ECE531 Lecture 4a: Neyman-Pearson Hypothesis Testing

D. Richard Brown III

Worcester Polytechnic Institute

12-February-2009

Hypothesis Testing: What We Know

- Bayesian decision rules: minimize the expected/weighted risk for a particular prior distribution π.
- Minimax decision rules: minimize the worst-case risk exposure over all possible prior distributions.
- Example: To approve a new flu test, the FDA requires the test to have a false positive rate of no worse than 10%.
  - Should we use the Bayes criterion?
  - Should we use the minimax criterion?
  - How do we assign a risk structure to this sort of problem?
- 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).

The Neyman-Pearson Criterion and Terminology

For now, we will focus on simple binary hypothesis testing under the UCA.

$$\begin{aligned} R_0(\rho) &= \mathrm{Prob}(\text{decide } \mathcal{H}_1 | \text{state is } x_0) = P_{\mathsf{fp}} \\ &= \text{probability of "false positive" or probability of "false alarm".} \end{aligned}$$

and

$$\begin{aligned} R_1(\rho) &= \mathrm{Prob}(\text{decide } \mathcal{H}_0 | \text{state is } x_1) = P_{\mathsf{fn}} \\ &= \text{probability of "false negative" or "missed detection".} \end{aligned}$$

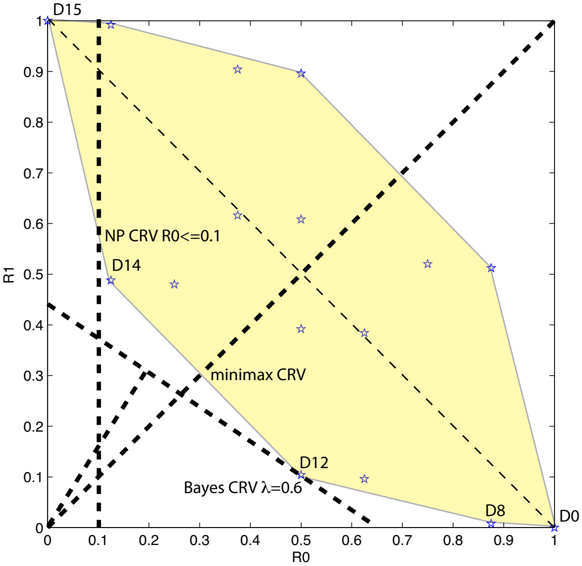
Definition

The Neyman-Pearson criterion decision rule is given as

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

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

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

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

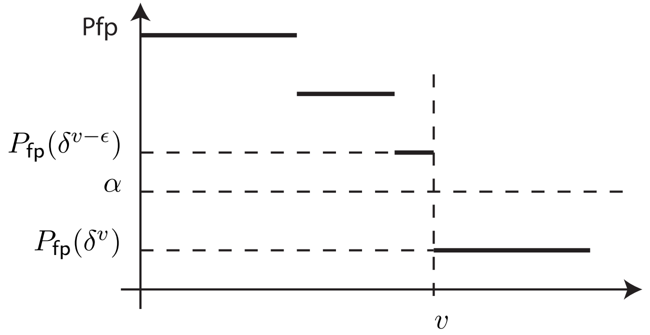
- If we can find a value of v such that $P_{\mathsf{fp}} = \sum_{\ell : L_\ell > v} P_{\ell,0} = \alpha$ then we are done. The probability of detection is then

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

N-P Hypothesis Testing With Discrete Observations

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

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



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

N-P Randomization

We form the usual convex combination between δ^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 α and solving for γ 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:

