CS/CNS/EE/IDS 165: Foundations in Machine Learning and Statistical Inference

Neyman Pearson Detection

http://tensorlab.cms.caltech.edu/users/anima/cms165-2020.html

Anima Anandkumar
Computing and Mathematical Sciences
California Institute of Technology, Pasadena CA 91125
anima@caltech.edu

Outline

Topics

- Size and power of detectors.
- Uniformly most powerful (UMP) detector.
- The Neyman-Pearson Lemma.
- Monotone likelihood ratio.
- Karlin Rubin Theorm.
- Examples.

References

1. H.V. Poor, An Introduction to Signal Detection and Estimation, 2nd Ed., Springer-Verlag, 1994, Chapter II.C.
2. L. L. Scharf, Statistical Signal Processing: Detection, Estimation and Time Series Analysis, Addison-Wesley, Publishing Company, Inc., 1991, Chapter 4.
3. S.M. Kay, Fundamentals of Statistical Signal Processing: Detection Theory, Prentice Hall, 1993, Chapter 3.
4. E.L. Lehmann, Testing Statistical Hypotheses, John Wiley & Sons, Inc., 1959, Chapter 3.
5. P.J. Bickel and K.A. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, Prentice Hall, 1977, Chapter 5.
6. T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, 1967, Chapter 5.
7. J. Shao, Mathematical Statistics, Springer-Verlag, 1999, Chap. 6.

Power and Size

Binary Hypotheses

Consider binary hypotheses $\mathcal{H}_i : Y \sim f(y|\theta), \quad \theta \in \Lambda_i$, $i = 0, 1$. Hypothesis $\mathcal{H}_0$ is called the null hypothesis and $\mathcal{H}_1$ the alternative. We consider the general (randomized) detector of the form $\delta(y) = \Pr(D = 1|Y = y)$.

False Alarm and Size

The false alarm, or the type I error, is defined as, $\forall\theta \in \Lambda_0$,

$$P_{\mathsf{F}}(\delta; \theta) \triangleq \Pr(D = 1; \theta) = \int \delta(\mathbf{y}) f(y|\theta) dy = \mathbb{E}_\theta(\delta(y)),$$

The size or level of a detector is defined by

$$\sup_{\theta \in \Lambda_0} P_F(\delta, \theta)$$

Miss Detection and Power

The miss detection, or type II error, is defined as, $\forall\theta \in \Lambda_1$,

$$P_{\mathsf{M}}(\delta; \theta) \triangleq \Pr(D = 0; \theta) = \int (1 - \delta(y)) f(y|\theta) dy = 1 - \mathbb{E}_\theta(\delta(Y))$$

The power of a detector is given by the probability of detection

$$P_{\mathsf{D}}(\delta, \theta) = \Pr(D = 1; \theta) = \mathbb{E}_\theta(\delta(Y)), \quad \theta \in \Lambda_1$$

Remark: We expect that power increases with size.

Receiver Operating Characteristic (ROC)

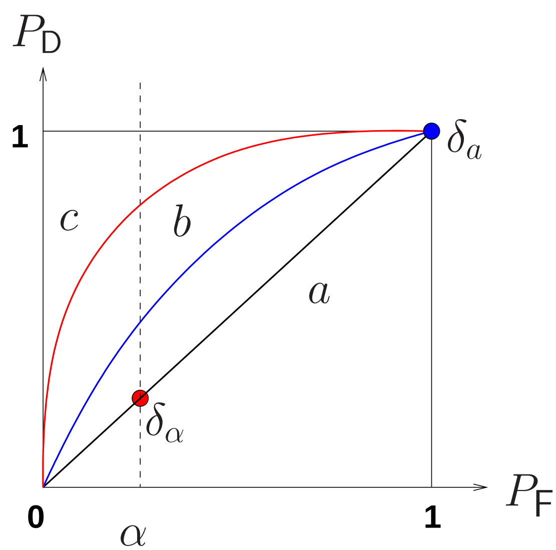
Consider simple binary hypotheses

$$\mathcal{H}_0 : Y \sim f(y|\theta_0) \quad \text{vs.} \quad \mathcal{H}_1 : Y \sim f(y|\theta_1).$$

The performance of a fixed detector is characterized by the false alarm and the detection power (P_F, P)

$$P_\mathsf{F} \triangleq \mathbb{E}_{\theta_0}(\delta(Y)) \qquad P_\mathsf{D} \triangleq \mathbb{E}_{\theta_1}(\delta(Y))$$

The ROC curve describes P as a functional P_F.



- Each curve represents a class of detectors indexed by the false alarm probability P_F. The detectors lie on curve (b) are all better than those on (a) and worse than those on (c).
- Detector δ_a is the constant detector $(\delta_a(y) = 1)$.
- What is detector δ_α?

