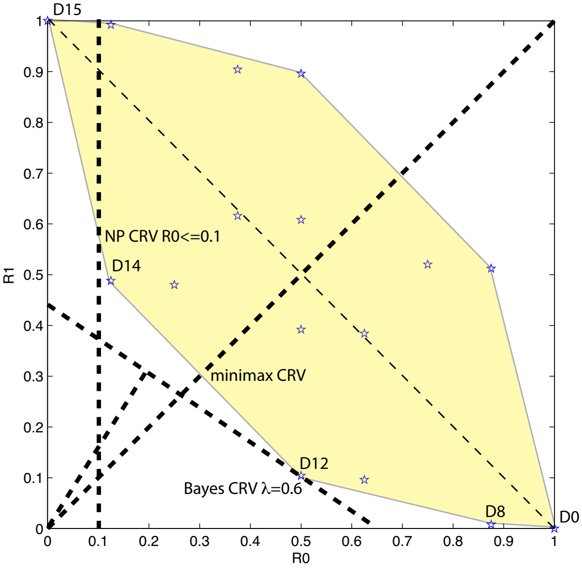
The Neyman-Pearson criterion decision rule is given as

$$\begin{aligned} D^{\mathsf{NP}} &= \arg\min_{D \in \mathcal{D}} P_{\mathsf{fn}}(D) \\ \text{subject to } P_{\mathsf{fp}}(D) &\leq \alpha \end{aligned}$$

where $\alpha \in [0, 1]$ is called the "significance level" of the test.

Note this is a "constrained optimization" problem.

# The N-P Criterion: 3 Coin Flips ($q_0 = 0.5$, $q_1 = 0.8$, $\alpha = 0.1$)



# Neyman-Pearson Hypothesis Testing Example

Coin flipping problem with a probability of heads of either $q_0 = 0.5$ or $q_1 = 0.8$. We observe three flips of the coin and count the number of heads. We can form our conditional probability matrix

$$P = \begin{bmatrix} 0.125 & 0.008 \\ 0.375 & 0.096 \\ 0.375 & 0.384 \\ 0.125 & 0.512 \end{bmatrix} \text{ where } P_{\ell j} = \mathrm{Prob}(\text{observe } \ell \text{ heads} | \text{state is } x_j).$$

Suppose we need a test with a significance level of $\alpha = 0.125$.

- What is the N-P decision rule in this case?
- What is the probability of correct detection if we use this N-P decision rule?

What happens if we relax the significance level to $\alpha = 0.5$?

# Intuition: The Hiker

You are going on a hike and you have a budget of $5 to buy food for the hike. The general store has the following food items for sale:

- One box of crackers: $1 and 60 calories
- One candy bar: $2 and 200 calories
- One bag of potato chips: $2 and 160 calories
- One bag of nuts: $3 and 270 calories

You would like to purchase the maximum calories subject to your $5 budget. What should you buy?

What if there were two candy bars available?

- The idea here is to rank the items by decreasing value (calories per dollar) and then purchase items with the most value until all the money is spent.
- The final purchase may only need to be a fraction of an item.

# N-P Hypothesis Testing With Discrete Observations

Basic idea:

- Sort the likelihood ratio $L_\ell = \frac{P_{\ell,1}}{P_{\ell,0}}$ by observation index in descending order. The order of $L$'s with the same value doesn't matter.
- Now pick $v$ to be the smallest value such that

$$P_{\mathsf{fp}} = \sum_{\ell : L_\ell > v} P_{\ell,0} \leq \alpha$$

- This defines a deterministic decision rule (binary HT notation)

$$\delta^v(y_\ell) = \begin{cases} 1 & L_\ell > v \\ 0 & \text{otherwise} \end{cases}$$

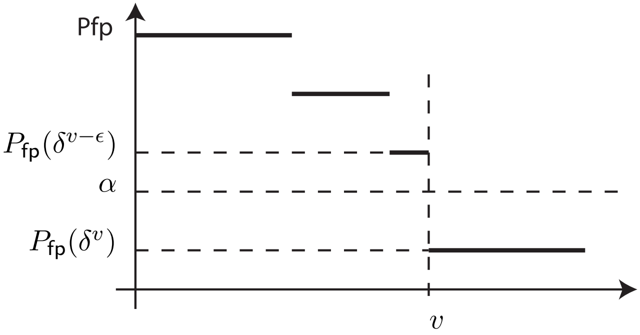
- If we can find a value of $v$ such that $P_{\mathsf{fp}} = \sum_{\ell : L_\ell > v} P_{\ell,0} = \alpha$ then we are done. The probability of detection is then