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

where

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

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

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

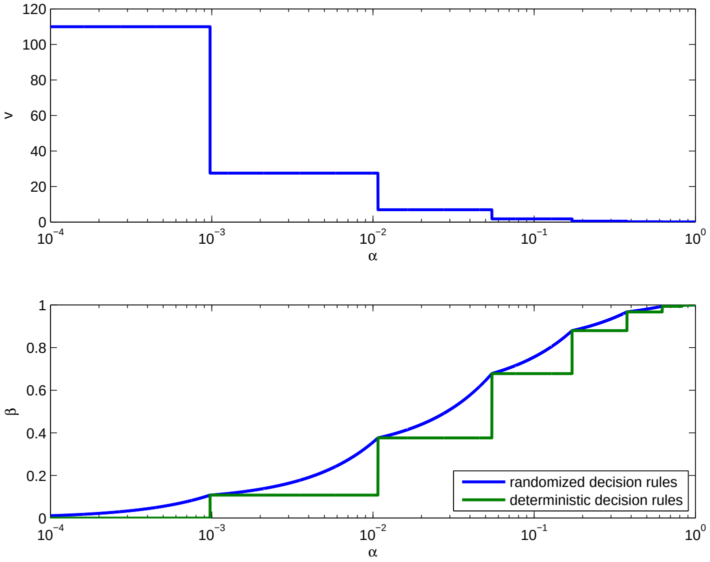
Example: 10 Coin Flips

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.

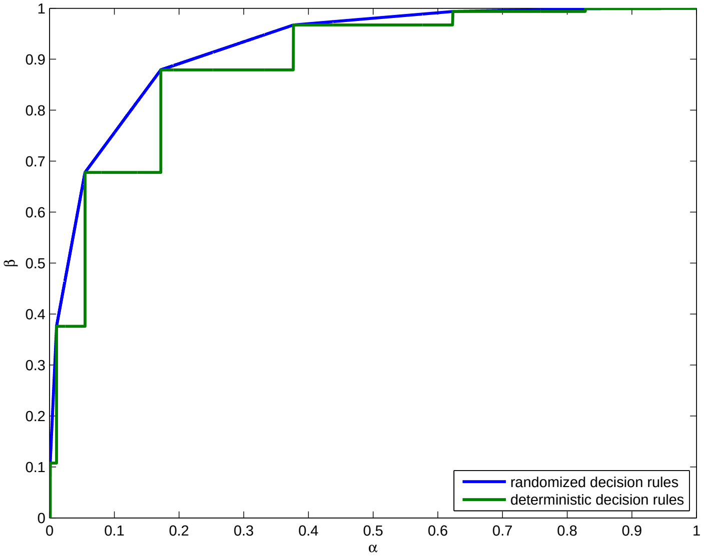
$$P = \begin{bmatrix} 0.0010 & 0.0000 \\ 0.0098 & 0.0000 \\ 0.0439 & 0.0001 \\ 0.1172 & 0.0008 \\ 0.2051 & 0.0055 \\ 0.2461 & 0.0264 \\ 0.2051 & 0.0881 \\ 0.1172 & 0.2013 \\ 0.0439 & 0.3020 \\ 0.0098 & 0.2684 \\ 0.0010 & 0.1074 \end{bmatrix} \text{ and } L = \begin{bmatrix} 0.0001 \\ 0.0004 \\ 0.0017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