The UMP Detector

Uniformly Most Powerful Detector

A size α detector δ_{UMP} is uniformly most powerful (UMP) if, for all δ of size no greater than α,

$$P_{\mathsf{D}}(\delta_{\mathsf{UMP}}, \theta) \geq P_{\mathsf{D}}(\delta, \theta), \quad \forall \theta \in \Lambda_1.$$

For simple binary hypotheses, the UMP detector is given by

$$\max P_{\mathsf{D}}(\delta, \theta_1) \quad \text{subject to} \quad P_{\mathsf{F}}(\delta, \theta_0) \leq \alpha.$$

The Neyman-Pearson Lemma

Consider simple binary hypotheses

$$\mathcal{H}_0 : \quad Y \sim f(y|\theta_0)$$
$$\mathcal{H}_1 : \quad Y \sim f(y|\theta_1)$$

Let δ^* be a likelihood ratio detector with threshold η

$$\delta^*(y) = \begin{cases} 1 & \frac{f(y|\theta_1)}{f(y|\theta_0)} \geq \eta \\ 0 & \text{otherwise} \end{cases}$$

Then for any deterministic detector δ of size less than that of δ^*,

$$P_{\mathsf{D}}(\delta) < P_{\mathsf{D}}(\delta^*)$$

Proof:

For δ^*, the acceptance region Γ_0^* and its complement Γ_1^*—the rejection region—are defined by

$$\Gamma_1^* \triangleq \{y : \frac{f(y|\theta_1)}{f(y|\theta_0)} \geq \eta\}, \quad \Gamma_0^* \triangleq \{y : \frac{f(y|\theta_1)}{f(y|\theta_0)} < \eta\}.$$

Let Γ_0 be the acceptance region for any deterministic detector δ with size less than α, and Γ_1 be the rejection region of δ.

$$\begin{aligned}
P_{\mathsf{D}}(\delta^*) - P_{\mathsf{D}}(\delta) &= \int_{\Gamma_1^*} f(y|\theta_1)dy - \int_{\Gamma_1} f(y|\theta_1)dy \\
&= \int_{\Gamma_1^* \bigcap \Gamma_0} f(y|\theta_1)dy - \int_{\Gamma_1 \bigcap \Gamma_0^*} f(y|\theta_1)dy \\
&> \int_{\Gamma_1^* \bigcap \Gamma_0} \eta f(y|\theta_0)dy - \int_{\Gamma_1 \bigcap \Gamma_0^*} \eta f(y|\theta_0)dy \\
&= \eta \left\{ \int_{\Gamma_1^*} f(y|\theta_0)dy - \int_{\Gamma_1} f(y|\theta_0)d\mathbf{y} \right\} \\
&= \eta(P_{\mathsf{F}}(\delta^*) - P_{\mathsf{F}}(\delta)) \geq 0
\end{aligned}$$

Remarks:

- The NP detector is also a likelihood ratio detector. The same is true for the Bayesian and Minimax detectors.
- The NP lemma in the above form does not allow the specification of the size of the NP detector.
- Is randomization necessary? The proof is restrictive since a randomized detector does not necessarily partition the observation space into Γ_0 and Γ_1.

Example

Let $\mathbf{Y} = [Y_1, \cdots, Y_n]^T$ and $Y_i \overset{i.i.d.}{\sim} \mathcal{N}(\theta, \sigma^2)$. The simple binary hypotheses are defined by

$$\begin{aligned}\mathcal{H}_0 : &\quad \theta = \theta_0 \\ \mathcal{H}_1 : &\quad \theta = \theta_1 > \theta_0\end{aligned}$$

Find the size α NP detector.

The likelihood ratio is given by

$$\begin{aligned}L(\mathbf{y}) &\triangleq \frac{f(\mathbf{y}|\theta_1)}{f(\mathbf{y}|\theta_0)} = \frac{\exp\{-\frac{1}{2\sigma^2}||\mathbf{y} - \boldsymbol{\theta}_1||^2\}}{\exp\{-\frac{1}{2\sigma^2}||\mathbf{y} - \boldsymbol{\theta}_0||^2\}} \\ &= \exp\{\frac{2\mathbf{y}^T(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) + ||\boldsymbol{\theta}_0||^2 - ||\boldsymbol{\theta}_1||^2}{2\sigma^2}\} = \exp\{\frac{(\theta_1 - \theta_0)\sum_i y_i}{\sigma^2} + \text{constant}\}\end{aligned}$$

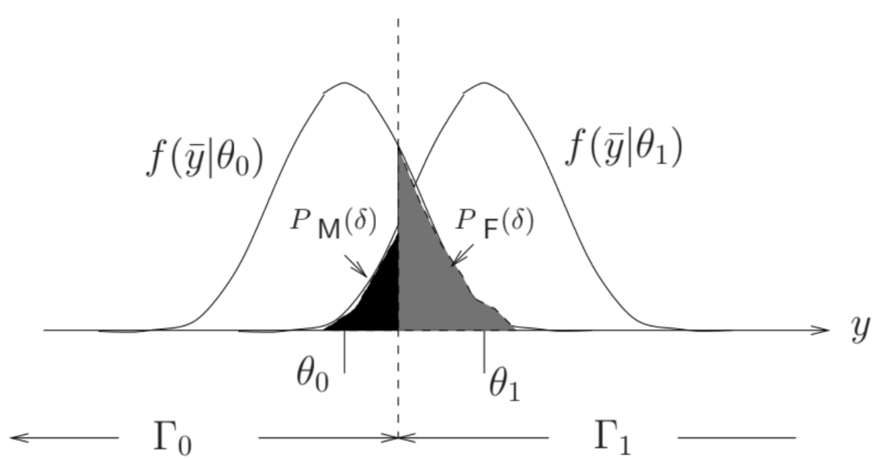
To maximize the power, we only need to find a size α likelihood ratio detector. For this we can work with the log-likelihood ratio