$$P_D = \sum_{\ell : L_\ell > v} P_{\ell,1} = \beta.$$

# N-P Hypothesis Testing With Discrete Observations

- If we cannot find a value of $v$ such that $P_{\sf fp} = \sum_{\ell:L_\ell > v} P_{\ell,0} = \alpha$ then it must be the case that, for any $\epsilon > 0$,

$$P_{\sf fp}(\delta^v) = \sum_{\ell:L_\ell>v} P_{\ell,0} < \alpha \quad \text{and} \quad P_{\sf fp}(\delta^{v-\epsilon}) = \sum_{\ell:L_\ell>v-\epsilon} P_{\ell,0} > \alpha$$




- In this case, we must randomize between decision rules $\delta^v$ and $\delta^{v-\epsilon}$.

# N-P Randomization

We form the convex combination between $\delta^v$ and $\delta^{v-\epsilon}$ as

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

for $\gamma \in [0,1]$. The false positive probability is then

$$P_{\mathsf{fp}} = (1-\gamma)P_{\mathsf{fp}}(\delta^v) + \gamma P_{\mathsf{fp}}(\delta^{v-\epsilon})$$

Setting this equal to $\alpha$ and solving for $\gamma$ yields

$$\begin{aligned} \gamma &= \frac{\alpha - P_{\mathsf{fp}}(\delta^v)}{P_{\mathsf{fp}}(\delta^{v-\epsilon}) - P_{\mathsf{fp}}(\delta^v)} \\ &= \frac{\alpha - \sum_{\ell:L_\ell>v} P_{\ell,0}}{\sum_{\ell:L_\ell=v} P_{\ell,0}} \end{aligned}$$

# N-P Decision Rule With Discrete Observations

The Neyman-Pearson decision rule for simple binary hypothesis testing with discrete observations is then:

$$\rho^{\mathsf{NP}}(y) = \begin{cases} 1 & \text{if } L_\ell > v \\ \gamma & \text{if } L_\ell = v \\ 0 & \text{if } L_\ell < v \end{cases}$$

where

$$L_\ell := \frac{\mathrm{Prob}(\text{observe } y \,|\, \text{state is } x_1)}{\mathrm{Prob}(\text{observe } y \,|\, \text{state is } x_0)} = \frac{P_{\ell,1}}{P_{\ell,0}}$$

and $v \geq 0$ is the minimum value such that

$$P_{\mathsf{fp}} = \sum_{\ell : L_\ell > v} P_{\ell,0} \leq \alpha.$$

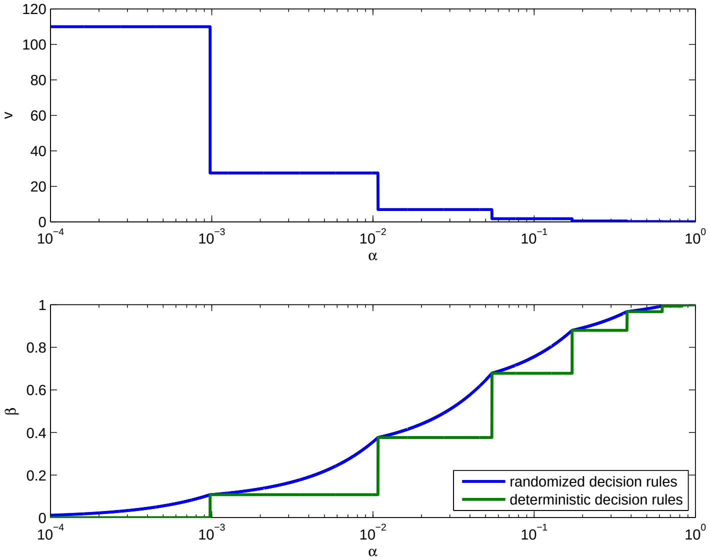
# Example: 10 Coin Flips

Coin flipping problem with a probability of heads of either $q_0 = 0.5$ or $q_1 = 0.8$. We observe ten flips of the coin and count the number of heads.

