Finite sample size optimality of GLR tests

George V. Moustakides
University of Patras, Greece

Outline

- Hypothesis testing and GLRT
- Randomized tests
  - Classical hypothesis testing with randomized tests
  - Alternative implementation of randomized tests
- A combined hypothesis testing/isolation problem – Optimality of GLRT
- Randomized estimators
- A combined hypothesis testing/estimation problem – Optimum GLR-type tests

Hypothesis testing and GLRT

Binary hypothesis testing (fixed sample size N):
Collection of observations $X=[x_1,\ldots,x_N]$

$$\mathbb{H}_0: X \sim f_0(X);$$
$$\mathbb{H}_1: X \sim f_1(X);$$

Decide between $\mathbb{H}_0$ and $\mathbb{H}_1$

Likelihood Ratio Rest (LRT) $\quad \dfrac{f_1(X)}{f_0(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$

Optimum according to different criteria:

Neyman-Pearson, Bayesian, Min-Max, Equalized error probabilities.

Composite binary hypothesis testing

Observations $X=[x_1,\ldots,x_N]$

$$
\begin{array}{llll}
\mathbb{H}_0 : X & \sim & f_{01}(X), & \pi_{01} \\
 & \sim & f_{02}(X), & \pi_{02} \\
 & \vdots & & \\
 & \sim & f_{0K_0}(X), & \pi_{0K_0}
\end{array}
$$

$$
\begin{array}{llll}
\mathbb{H}_1 : X & \sim & f_{11}(X), & \pi_{11} \\
 & \sim & f_{12}(X), & \pi_{12} \\
 & \vdots & & \\
 & \sim & f_{1K_1}(X), & \pi_{1K_1}
\end{array}
$$

One subcase responsible for the data

If we knew before hand the two possible subcases, for example: f_{0i} and f_{1j} then

$$\frac{f_{1j}(X)}{f_{0i}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

When this knowledge is not possible:

$$\frac{f_1(X)}{f_0(X)} = \frac{\pi_{11} f_{11}(X) + \cdots + \pi_{1K_1} f_{1K_1}(X)}{\pi_{01} f_{01}(X) + \cdots + \pi_{0K_0} f_{0K_0}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

If in addition to detection, we like to **isolate** the subcase responsible for the data X

GLRT

$$\frac{f_{1\hat{j}}(X)}{f_{0\hat{i}}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

$$\hat{i} = \arg \max_{1 \le i \le K_0} f_{0i}(X)$$

$$\hat{j} = \arg \max_{1 \le j \le K_1} f_{1j}(X)$$

ML isolation

Composite binary hypothesis testing (cond.)

Observations $X=[x_1,\dots,x_N]$

$$\mathbb{H}_0 : X \sim f_0(X|\theta_0), \quad \pi_0(\theta_0)$$

$$\mathbb{H}_1 : X \sim f_1(X|\theta_1), \quad \pi_1(\theta_1)$$

Only detection

$$\frac{f_1(X)}{f_0(X)} = \frac{\int \pi_1(\theta_1) f_1(X|\theta_1)\, d\theta_1}{\int \pi_0(\theta_0) f_0(X|\theta_0)\, d\theta_0} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

If we like, in addition to detection to **estimate** the parameter vector θ_i

GLRT

$$\frac{f_1(X|\hat{\theta}_1)}{f_0(X|\hat{\theta}_0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

$$\hat{\theta}_0 = \arg\max_{\theta_0} f_0(X|\theta_0)$$

$$\hat{\theta}_1 = \arg\max_{\theta_1} f_1(X|\theta_1)$$

ML estimate

Optimality of GLRT

Wald (1943)

Le Cam (1953)

Lehmann (1959)

Hoeffding (1965)

Neyman (1965)

⋮

Asymptotic optimality under **regularity conditions**. As the number of samples increases without limit, performance of GLRT approaches the optimum test (with true parameters considered known).

Randomized tests

Binary hypothesis testing (fixed sample size N):
Collection of observations $X = [x_1, \ldots, x_N]$

$$\mathbb{H}_0 : X \sim f_0(X);$$

$$\mathbb{H}_1 : X \sim f_1(X);$$

Decide between $\mathbb{H}_0$ and $\mathbb{H}_1$

Deterministic tests

$X \in \mathbb{R}^N$

Two complementary sets in $\mathbb{R}^N$: A_0, A_1 with $A_1^c = A_0$

When $X \in A_i$ decide for $\mathbb{H}_i$.