$$l(\mathbf{y}) \triangleq \ln L(\mathbf{y}) = \frac{(\theta_1 - \theta_0)}{\sigma^2}\sum_i y_i + \text{constant}$$

and consider the following detector

$$\bar{y} \triangleq \frac{1}{n}\sum_i y_i \geq \tau, \leftrightarrow \delta(\mathbf{y}) = \begin{cases} 1 & \bar{y} \geq \tau \\ 0 & \bar{y} < \tau \end{cases}$$

where τ is chosen to satisfy the size requirement.



Because under $\mathcal{H}_0$, $\bar{y} \sim \mathcal{N}(\theta_0, \frac{\sigma^2}{n})$,

$$P_{\mathsf{F}}(\delta) = \Pr(\bar{y} \geq \tau) = Q(\frac{\sqrt{n}(\tau - \theta_0)}{\sigma}) = \alpha$$

Hence

$$\tau = \theta_0 + \frac{\sigma Q^{-1}(\alpha)}{\sqrt{n}}$$

To compute the power of the detector, under $\mathcal{H}_1$,

$$P_{\mathsf{D}}(\delta) = \Pr(\bar{y} \geq \tau) = Q(\frac{\sqrt{n}(\tau - \theta_1)}{\sigma})$$

Remarks:

- As $\sigma \to 0$, or $n \to \infty$, $\tau \to \theta_0$. This is in contrast to the Bayesian detector with equal prior which is also the minimax detector.
- As $\sigma \to \infty$, $\tau \to \infty$. The probability that the detection is $\mathcal{H}_0$ approaches to 1. Again, this is different from the Bayesian and minimax detectors.
- Notice that the threshold does not depend on θ_1! This implies that if we consider the following simple hypothesis vs. composite alternative

$$\begin{aligned} \mathcal{H}_0: &\quad \theta = \theta_0 \\ \mathcal{H}_1: &\quad \theta \in (\theta_0, \infty) = \Lambda_1 \end{aligned}$$

  The same detector is optimal for all $\theta \in \Lambda_1$. Therefore, δ is a UMP detector.

Example: Fair vs. Double-headed Coins

Example: Recall the the coin tossing problem:

$$\begin{aligned} \mathcal{H}_0: &\quad Y \sim f(\mathsf{H}|\theta_0) = f(\mathsf{T}|\theta_0) = 0.5, \\ \mathcal{H}_1: &\quad Y \sim f(\mathsf{H}|\theta_1) = 1, f(\mathsf{T}|\theta_1) = 0. \end{aligned}$$

Suppose that we want to find the most powerful detector of size α. The Neyman-Pearson detector detector is based on testing the likelihood ratio $L(y)$ against certain threshold τ, where

$$L(y) = \frac{f(y|\theta_1)}{f(y|\theta_0)} = \begin{cases} 0 & y = \mathsf{T} \\ 2 & y = \mathsf{H} \end{cases}$$

Here we must be careful to specify the test because $L(y)$ is discrete. Regardless what the threshold is used, there are three possibilities:

1. $\tau = 0, \delta_1(y) = 1$ for all y.

$$P_{\mathsf{F}}(\delta_1) = 1, P_{\mathsf{D}}(\delta_1) = 1$$

   This detector is optimal only if $\alpha = 1$.

2. $\tau \in (0, 2], \delta_2(\mathsf{H}) = 1, \delta_2(\mathsf{T}) = 0$

$$P_{\mathsf{F}}(\delta_2) = 0.5, P_{\mathsf{D}}(\delta_2) = 1$$

   This detector is optimal if $\alpha = 0.5$.

3. $\tau > 2, \delta_3(y) = 0$ for all y

$$P_{\mathsf{F}}(\delta_3) = 0, P_{\mathsf{D}}(\delta_3) = 0$$

   This detector is optimal if $\alpha = 0$.

Randomization can be used to improve the performance. For example, the following detector has size $\alpha = 0.1$

$$\delta_4 = \begin{cases} \delta_2 & \text{with probability } 0.2 \\ \delta_3 & \text{with probability } 0.8 \end{cases} \rightarrow \delta_4(y) = \begin{cases} 0 & y = \mathsf{T} \\ 0.2 & y = \mathsf{H} \end{cases}$$

The power of this detector is given by

$$P_D(\delta_4) = 0.2$$

Is this best we can do?

Clearly, there are other choices for the same size $\alpha = 0.1$:

$$\delta_5 = \begin{cases} \delta_1 & \text{with probability } 0.1 \\ \delta_3 & \text{with probability } 0.9 \end{cases} \qquad \delta_6 = \begin{cases} \delta_1 & \text{with probability } 0.05 \\ \delta_2 & \text{with probability } 0.1 \\ \delta_3 & \text{with probability } 0.8 \end{cases}$$

but these have smaller power than that of δ_4.