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

Example: Randomized vs. Deterministic Decision Rules

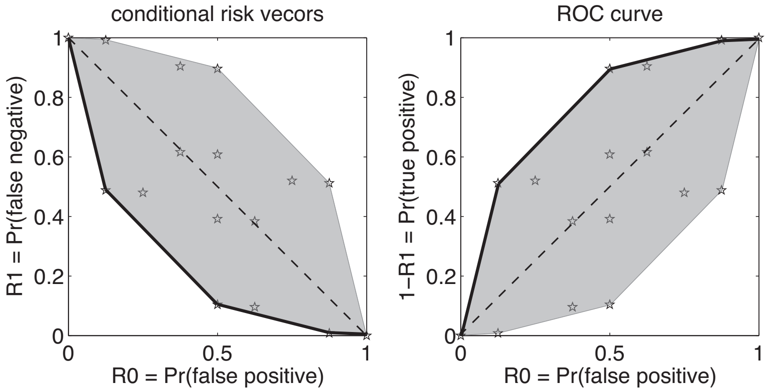


Example: Same Results Except Linear Scale

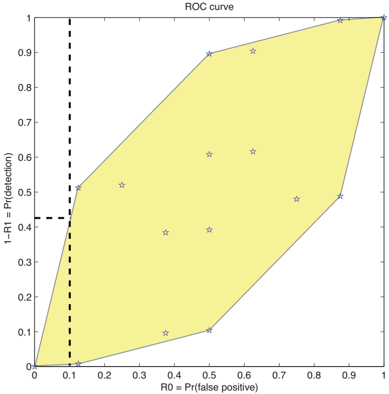


Remarks: 1 of 3

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 3



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 α.

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

- 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 α".
- Intuitively, we can expect that the power of a test will increase with the significance level of the test.

Remarks: 3 of 3

- Like Bayes and minimax, the N-P decision rule for simple binary hypothesis testing problems is just a likelihood ratio comparison (possibly with randomization).
- Can same intuition that you developed for the discrete observation case be applied in the continuous observation case?
  - Form $L(y) = \frac{p_1(y)}{p_0(y)}$.
  - Find the smallest v such that the decision rule

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

    has $P_{\mathsf{fp}} \leq \alpha$.
- The answer is "yes", but we need to formalize this claim by understanding the fundamental Neyman-Pearson lemma...

The Neyman-Pearson Lemma: Part 1 of 3: Optimality

Recall $p_j(y)$ for $j \in \{0,1\}$ and $y \in \mathcal{Y}$ is the conditional pmf or pdf of the observation y given that the state is x_j .

Lemma

Let ρ *be any decision rule satisfying* $P_{fp}(\rho) \le \alpha$ *and let* ρ' *be any decision rule of the form*

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

where $v \ge 0$ *and* $0 \le \gamma(y) \le 1$ *are such that* $P_{fp}(\rho') = \alpha$*. Then* $P_D(\rho') \ge P_D(\rho)$*.*

The Neyman-Pearson Lemma: Part 1 of 3: Optimality

Proof.

By the definitions of ρ and ρ', we always have $[\rho'(y) - \rho(y)][p_1(y) - vp_0(y)] \geq 0$. Hence

$$\int_{\mathcal{Y}} [\rho'(y) - \rho(y)][p_1(y) - vp_0(y)]\, dy \geq 0$$

Rearranging terms, we can write

$$\begin{aligned}
\int_{\mathcal{Y}} \rho'(y) p_1(y)\, dy - \int_{\mathcal{Y}} \rho(y) p_1(y)\, dy &\geq v \left[\int_{\mathcal{Y}} \rho'(y) p_0(y)\, dy - \int_{\mathcal{Y}} \rho(y) p_0(y)\, dy \right] \\
P_D(\rho') - P_D(\rho) &\geq v \left[P_{\mathsf{fp}}(\rho') - P_{\mathsf{fp}}(\rho) \right] \\
P_D(\rho') - P_D(\rho) &\geq v \left[\alpha - P_{\mathsf{fp}}(\rho) \right]
\end{aligned}$$

But $v \geq 0$ and $P_{\mathsf{fp}}(\rho) \leq \alpha$ implies that the RHS is non-negative. Hence

$$P_D(\rho') \geq P_D(\rho).$$

□

The Neyman-Pearson Lemma: Part 2 of 3: Existence

Lemma

For every $\alpha \in [0,1]$ there exists a decision rule ρ^{NP} of the form

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

where $v \geq 0$ and $\gamma(y) = \gamma \in [0,1]$ (a constant) such that $P_{fp}(\rho^{NP}) = \alpha$.

The Neyman-Pearson Lemma: Part 2 of 3: Existence

Proof by construction.