Randomized tests

Two complementary probabilities: $\delta_0(X), \delta_1(X)$ with $\delta_0(X) + \delta_1(X) = 1$.

In a **random game** select $\mathbb{H}_i$ with probability $\delta_i(X)$.

Randomized more general than Deterministic

$$\delta_i(X) = 1_{A_i}(X)$$

When $X \in A_1$ then we decide with **probability 1** for $\mathbb{H}_1$ and with probability 0 for $\mathbb{H}_0$ (the equivalent of a deterministic decision).

Neyman-Pearson:

$$\max_{\delta_0,\delta_1} \mathbb{P}[d = 1|\mathbb{H}_1]; \quad \text{subject to} \quad \mathbb{P}[d = 1|\mathbb{H}_0] \leq \alpha$$

$$\int \delta_1(X) f_1(X)\, dX \quad - \quad \lambda \int \delta_1(X) f_0(X)\, dX$$

$$= \int \delta_1(X)[f_1(X) - \lambda f_0(X)]\, dX$$

$$\leq \int \max\{f_1(X) - \lambda f_0(X), 0\}\, dX$$

$$\delta_1(X) = \begin{cases} 1 & f_1(X) - \lambda f_0(X) > 0 \\ 0 & f_1(X) - \lambda f_0(X) < 0 \\ \gamma & f_1(X) - \lambda f_0(X) = 0. \end{cases}$$

$$\delta_0(X) = 1 - \delta_1(X) = \begin{cases} 0 & f_1(X) - \lambda f_0(X) > 0 \\ 1 & f_1(X) - \lambda f_0(X) < 0 \\ 1-\gamma & f_1(X) - \lambda f_0(X) = 0. \end{cases}$$

Likelihood Ratio Rest (LRT) $$\frac{f_1(X)}{f_0(X)} \overset{\mathbb{H}_1}{\underset{\mathbb{H}_0}{\gtreqless}} \lambda$$

Composite binary hypothesis testing

Observations $X = [x_1, \ldots, x_N]$

$$
\begin{array}{llllll}
\mathbb{H}_0 : X & \sim & f_{01}(X), & \pi_{01} & & \delta_{01}(X) \\
 & \sim & f_{02}(X), & \pi_{02} & & \delta_{02}(X) \\
 & \vdots & & & & \vdots \\
 & \sim & f_{0K_0}(X), & \pi_{0K_0} & & \delta_{0K_0}(X) \\
 & & & & & \\
\mathbb{H}_1 : X & \sim & f_{11}(X), & \pi_{11} & & \delta_{11}(X) \\
 & \sim & f_{12}(X), & \pi_{12} & & \delta_{12}(X) \\
 & \vdots & & & & \vdots \\
 & \sim & f_{1K_1}(X), & \pi_{1K_1} & & \delta_{1K_1}(X)
\end{array}
$$

$$\delta_{01}(X)\ \delta_{02}(X)\cdots\delta_{0K_0}(X)\quad \delta_{11}(X)\ \delta_{12}(X)\cdots\delta_{1K_1}(X)$$

$$\sum_{i=0,1}\sum_{j=1}^{K_i}\delta_{ij}(X)=1$$

In a random game, with probability $\delta_{ij}(X)$:
decide for Hypothesis $\mathbb{H}_i$ **and** subcase j.

We **SIMULTANEOUSLY** Detect and Isolate

Alternative two-step implementation

$$\underbrace{\delta_{01}(X), \delta_{02}(X), \ldots, \delta_{0K_0}(X)}_{\delta_0(X)} \quad \underbrace{\delta_{11}(X), \delta_{12}(X), \ldots, \delta_{1K_1}(X)}_{\delta_1(X)}$$

$$q_{ij}(X) = \frac{\delta_{ij}(X)}{\delta_i(X)}$$

$$\delta_0(X) + \delta_1(X) = 1$$

$$q_{01}(X) + \cdots + q_{0K_0}(X) = 1$$

$$q_{11}(X) + \cdots + q_{1K_1}(X) = 1$$

$$\delta_0(x), \delta_1(X), [q_{01}(X), \ldots, q_{0K_0}(X)], [q_{11}(X), \ldots, q_{1K_1}(X)]$$

- With probabilities $\delta_0(X)$,$\delta_1(X)$ decide between $\mathbb{H}_0$,$\mathbb{H}_1$.
- Given we selected $\mathbb{H}_i$, isolate among the subcases of $\mathbb{H}_i$ with probabilities $q_{i1}(X)$, $q_{i2}(X)$,..., $q_{iK_i}(X)$.