The Neyman-Pearson Lemma

Theorem

Consider the simple binary hypothesis testing

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} \sim f(\mathbf{y}|\theta_0) \\ \mathcal{H}_1 : &\quad \mathbf{y} \sim f(\mathbf{y}|\theta_1). \end{aligned}$$

1. Optimality. Any likelihood ratio detector of the form

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & f(\mathbf{y}|\theta_1) > \eta f(\mathbf{y}|\theta_0), \\ \gamma(\mathbf{y}) & f(\mathbf{y}|\theta_1) = \eta f(\mathbf{y}|\theta_0), \\ 0 & f(\mathbf{y}|\theta_1) < \eta f(\mathbf{y}|\theta_0), \end{cases} \tag{1}$$

   for some $\eta \geq 0$ and $\gamma(\mathbf{y}) \in [0, 1]$, is the best of its size.

2. Existence. For every $\alpha \in [0, 1]$, there exists a detector of the form in (1). The threshold η_0 for the likelihood ratio test is the the smallest number η such that $\Pr(L(\mathbf{Y}) > \eta; \theta_0) \leq \alpha$, *i.e.*,

$$\eta_0 = \min \eta \;\; \text{subject to} \;\; \Pr(L(\mathbf{Y}) > \eta; \theta_0) \leq \alpha.$$

   and the randomization $\gamma(\mathbf{y})$ is a constant defined by

$$\gamma(\mathbf{y}) = \begin{cases} \frac{\alpha - \Pr(L(\mathbf{Y}) > \eta_0; \theta_0)}{\Pr(L(\mathbf{Y}) = \eta_0; \theta_0)} \triangleq \gamma_0 & \Pr(L(\mathbf{Y}) = \eta_0; \theta_0) \neq 0 \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

3. Uniqueness. If δ' is a size α NP detector, then $\delta'(\mathbf{y})$ has the form in (1) except perhaps for a set of $\mathbf{y}$ with zero probability under both $\mathcal{H}_0$ and $\mathcal{H}_1$.

Remarks and Proof

Remarks:

- Randomization is only necessary for a constant γ and only on the boundary between the two decision regions.
- The test in (1) is optimal even if the functions $f(\mathbf{y}|\theta_0)$ and $f(\mathbf{y}|\theta_1)$ may take negative values.

Proof:

Optimality: Let $\delta^*(\mathbf{y})$ have the form in (1), and let $\delta(\mathbf{y})$ be any other detector. Note that, for any $\mathbf{y}$,

$$(\delta^*(\mathbf{y}) - \delta(\mathbf{y}))(f(\mathbf{y}|\theta_1) - \eta f(\mathbf{y}|\theta_0)) \geq 0$$

Thus

$$\underbrace{\int \delta^*(\mathbf{y}) f(\mathbf{y}|\theta_1) d\mathbf{y}}_{P_D(\delta^*)} - \underbrace{\int \delta(\mathbf{y}) f(\mathbf{y}|\theta_1) d\mathbf{y}}_{P_D(\delta)} \geq \eta \left[\underbrace{\int \delta^*(\mathbf{y}) f(\mathbf{y}|\theta_0) d\mathbf{y}}_{P_F(\delta^*)} - \underbrace{\int \delta^*(\mathbf{y}) f(\mathbf{y}|\theta_0) d\mathbf{y}}_{P_F(\delta)} \right].$$

So for any detector δ of size no greater than δ^*, it has power no greater than that of δ^*.

Uniqueness: If there is another detector δ of size α and power $P_D(\delta^*)$, then

$$\int (\delta^*(\mathbf{y}) - \delta(\mathbf{y}))(f(\mathbf{y}|\theta_1) - \eta f(\mathbf{y}|\theta_0)) d\mathbf{y} = 0$$

which is only possible if $\delta(\mathbf{y}) = \delta^*(\mathbf{y})$ with, perhaps, an exception of a set of points with zero probability whenever $f(\mathbf{y};\theta_1) \neq \eta f(\mathbf{y}|\theta_0)$ or on the boundary $\partial\Gamma = \{\mathbf{y} : f(\mathbf{y}|\theta_1) = \eta f(\mathbf{y}|\theta_0)\}$. This implies that δ must be of the same form in (1).

Existence: When we deal with finding a likelihood ratio detector for a fixed size, we will work under $\mathcal{H}_0$, and we only consider those $\mathbf{y}$ such that $f(\mathbf{y}|\theta_0) \neq 0$.