Let $\nu \geq 0$, $\mathcal{Y}_\nu = \{y \in \mathcal{Y} : p_1(y) > \nu p_0(y)\}$ and $\mathcal{Z}_\nu = \{y \in \mathcal{Y} : p_1(y) = \nu p_0(y)\}$. For $\nu_2 \geq \nu_1$, $\mathcal{Y}_{\nu_2} \subseteq \mathcal{Y}_{\nu_1}$ and $\int_{\mathcal{Y}_{\nu_2}} p_0(y)\,dy \leq \int_{\mathcal{Y}_{\nu_1}} p_0(y)\,dy$.
Let v be the smallest value of ν such that

$$\int_{\mathcal{Y}_v} p_0(y)\,dy \leq \alpha$$

Choose

$$\gamma = \begin{cases} \frac{\alpha - \int_{\mathcal{Y}_v} p_0(y)\,dy}{\int_{\mathcal{Z}_v} p_0(y)\,dy} & \text{if } \int_{\mathcal{Y}_v} p_0(y)\,dy < \alpha \\ \text{any arbitrary number in } [0,1] & \text{otherwise} \end{cases}$$

Then

$$P_{\text{fp}}(\rho^{\text{NP}}) = \int_{\mathcal{Y}_v} p_0(y)\,dy + \gamma \int_{\mathcal{Z}_v} p_0(y)\,dy = \alpha$$

□

The Neyman-Pearson Lemma: Part 3 of 3: Uniqueness

Lemma

Suppose that $\rho''(y)$ is any N-P decision rule for $\mathcal{H}_0$ versus $\mathcal{H}_1$ with significance level α. Then $\rho''(y)$ must be of the same form a $\rho^{NP}(y)$ except possibly on a subset of $\mathcal{Y}$ having zero probability under $\mathcal{H}_0$ and $\mathcal{H}_1$.

The Neyman-Pearson Lemma: Part 3 of 3: Uniqueness

Proof.

If ρ'' is a N-P decision rule with significance level α, then it must be true that $P_D(\rho'') = P_D(\rho^{\rm NP})$. From part 1 of the Lemma, we know that

$$P_D(\rho^{\rm NP}) - P_D(\rho'') \quad \geq \quad v\left[\alpha - P_{\rm fp}(\rho'')\right]$$

which implies that $P_{\rm fp}(\rho'') = \alpha$ since the LHS of the inequality is zero. So $P_D(\rho'') = P_D(\rho^{\rm NP})$ and $P_{\rm fp}(\rho'') = P_{\rm fp}(\rho^{\rm NP})$. We can work the proof of part 1 of the Lemma back to write

$$\int_{\mathcal{Y}} [\rho^{\rm NP}(y) - \rho''(y)][p_1(y) - vp_0(y)]\,dy = 0$$

Note that the integrand here is non-negative. This implies that $\rho^{\rm NP}(y)$ and $\rho''(y)$ can differ only on the set $\mathcal{Z}_v = \{y \in \mathcal{Y} : p_1(y) = vp_0(y)\}$. This then implies that $\rho^{\rm NP}(y)$ and $\rho''(y)$ must have the same form and can differ only in the choice of γ.

From part 2 of the lemma, we know that γ is arbitrary when $\int_{\mathcal{Z}_v} p_0(y)\,dy = 0$. Otherwise, if $\int_{\mathcal{Z}_v} p_0(y)\,dy > 0$, $\rho^{\rm NP}(y)$ and $\rho''(y)$ must share the same value of γ. □

Example: Coherent Detection of BPSK

Suppose a transmitter sends one of two scalar signals a_0 or a_1 and the signals arrive at a receiver corrupted by zero-mean additive white Gaussian noise (AWGN) with variance σ^2.
We want to use N-P hypothesis maximize

$$P_D = \mathrm{Prob}(\text{decide } \mathcal{H}_1 \,|\, a_1 \text{ was sent})$$

subject to the constraint

$$P_{\mathsf{fp}} = \mathrm{Prob}(\text{decide } \mathcal{H}_1 \,|\, a_0 \text{ was sent}) \le \alpha.$$

