6: Neyman-Pearson Detectors

ECE 830, Spring 2014

Main result from Lecture 5

The likelihood ratio statistic is optimal for testing between two simple hypotheses.

The test simply compares the likelihood ratio to a threshold:

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma = \frac{\pi_0(c_{1,0} - c_{0,0})}{\pi_1(c_{0,1} - c_{1,1})} \,.$$

How do we choose the best threshold?

The "optimal" threshold is a function of the prior probabilities and the costs assigned to different errors. The choice of costs is subjective and depends on the nature of the problem, but the prior probabilities must be known. Unfortunately, often the prior probabilities are not known precisely, and thus the correct setting for the threshold is unclear.

Example: Search for extra-terrestrial intelligence

$$H_0 : X \sim \text{cosmic radiation}$$
$$H_1 : X \sim \text{cosmic radiation} + \text{intelligent signal}$$

What is π_1?

Alternative design specification

Let's design a test that minimizes one type of error subject to a constraint on the other type of error.

This constrained optimization criterion does not require knowledge of prior probabilities or cost assignments. It only requires a specification of the maximum allowable value for one type of error, which is sometimes even more natural than assigning costs to the different errors.

Neyman Pearson testing

A classic result due to Neyman and Pearson shows that the solution to this type of optimization is again a likelihood ratio test.

Neyman-Pearson Lemma

Assume that we observe a random variable distributed according to one of two distributions.

$$
\begin{aligned}
H_0 : X &\sim p_0 \\
H_1 : X &\sim p_1
\end{aligned}
$$

Definition: null and alternative hypotheses

H_0 is consider to be a sort of baseline or default model and is called the null hypothesis. H_1 is a different model and is called the alternative hypothesis.

Definition: false alarms and misses

If a test chooses H_1 when in fact the data were generated by H_0 the error is called a false-positive or false-alarm, since we mistakenly accepted the alternative hypothesis. The error of deciding H_0 when H_1 was the correct model is called a false-negative or miss.

Let T denote a testing procedure based on an observation of X. and let R_T denote the subset of the range of X where the test chooses H_1. The probability of a false-positive is denoted by

$$P_{FA} = P_0(R_T) \; := \; \int_{R_T} p_0(x)\, dx \; .$$

The probability of a false-negative is $P_M = 1 - P_1(R_T)$, where

$$P_D = P_1(R_T) \; := \; \int_{R_T} p_1(x)\, dx \; ,$$

is the probability of correctly deciding H_1, often called the probability of detection.

Note that P_{FA} and P_D do not depend on prior probabilities π_0 and π_1.

Statistics vocabulary

These same ideas all arise in statistics, but that community uses different language.

- $1 - P_{FA}$ is call the specificity of a test.
- $P_D = 1 - P_M$ is called the power or sensitivity of a test.
- Engineers talk about keeping P_{FA} small while maximizing P_D.
- Statisticians talk about keeping both the specificity and power of a test large.
- These are equivalent ideas.

Consider likelihood ratio tests of the form

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \ .$$

The subset of the range of X where this test decides H_1 is denoted

$$R_{LR}(\lambda) \ := \ \{x \,:\, p_1(x) > \lambda \, p_0(x)\} \ ,$$

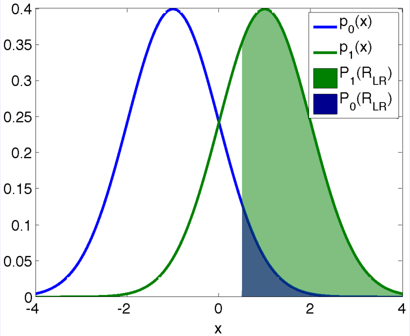
and therefore the probability of a false-positive decision is

$$P_0(R_{LR}(\lambda)) \ := \ \int_{R_{LR}(\lambda)} p_0(x)\, dx \ = \ \int_{\{x:p_1(x)>\lambda p_0(x)\}} p_0(x)\, dx$$

This probability is a function of the threshold λ; the set $R_{LR}(\lambda)$ shrinks/grows as λ increases/decreases. We can select λ to achieve a desired probability of error.