We need to find $\gamma(\mathbf{y})$ and η such that

$$\Pr(L(\mathbf{Y}) > \eta; \theta_0) + \int_{\mathbf{y}:L(\mathbf{Y})=\eta} \gamma(\mathbf{y}) d\mathbf{y} = \alpha$$

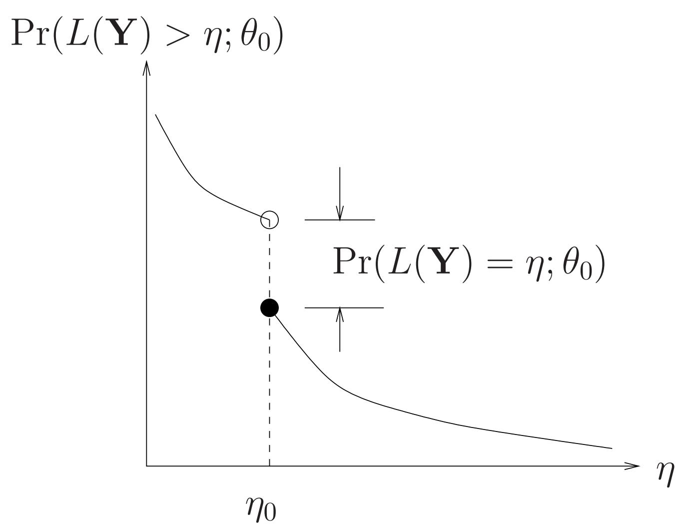
Treating the likelihood ratio $L(\mathbf{Y})$ as a random variable, then

$$\beta(\eta) \triangleq \Pr(L(\mathbf{Y}) > \eta; \theta_0) = 1 - \Pr(L(\mathbf{Y}) \leq \eta; \theta_0)$$

is the complimentary distribution function and is right continuous and monotonically decreasing. Hence for any α, we can always find the smallest η_0 such that $\Pr(L(\mathbf{Y}) > \eta_0; \theta_0) \leq \alpha$. When the equality holds, $\Pr(L(\mathbf{Y}) = \eta; \theta_0) = 0$, and we have $\eta = \eta_0$ and γ arbitrary. Otherwise, we solve for γ from

$$\alpha = \beta(\eta_0) + \gamma \Pr(L(\mathbf{Y}) = \eta; \theta_0),$$

which gives the desired threshold η_0 and randomization probability γ.



Example: Gaussian Signal in Gaussian Noise

We consider the noncoherent detection of Gaussian signal in Gaussian noise. The simple binary hypotheses are defined by

$$\mathcal{H}_i : Y_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_i^2), \quad i = 1, 2, k = 1, \cdots n$$

where σ_i^2 are known. Denote $\mathbf{Y} = (Y_1, \cdots, Y_n)$. The log-likelihood ratio is given by

$$l(\mathbf{Y}) = n \log \frac{\sigma_0}{\sigma_1} + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} ||\mathbf{Y}||^2$$

Thus the log-likelihood detector is an "energy detector"

$$\sum_{k=1}^{n} Y_i^2 \gtrless \tau.$$

For the NP detector of size α, τ should be such that

$$\Pr(\sum_{k=1}^{n} Y_i^2 \geq \tau | \mathcal{H}_0) = \alpha$$

The energy $E = \sum_{k=1}^{n} Y_i^2$ has the χ_n^2 distribution under $\mathcal{H}_i$ can be expressed as

$$\Pr(E < \tau | \mathcal{H}_i) = \Gamma(\frac{n}{2}; \frac{\tau}{2\sigma_i^2})$$

where $\Gamma(x; t)$ is the incomplete gamma function

$$\Gamma(x; t) \triangleq \frac{1}{\Gamma(x)} \int_0^t e^{-u} u^{x-1} du$$

Thus the size α NP detector is given by the threshold

$$\tau^* = 2\sigma_0^2 \Gamma^{-1}(\frac{n}{2}; 1 - \alpha)$$

and the ROC curve is given by

$$P_D(\alpha) = 1 - \Gamma(\frac{n}{2}; \frac{\sigma_0^2}{\sigma_1^2} \tau^*)$$

The Coin Tossing Problem Revisited

Suppose that we want to find the NP detector of size $\alpha = 0.1$. Under $\mathcal{H}_0$, $L(\mathbf{y})$ has the following probability mass function:

$$\Pr(L(T) = 0; \theta_0) = \Pr(L(H) = 2; \theta_0) = 0.5$$

Hence

$$\eta_0 = \min_{\Pr(L(Y) > \eta; \theta_0) \leq 0.1} \eta = 2.$$

The NP detector is then given by

$$\delta^*(y) = \begin{cases} 0 & L(y) < 2 \\ \gamma_0 = 0.2 & L(y) = 2 \end{cases}$$

This is the same detector as δ_4.

Example: A UMP Detector

The Binary Composite Hypotheses

$$\mathcal{H}_0 \ : \ Y \sim f(y|\theta) = \begin{pmatrix} n \\ y \end{pmatrix} \theta^y (1-\theta)^{n-y}, \quad \theta \in [0, \theta_0]$$
$$\mathcal{H}_1 \ : \ Y \sim f(y|\theta), \quad \theta \in (\theta_0, 1]$$

Let $\alpha = \sum_{k=0}^{K_0} \begin{pmatrix} n \\ k \end{pmatrix} \theta_0^k (1-\theta_0)^{n-k}$. Does there exist a UMP detector of size α?