Optimality of GLRT

$$\mathbb{H}_0 : X \sim f_{01}(X), \pi_{01} \quad \mathbb{H}_1 : X \sim f_{11}(X), \pi_{11}$$
$$\sim f_{02}(X), \pi_{02} \quad \sim f_{12}(X), \pi_{12}$$
$$\vdots \quad \vdots$$
$$\sim f_{0K_0}(X), \pi_{0K_0} \quad \sim f_{1K_1}(X), \pi_{1K_1}$$

A Combined Detection / Isolation problem

Following a Neyman-Pearson like approach, we propose

$$\max \mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_1]$$

subject to the “false alarm” constraint

$$\mathbb{P}[\text{Miss-detection/isolation}|\mathbb{H}_0] \leq \alpha$$

THEOREM: The test that solves the combined detection/isolation problem is the following

$$\frac{\max\limits_{1\le j\le K_1} \pi_{1j} f_{1j}(X)}{\max\limits_{1\le j\le K_0} \pi_{0j} f_{0j}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

(Randomization probability γ)

Threshold λ and randomization probability γ are selected to satisfy the false alarm constraint with equality.

PROOF

$$\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_i]$$

$$= \sum_{j=1}^{K_i} \mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_{ij}]\pi_{ij}$$

$$\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_{ij}]$$

$$= \int \delta_i(X) q_{ij}(X) f_{ij}(X)\, dX.$$

Constrained optimization problem: Lagrange multiplier

$$\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_1] - \lambda\mathbb{P}[\text{Miss-detection/isolation}|\mathbb{H}_0]$$

$$= \mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_1] + \lambda\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_0] - \lambda$$

$$\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_1] + \lambda\mathbb{P}[\text{Correct-detection/isolation}|\mathbb{H}_0]$$

$$= \int \delta_1(X)\left\{\sum_{j=1}^{K_1} q_{1j}(X)\pi_{1j}f_{1j}(X)\right\}dX + \lambda \int \delta_0(X)\left\{\sum_{j=1}^{K_0} q_{0j}(X)\pi_{0j}f_{0j}(X)\right\}dX$$

$$\leq \int \delta_1(X) \max_{1\leq j\leq K_1}\{\pi_{1j}f_{1j}(X)\}dX + \lambda \int \delta_0(X) \max_{1\leq j\leq K_0}\{\pi_{0j}f_{0j}(X)\}dX$$

$$= \int \Big[\delta_1(X) \max_{1\le j\le K_1}\{\pi_{1j} f_{1j}(X)\}$$
$$+ \delta_0(X)\lambda \max_{1\le j\le K_0}\{\pi_{0j} f_{0j}(X)\}\Big] dX$$

$$\le \int \max\left\{ \max_{1\le j\le K_1}\{\pi_{1j} f_{1j}(X)\}, \lambda \max_{1\le j\le K_0}\{\pi_{0j} f_{0j}(X)\} \right\} dX$$

$$\frac{\displaystyle\max_{1\le j\le K_1} \pi_{1j} f_{1j}(X)}{\displaystyle\max_{1\le j\le K_0} \pi_{0j} f_{0j}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

If priors not known, assume equiprobable subcases:

$$\frac{\displaystyle\max_{1\le j\le K_1} K_1^{-1} f_{1j}(X)}{\displaystyle\max_{1\le j\le K_0} K_0^{-1} f_{0j}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda \Rightarrow \boxed{\frac{\displaystyle\max_{1\le j\le K_1} f_{1j}(X)}{\displaystyle\max_{1\le j\le K_0} f_{0j}(X)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda \frac{K_1}{K_2}}$$

Classical GLRT

Randomized estimators

Observations $X=[x_1,\ldots,x_N]$

$$X \sim f(X|\theta),\ \pi(\theta)$$

We would like to obtain an estimate $\hat{\theta}$ of θ

Deterministic estimator: $\hat{\theta} = G(X)$

Randomized estimator:

$$\delta_{\hat{\theta}}(X) \Rightarrow \delta(\hat{\theta}|X); \qquad \int \delta(\hat{\theta}|X)\, d\hat{\theta} = 1;$$ **a PDF!**

How do we obtain a randomized estimate?