Example:

$$H_0 : X \sim \mathcal{N}(-1, 1)$$
$$H_1 : X \sim \mathcal{N}(1, 1)$$



As γ increases, P_{FA} decreases (good) and P_D decreases (bad).

Neyman-Pearson Lemma

Consider the likelihood ratio test

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

with $\lambda > 0$ chosen so that $P_0(R_{LR}(\lambda)) = \alpha$. There does not exist another test T with $P_0(R_T) \leq \alpha$ and $P_1(R_T) > P_1(R_{LR}(\lambda))$. That is, the LRT is the **most powerful test** with probability of false-negative less than or equal to α.

In other words, the NP test maximizes P_D subject to a constraint on P_{FA}:

$$\lambda_{NP} = \arg\max_{\lambda} P_D(\lambda) \text{ subject to } P_{FA}(\lambda) \leq \alpha$$

Proof: Let T be any test with $P_0(R_T) = \alpha$. Let NP denote the LRT with λ chosen so that $P_0(R_{LR}(\lambda)) = \alpha$. To simplify the notation we will denote use R_{NP} to denote the region $R_{LR}(\lambda)$.
We want to show that

For any subset R of the range of X define

$$P_i(R) := \int_R p_i(x)\, dx,$$

This is simply the probability of $X \in R$ under hypothesis H_i. Note that

$$\begin{aligned} P_i(R_{NP}) &= P_i(R_{NP} \cap R_T) + \\ P_i(R_T) &= P_i(R_{NP} \cap R_T) + \end{aligned}$$

where the superscript c indicates the complement of the set.

By assumption $P_0(R_{NP}) = P_0(R_T) = \alpha$, therefore

$$\begin{aligned} P_0(R_{NP} \cap R_T^c) =& P_0(R_{NP}) - P_0(R_{NP} \cap R_T) \\ =& P_0(R_T) - P_0(R_{NP} \cap R_T) \\ =& P_0(R_{NP}^c \cap R_T). \end{aligned}$$

Now, we will show

$$\begin{aligned} P_1(R_{NP}) =& P_1(R_{NP} \cap R_T) + P_1(R_{NP} \cap R_T^c) \\ \geq& P_1(R_{NP} \cap R_T) + P_1(R_{NP}^c \cap R_T) \\ =& P_1(R_T). \end{aligned}$$

which holds if

$$P_1(R_{NP} \cap R_T^c) \geq P_1(R_{NP}^c \cap R_T) \ .$$

To see that this is indeed the case,

$$
\begin{aligned}
P_1(R_{NP} \cap R_T^c) &= \int_{R_{NP} \cap R_T^c} p_1(x)\,dx \\
&\geq \lambda \int_{R_{NP} \cap R_T^c} p_0(x)\,dx \\
&= \lambda\, p_0(R_{NP} \cap R_T^c) \\
&= \lambda\, p_0(R_{NP}^c \cap R_T) \\
&= \lambda \int_{R_{NP}^c \cap R_T} p_0(x)\,dx \\
&\geq \int_{R_{NP}^c \cap R_T} p_1(x)\,dx \\
&= P_1(R_{NP}^c \cap R_T).
\end{aligned}
$$

Detecting a DC Signal in Additive White Gaussian Noise

Consider the binary hypotheses

$$\begin{aligned} H_0 :&\ X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \\ H_1 :&\ X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)\ ,\ \mu > 0 \end{aligned}$$

and assume that $\sigma^2 > 0$ is known. The first hypothesis is simple. It involves a fixed and known distribution. The second hypothesis is simple if μ is known. However, if all we know is that $\mu > 0$, then the second hypothesis is the composite of many alternative distributions, i.e., the collection $\{N(\mu, \sigma^2)\}_{\mu>0}$. In this case, H_1 is called a composite hypothesis.