If a UMP detector exists, it must be an NP detector for the simple hyptheses $\mathcal{H}_i : Y \sim f(y|\theta_i), \theta_1 > \theta_0$. From the NP lemma, we consider the test based on the likelihood ratio

$$L(y) = \rho^y (\frac{1-\theta_1}{1-\theta_0})^n, \quad \rho = \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} > 1$$

The threshold is chosen to be the the smallest η such that

$$\Pr\{\rho^Y (\frac{1-\theta_1}{1-\theta_0})^n > \eta; \theta_0\} \leq \alpha$$

which is equivalent to finding the smallest K such that

$$\Pr\{Y > K; \theta_0\} \leq \alpha \to K = K_0$$

The size α detector is given by

$$\delta(y) = \begin{cases} 1 & y > K_0 \\ 0 & o.w. \end{cases}$$

Note that the size of δ for the original hypotheses is also α and that the threshold does not depend on θ_1. Therefore, δ is UMP.

Remarks When $L(y)$ is a monotonic function of y, the detector can be defined by the test on y directly.

Monotone Likelihood Ratio

Monotone Likelihood Ratio The real-parameter family $f(\mathbf{y}|\theta)$ is said to have monotone likelihood ratio if for any $\theta < \theta'$, the distributions P_θ and $P_{\theta'}$ are distinct and the likelihood ratio

$$L(\mathbf{y};\theta',\theta) \triangleq \frac{f(\mathbf{y}|\theta')}{f(\mathbf{y}|\theta)}$$

is a nondecreasing function of some real valued function $T(\mathbf{y})$.

Example: Consider $\mathbf{y} \sim \mathcal{N}(\theta\boldsymbol{\mu}, \sigma^2\mathbf{I})$. We have

$$f(\mathbf{y}|\theta) = C\exp\{-\frac{||\mathbf{y}-\theta\boldsymbol{\mu}||^2}{2\sigma^2}\}$$

where C is a constant. The likelihood ratio is given by

$$\begin{aligned} L(\mathbf{y};\theta',\theta) &\triangleq \frac{f(\mathbf{y}|\theta')}{f(\mathbf{y}|\theta)} = \frac{\exp\{-\frac{1}{2\sigma^2}||\mathbf{y}-\theta'\boldsymbol{\mu}||^2\}}{\exp\{-\frac{1}{2\sigma^2}||\mathbf{y}-\theta\boldsymbol{\mu}||^2\}} \\ &= \exp\{\frac{2(\theta'-\theta)\mathbf{y}^T\boldsymbol{\mu} + (\theta^2-(\theta')^2)||\boldsymbol{\mu}||^2}{2\sigma^2}\} = \exp\{\frac{(\theta'-\theta)\mathbf{y}^T\boldsymbol{\mu}}{\sigma^2} + \text{constant}\} \end{aligned}$$

For any $\theta' > \theta$, $L(\mathbf{y})$ is nondecreasing function of $T(\mathbf{y}) = \mathbf{y}^T\boldsymbol{\mu}$. Therefore, $f(\mathbf{y}|\theta)$ has monotonic likelihood ratio.

Example: Consider the one-parameter exponential family

$$f(\mathbf{y}|\theta) = c(\theta)\exp\{Q(\theta)T(\mathbf{y})\}h(\mathbf{y}).$$

The likelihood ratio with $\theta' > \theta$ is

$$L(\mathbf{y}) = \frac{c(\theta')}{c(\theta)}\exp\{[Q(\theta')-Q(\theta)]T(\mathbf{y})\}$$

which is monotone if Q is monotone.

Karlin Rubin Theorem

One-sided Hypotheses Testing

The one-sided hypotheses testing problem is defined by

$$\begin{aligned} \mathcal{H}_0 : \quad & \mathbf{y} \sim f(\mathbf{y}|\theta) \quad \theta \leq \theta_* \\ \mathcal{H}_1 : \quad & \mathbf{y} \sim f(\mathbf{y}|\theta) \quad \theta > \theta_*. \end{aligned}$$

The Karlin Rubin Theorem: Let θ be a real parameter and let $f(\mathbf{y}|\theta)$ have monotone likelihood ratio in $T(\mathbf{y})$. For testing the one-sided hypotheses, there exists a size α UMP detector of the form

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & T(\mathbf{y}) > \tau \\ \gamma & T(\mathbf{y}) = \tau \\ 0 & T(\mathbf{y}) < \tau \end{cases} \tag{2}$$

where τ and γ are determined by the size constraint

$$\mathbb{E}_{\theta_*}(\delta(\mathbf{y})) = \int \delta(\mathbf{y}) f(\mathbf{y}|\theta_*) d\mathbf{y} = \alpha$$

Remark:

- The likelihood ratio test is now changed to the test of statistic $T(\mathbf{y})$.