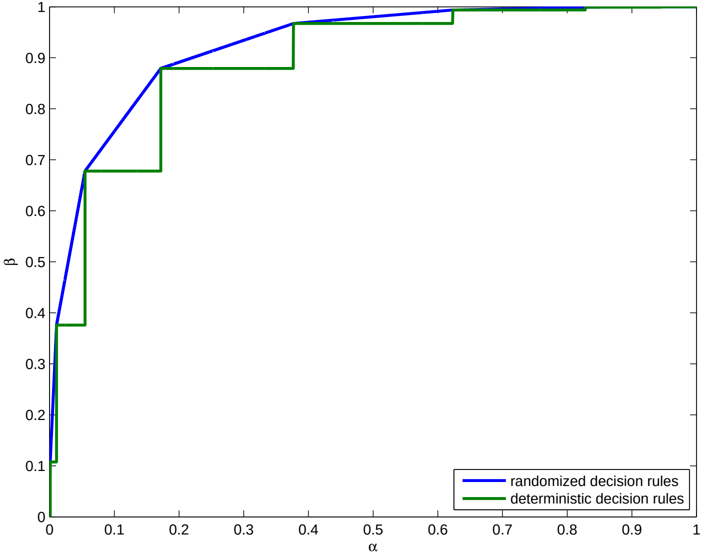
$$P = \begin{bmatrix} 0.0010 & 0.0000 \\ 0.0098 & 0.0000 \\ 0.0439 & 0.0001 \\ 0.1172 & 0.0008 \\ 0.2051 & 0.0055 \\ 0.2461 & 0.0264 \\ 0.2051 & 0.0881 \\ 0.1172 & 0.2013 \\ 0.0439 & 0.3020 \\ 0.0098 & 0.2684 \\ 0.0010 & 0.1074 \end{bmatrix} \text{ and } L = \begin{bmatrix} 0.0001 \\ 0.0004 \\ 0.0017 \\ 0.0067 \\ 0.0268 \\ 0.1074 \\ 0.4295 \\ 1.7180 \\ 6.8719 \\ 27.4878 \\ 109.9512 \end{bmatrix}$$

What is $v$, $\rho^{\mathsf{NP}}(y)$, and $\beta$ when $\alpha = 0.001$, $\alpha = 0.01$, $\alpha = 0.1$?

# Example: Randomized vs. Deterministic Decision Rules

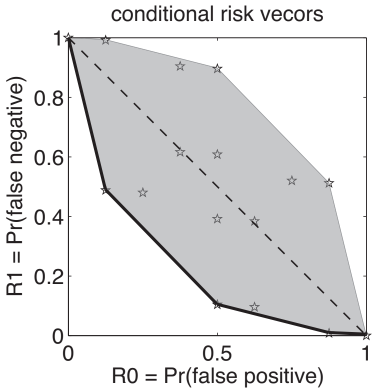


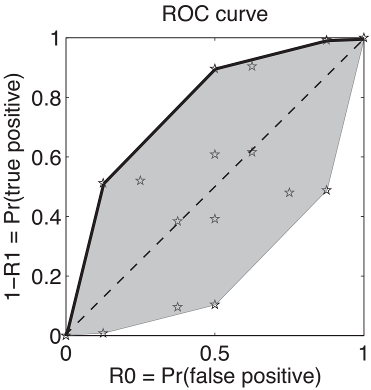
# Example: Same Results Except Linear Scale



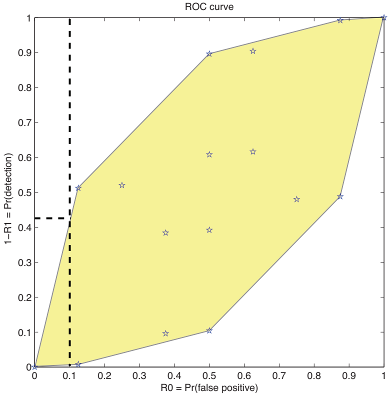
# Remarks: 1 of 2

The blue line on the previous slide is called the **Receiver Operating Characteristic** (ROC). An ROC plot shows the probability of detection $P_D = 1 - R_1$ as a function of $\alpha = R_0$. The ROC plot is directly related to our conditional risk vector plot.



# Remarks: 2 of 2




The N-P criterion seeks a decision rule that **maximizes the probability of detection** subject to the constraint that the probability of false alarm must be no greater than $\alpha$.

$$\rho^{\text{NP}} = \arg\max_{\rho} P_D(\rho)$$
$$\text{s.t. } P_{\text{fp}}(\rho) \leq \alpha$$

- The term **power** is often used instead of "probability of detection". The N-P decision rule is sometimes called the "most powerful test of significance level $\alpha$".
- Intuitively, we can expect that the power of a test will increase with the significance level of the test.