Signal model conditioned on state x_j:

$$Y = a_j + \eta$$

where a_j is the scalar signal and $\eta \sim \mathcal{N}(0, \sigma^2)$. Hence

$$p_j(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y-a_j)^2}{2\sigma^2}\right)$$

Example: Coherent Detection of BPSK

How should we approach this problem? We know from the N-P Lemma that the optimum decision rule will be of the form

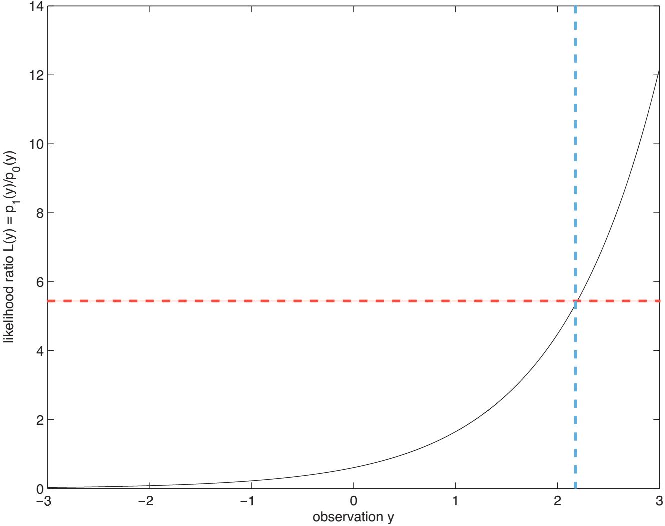
$$\rho^{\mathsf{NP}}(y) = \begin{cases} 1 & \text{if } p_1(y) > vp_0(y) \\ \gamma(y) & \text{if } p_1(y) = vp_0(y) \\ 0 & \text{if } p_1(y) < vp_0(y) \end{cases}$$

where $v \geq 0$ and $0 \leq \gamma(y) \leq 1$ are such that $P_{\mathsf{fp}}(\rho^{\mathsf{NP}}) = \alpha$. How should we choose our threshold v?
We need to find the smallest v such that

$$\int_{\mathcal{Y}_v} p_0(y)\, dy \leq \alpha$$

where $\mathcal{Y}_v = \{y \in \mathcal{Y} \, : \, p_1(y) > vp_0(y)\}$.

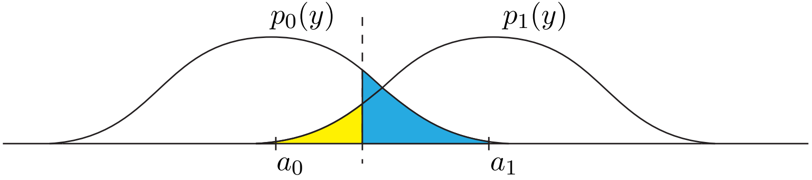
Example: Likelihood Ratio for $a_0 = 0$, $a_1 = 1$, $\sigma^2 = 1$



Example: Coherent Detection of BPSK

Note that, since $a_1 > a_0$, the likelihood ratio $L(y) = \frac{p_1(y)}{p_0(y)}$ is monotonically increasing. This means that finding v is equivalent to finding a threshold τ so that

$$\int_\tau^\infty p_0(y)\,dy \le \alpha \quad \Leftrightarrow \quad Q\left(\frac{\tau - a_0}{\sigma}\right) \le \alpha$$



How are τ and v related? Once we find τ, we can determine v by computing

$$v \;=\; L(\tau) = \frac{p_1(\tau)}{p_0(\tau)}.$$

Example: Coherent Detection of BPSK: Finding τ

Unfortunately, no "closed form" solution exists to exactly solve the inverse of a Q function. We can use the fact that $Q(x) = \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right)$ to write

$$Q\left(\frac{\tau - a_0}{\sigma}\right) = \frac{1}{2}\text{erfc}\left(\frac{\tau - a_0}{\sqrt{2}\sigma}\right) \quad \leq \quad \alpha.$$