Proof and Remarks

1. $\mathcal{H}_0' : \theta = \theta_*$ vs. $\mathcal{H}_1' : \theta = \theta_1 > \theta_*$.
   For any $\theta_1 > \theta_*$, it follows from the Neyman-Pearson lemma that there exists a most powerful detector δ^* of size α. Because $L(\mathbf{y})$ is monotone in $T(\mathbf{y})$, any test on $T(\mathbf{y})$ of the form (2) is also of the form (1), and therefore most powerful for its size.
2. $\mathcal{H}_0' : \theta = \theta_*$ vs. $\mathcal{H}_1 : \theta \in (\theta_*, \infty)$.
   Since τ and γ in (2) are only functions of θ_*, δ^* is UMP for $\mathcal{H}_0'$ vs. $\mathcal{H}_1$. What remains to be shown is that δ^* has size α under $\mathcal{H}_0$.
3. $\mathcal{H}_0$ vs. $\mathcal{H}_1$.
   For any $\theta'' > \theta'$, δ^* is the size $\alpha' \triangleq \mathbb{E}_{\theta'}(\delta^*(\mathbf{Y}))$ NP detector for the simple hypotheses test θ' vs. θ''. Since $\delta(\mathbf{y}) = \alpha'$ is a size α' detector with power α', we must have

   $$\mathbb{E}_{\theta''}(\delta^*(\mathbf{Y})) \geq \alpha' = \mathbb{E}_{\theta'}(\delta^*(\mathbf{Y}))$$

   which implies that $\mathbb{E}_{\theta}(\delta^*(\mathbf{Y}))$ is a nondecreasing function of θ. Therefore

   $$P_{\mathsf{F}}(\delta^*; \theta) \leq P_{\mathsf{F}}(\delta^*; \theta_*) = \alpha, \forall \theta \leq \theta^*$$

Remarks

- The monotone assumption of the likelihood ratio imposes strong ordering of a family of distributions. Because of this, the test statistic does not depend on θ. Without this assumption, neither step (1) nor step (2) is valid. However, the monotonic nature of the power function used in (3) is true general.

Example

One-sided Hypotheses

Let $Y \sim f(y|\theta) = \theta e^{-\theta y}$ for $y > 0$. Consider binary hypotheses

$$\mathcal{H}_0 : \theta \leq \theta_0, \quad \mathcal{H}_1 : \theta > \theta_0$$

For any $\theta' > \theta$, the likelihood ratio is given by

$$L(y) = \frac{\theta'}{\theta} e^{-(\theta' - \theta)y}$$

which is monotone with respect to $T(y) = -y$. For the one sided hypotheses, the UMP is given by

$$\delta(y) = \begin{cases} 1 & y < \tau \\ 0 & o.w. \end{cases}$$

To obtain the threshold, we impose the condition on size

$$\alpha = \int_0^\tau \theta_0 e^{-\theta_0 y} dy \to \tau = \frac{1}{\theta_0} \ln \frac{1}{1-\alpha}$$

Two-sided Hypotheses:

- $\mathcal{H}_0 : \theta = \theta_0, \quad \mathcal{H}_1 : \theta \neq \theta_0.$
- $\mathcal{H}_0 : \theta \in [\theta_1, \theta_2], \quad \mathcal{H}_1 : \theta < \theta_1$ or $\theta > \theta_2$
- $\mathcal{H}_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2, \quad \mathcal{H}_1 : \theta \in (\theta_1, \theta_2)$

UMP detector cannot exist for the first two cases. There exists a UMP detector for the third case.

UMP Detector for Two-sided Hypothesis

Theorem

Consider the one-parameter exponential family

$$f(y|\theta) = h(x)\exp\{a(\theta)T(y) - b(\theta)\}$$

with nondecreasing $a(\theta)$. There exists a UMP detector for the two-sided binary hypotheses

$$\mathcal{H}_0 : \theta \leq \theta_1 \text{ or } \theta \geq \theta_2, \quad \mathcal{H}_1 : \theta \in (\theta_1, \theta_2)$$

of the form

$$\delta(y) = \begin{cases} 1 & c_1 < T(y) < c_2 \\ \gamma_i & T(y) = c_i \\ 0 & o.w. \end{cases}$$

where $c_1 < c_2$ and γ_i are determined by

$$\mathbb{E}_{\theta_1}(\delta(Y)) = \mathbb{E}_{\theta_2}(\delta(Y)) = \alpha$$

Proof: See Lehmann: Testing Statistical Hypotheses.

Example For the two-sided test of the exponential distribution $f(y|\theta) = \theta e^{-\theta y}$, the UMP detector is given by

$$\delta(y) = \begin{cases} 1 & \tau_1 < y < \tau_2 \\ 0 & o.w. \end{cases}$$

where τ_i are determined by

$$\int_{\tau_1}^{\tau_2} \theta_1 e^{-\theta_1 y} dy = \int_{\tau_1}^{\tau_2} \theta_2 e^{-\theta_2 y} dy = \alpha$$

Locally Most Powerful Test

Intuition Given $\mathbf{Y} \sim f(\mathbf{y}|\theta)$, consider the one-sided hypotheses testing

$$\mathcal{H}_0 : \quad \theta \leq \theta_* \quad \text{vs.} \quad \theta > \theta_*.$$

Assume that the following expansion around θ_* is valid

$$P_{\mathsf{D}}(\delta;\theta) \triangleq \mathbb{E}(\delta(\mathbf{Y})) = P_{\mathsf{D}}(\delta;\theta_*) + (\theta-\theta_*)P_D'(\delta;\theta_*) + O((\theta-\theta_*)^2)$$

Under the constraint that $P_{\mathsf{D}}(\delta;\theta_*) \leq \alpha$, maximizing the power around θ_* is equivalent to maximze $P_D'(\delta;\theta_*)$.