The likelihood ratio test takes the form

$$\frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x_i^2}} = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}}{\frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

The inequalities are preserved if we apply a monotonic transformation to both sides, so we can simplify the expression by taking the logarithm, giving us the log-likelihood ratio test

$$\frac{-1}{2\sigma^2}\left(-2\mu\sum_{i=1}^{n} x_i + n\mu^2\right) \underset{H_0}{\overset{H_1}{\gtrless}} \log(\gamma)$$

Assuming $\mu > 0$, this is equivalent to

$$\sum_{i=1}^{n} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \nu \ ,$$

with $\nu = \frac{\sigma^2}{\mu} \ln \gamma + \frac{n\mu}{2}$, and since γ was ours to choose, we can equivalently choose ν to trade-off between the two types of error. Note that $t := \sum_{i=1}^{n} x_i$ is simply the sufficient statastic for the mean of a normal distribution. Let's rewrite our hypotheses in terms of the sufficient statistic:

$$H_0 : t \sim \qquad H_1 : t \sim \qquad \mu > 0$$

Let's now determine P_{FA} and P_D for the log-likelihood ratio test.

$$P_{FA} \;=\; \int_{\nu}^{\infty} \frac{1}{\sqrt{2n\pi\sigma^2}}\, e^{-\frac{t^2}{2n\sigma^2}}\, dt \;=\; Q\left(\frac{\nu}{\sqrt{n\sigma^2}}\right) \;,$$

where $Q(z) = \int_{u \geq z} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}\, du$, the tail probability of the standard normal distribution. Similarly,

$$P_D \;=\; \int_{\nu}^{\infty} \frac{1}{\sqrt{2n\pi\sigma^2}}\, e^{-\frac{(t-n\mu)^2}{2n\sigma^2}}\, dt \;=\; Q\left(\frac{\nu - n\mu}{\sqrt{n\sigma^2}}\right) \;.$$

In both cases the expression in terms of the Q function is the result of a simple change of variables in the integration. (See earlier notes on the error function.)

The Q function is invertible, so we can solve for the value of ν in terms of P_{FA}, that is

$$\nu = \qquad\qquad\qquad\qquad .$$

Note that we are choosing the threshold ν based only on the desired P_{FA}; we don't need to know μ! Using this we can express P_D as

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{n\mu^2}{\sigma^2}}\right),$$

where $\sqrt{\frac{n\mu^2}{\sigma^2}}$ is simply the signal-to-noise ratio (SNR). Since $Q(z) \to 1$ as $z \to -\infty$, it is easy to see that the probability of detection increases as μ and/or n increase.

Detecting a Change in Variance

Consider the binary hypotheses

$$\begin{aligned} H_0 :\ & X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0, \sigma_0^2) \\ H_1 :\ & X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0, \sigma_1^2)\ ,\ \sigma_1 > \sigma_0 \end{aligned}$$

The log-likelihood ratio test is

$$\frac{n}{2} \log \left(\frac{\sigma_0^2}{\sigma_1^2} \right) + \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2} \right) \sum_{i=1}^{n} x_i^2 \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \ln(\gamma)\ .$$

Some simple algebra shows

$$\sum_{i=1}^{n} x_i^2 \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \nu$$

with $\nu = 2 \left(\frac{\sigma_1^2 \sigma_o^2}{\sigma_1^2 - \sigma_0^2} \right) (\log(\gamma) + n \ln(\frac{\sigma_1}{\sigma_o}))$. Note that $t := \sum_{i=1}^{n} x_i^2$ is the sufficient statistic for variance of a zero-mean normal distribution.

Now recall that if $X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$, then $\sum_{i=1}^n X_i^2 \sim \chi_n^2$ (chi-square distributed with n degrees of freedom). Let's rewrite our null hypothesis test using the sufficient statistic:

$$H_0 \;:\; t = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_0^2} \sim \chi_n^2$$

The probability of false alarm is just the probability that a χ_n^2 random variable exceeds ν/σ_0^2. This can be easily computed numerically. For example, if we have $n = 20$ and set $P_{FA} = 0.01$, then the correct threshold is $\nu = 37.57\sigma_0^2$.

Applying the NP test

1. Decide which hypothesis to call the null and which to call the alternative. Usually we choose for the null the hypothesis where we have more serious consequences if it's rejected (because with NP testing we control the probability of rejecting the null).
2. Select the power of the test. Often we see $\alpha = 0.05$, meaning we will accept a 5% chance of falsely rejecting the null (e.g., falsely saying a target is present when it is not).
3. Use the likelihood ratio and α to choose the threshold level ν. Often this can be made simpler by expressing the problem in terms of sufficient statistics.