This can be rewritten as

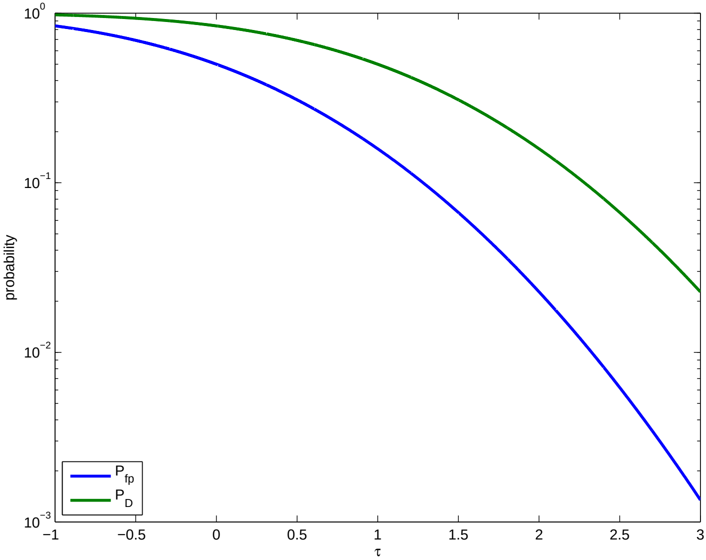
$$\tau \quad \geq \quad \sqrt{2}\sigma\text{erfc}^{-1}(2\alpha) + a_0$$

and we can use Matlab's handy `erfcinv` function to compute the lower bound on τ. It turns out that we are going to always be able to find a value of τ such that

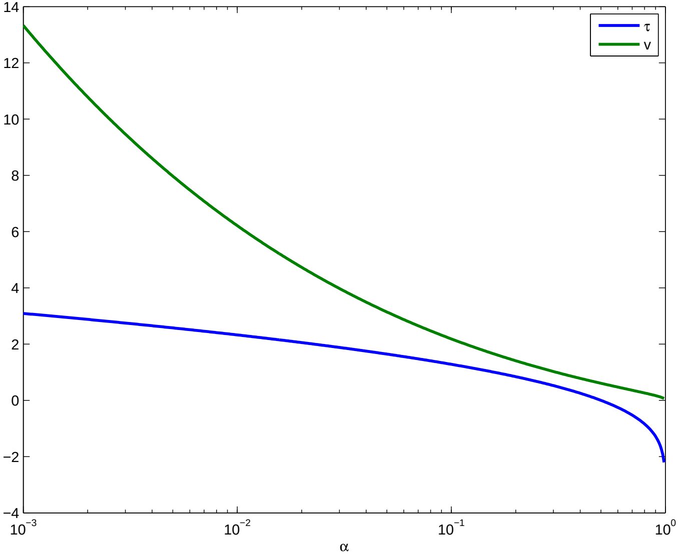
$$Q\left(\frac{\tau - a_0}{\sigma}\right) = \alpha$$

so we won't have to worry about randomization here.

Example: Coherent Detection of BPSK



Example: Coherent Detection of BPSK



Final Comments on Neyman-Pearson Hypothesis Testing

1. N-P decision rules are useful in asymmetric risk scenarios or in scenarios where one has to guarantee a certain probability of false detection.
2. N-P decision rules are simply likelihood ratio comparisons, just like Bayes and minimax. The comparison threshold in this case is chosen to satisfy the significance level constraint.
3. Like minimax, randomization is often necessary for N-P decision rules. Without randomization, the power of the test may not be maximized for the significance level constraint.
4. The original N-P paper: "On the Problem of the Most Efficient Tests of Statistical Hypotheses," J. Neyman and E.S. Pearson, *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, Vol. 231 (1933), pp. 289-337. Available on jstor.org.