Locally Most Powerful Test

A test δ^* is a locally most powerful test of size α if for any test δ for which $\mathbb{E}_{\theta_*}(\delta) = \alpha$, $P_D'(\delta^*,;\theta_*) > P_D'(\delta,\theta_*)$.

Suppose that the distribution of $\mathbf{Y}$ is such that the power function of any test δ is such that

$$P_D'(\delta;\theta) = \frac{d}{d\theta}P_D(\delta;\theta) = \int \delta(\mathbf{Y})\frac{\partial}{\partial\theta}f(\mathbf{y}|\theta)d\mathbf{y}$$

The problem is equivalent to testing $\mathbf{Y} \sim f(\mathbf{y}|\theta_*)$ vs. $\mathbf{Y} \sim \frac{\partial}{\partial\theta}f(\mathbf{y}|\theta_*)$. From the NP lemma, the locally most powerful test is given by

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \frac{\partial}{\partial\theta}\log f(\mathbf{y}|\theta_*) > \eta, \\ \gamma & \frac{\partial}{\partial\theta}\log f(\mathbf{y}|\theta_*) = \eta, \\ 0 & o.w. \end{cases} \tag{3}$$

where γ and η satisfy the size constraint.

Locally Most Powerful Test: Example

Consider n i.i.d. Cauchy random variables

$$Y_i \sim f(y|\theta) = \frac{1}{\pi(1+(y-\theta)^2)}$$

for the one-sided test $\mathcal{H}_0 : \theta \leq 0$ vs. $\mathcal{H}_1 : \theta > 0$.

- We can check that there is no UMP for $\theta = 0$ vs. $\theta > 0$.
- To find the LMP test, we consider

  $$\frac{\partial}{\partial \theta} \log f(\mathbf{y}|\theta) = \sum_{i=1}^{n} \frac{2(y_i - \theta)}{1 + (y_i - \theta)^2}.$$

  The LMP is then given by

  $$\delta^*(\mathbf{y}) = \begin{cases} 1 & \sum_{i=1}^{n} \frac{2y_i}{1+y_i^2} > \eta \\ 0 & o.w. \end{cases}$$

- Let $T_i = \frac{2y_i}{1+y_i^2}$. It can be shown that $\mathbb{E}(T_i) = 0$, and

  $$\mathbb{E}(T_i^2) = \frac{8}{\pi} \int_0^{\pi/2} \cos^2(\theta) \sin^2 \theta d\theta = \frac{1}{2}$$

  We can then use the Gaussian approximation to find η.

Summary

The Problem $\mathbf{Y} \sim f(\mathbf{y}|\theta), \theta \in \Lambda_0 \bigoplus \Lambda_1$

$$\mathcal{H}_0 : \theta \in \Lambda_0 \quad \text{vs.} \quad \mathcal{H}_1 : \theta \in \Lambda_1$$

UMP Detection Maximize power for all $\theta \in \Lambda_1$ subject to size constraint α, *i.e.,*

$$\max_{\delta} P_{\mathsf{D}}(\delta, \theta), \quad \forall \theta \in \Lambda_1 \quad \text{subject to} \quad P_{\mathsf{F}}(\delta, \theta) \leq \alpha \quad \forall \theta \in \Lambda_0$$

LMP Detection For one-sided test $\theta \leq \theta_*$ vs. $\theta > \theta_*$, maximize power near $\theta = \theta_*$ subject to the size constraint.

Techniques

The Neyman-Pearson Lemma For simple hypotheses, the optimal detector is a test on the likelihood ratio $L(\mathbf{y})$ with possible randomization. The threshold is chosen to satisfy size constraint α.

The Karlin-Rubin Theorem When the likelihood function is monotone with respect to $T(\mathbf{y})$, the UMP detector is given by the test on $T(\mathbf{y})$ with possible randomization. The threshold is chosen to satisfy size constraint α.

The LMP Detector The LMP detector is given by the test on the score function $\frac{\partial}{\partial \theta} \log f(\mathbf{y}|\theta)$ at $\theta = \theta_*$ with possible randomization. The threshold is chosen to satisfy size constraint α.

Notations

$P_{\mathsf{D}}(\delta, \theta)$	The probability of detection of detector δ for $\theta \in \Lambda_1$.
$P_{\mathsf{D}}(\delta)$	The probability of detection of detector δ for binary simple hypotheses. $P_D(\delta) = P_D(\delta, \theta_1)$
$P_{\mathsf{F}}(\delta, \theta)$	The probability of false alarm of detector δ for $\theta \in \Lambda_1$.
$P_{\mathsf{F}}(\delta)$	The probability of false alarm of detector δ for binary simple hypotheses. $P_F(\delta) = P_F(\delta, \theta_0)$
$P_{\mathsf{M}}(\delta, \theta)$	The probability of miss of detector δ for $\theta \in \Lambda_1$.
$f(\mathbf{y}\|\theta)$	Likelihood function of θ.
$\delta(\mathbf{y})$	Decision function. For binary hypothesis testing, $\mathcal{H}_1$ is accepted with probability $\delta(\mathbf{y})$.