We generate a r.v. $\hat{\theta}$ distributed according to $\delta(\hat{\theta}|X)$

"Deterministic" estimator $$\delta(\hat{\theta}|X) = \mathrm{Dirac}\left\{\hat{\theta} - G(X)\right\}$$

Optimum Bayes estimation

If true parameter is θ and estimate $\hat{\theta}$, cost is: $\mathcal{C}(\hat{\theta},\theta)$

Select randomized estimator to **minimize average cost**

$$\mathbb{E}[\mathcal{C}(\hat{\theta},\theta)] = \iiint \mathcal{C}(\hat{\theta},\theta)\delta(\hat{\theta}|X)f(X|\theta)\pi(\theta)\,dX\,d\hat{\theta}\,d\theta$$

$$= \int\left[\int \delta(\hat{\theta}|X)\left\{\underbrace{\int \mathcal{C}(\hat{\theta},\theta)f(X|\theta)\pi(\theta)\,d\theta}_{\mathcal{D}(\hat{\theta},X) \geq \min_U \mathcal{D}(U,X)}\right\}d\hat{\theta}\right]dX$$

$$\geq \int \min_U \mathcal{D}(U,X)\,dX$$

$$\delta(\hat{\theta}|X) = \mathrm{Dirac}\left\{\hat{\theta} - \arg\min_U \mathcal{D}(U,X)\right\}$$

Combined detection/estimation

Observations $X=[x_1,\ldots,x_N]$

$$\mathbb{H}_0 : X \sim f(X|\theta=0)$$

$$\mathbb{H}_1 : X \sim f(X|\theta),\ \pi(\theta)$$

- Decide between $\mathbb{H}_0, \mathbb{H}_1$
- Whenever we decide in favor of $\mathbb{H}_1$ we also like to provide an estimate $\hat{\theta}$ of θ

Detection/estimation procedure: Two-step strategy

$$\delta_0(X), \delta_1(X), q(\hat{\theta}|X)$$

Following a Neyman-Pearson like approach, we propose

$$\min\ \mathbb{E}[\mathcal{C}(\hat{\theta},\theta)|\mathbb{H}_1]$$

subject to: $\mathbb{P}[\text{Miss-detection}|\mathbb{H}_0] \le \alpha$

$$\mathbb{P}[\text{Miss-detection}|\mathbb{H}_0] = \int \delta_1(X) f(X|0)\,dX$$

Cost computation

Correct Hypothesis:

$$\int\!\!\int\!\!\int \mathcal{C}(\hat{\theta},\theta)\delta_1(X)q(\hat{\theta}|X)\pi(\theta)f(X|\theta)\,d\theta\,d\hat{\theta}\,dX$$

Wrong Hypothesis:

$$\int\!\!\int \mathcal{C}(0,\theta)\delta_0(X)\pi(\theta)f(X|\theta)\,d\theta\,dX$$

$$\iiint \mathcal{C}(\hat\theta,\theta)\delta_1(X)q(\hat\theta|X)\pi(\theta)f(X|\theta)d\theta\,d\hat\theta\,dX$$
$$+\iint \mathcal{C}(0,\theta)\delta_0(X)\pi(\theta)f(X|\theta)d\theta\,dX$$
$$+\lambda\int \delta_1(X)f(X|0)dX$$

$$\geq \int \delta_1(X)\min_U\left\{\overbrace{\int \mathcal{C}(U,\theta)\pi(\theta)f(X|\theta)d\theta}^{\mathcal{D}(U,X)}\right\}dX$$
$$+\int \delta_0(X)\left\{\underbrace{\int \mathcal{C}(0,\theta)\pi(\theta)f(X|\theta)d\theta}_{\mathcal{D}(0,X)}\right\}dX$$
$$+\lambda\int \delta_1(X)f(X|0)dX$$

$$= \int \Big\{ \delta_1(X)[\lambda f(X|0) + \min_U \mathcal{D}(U,X)] + \delta_0(X)\mathcal{D}(0,X) \Big\} dX$$

$$\geq \int \min \Big\{ \lambda f(X|0) + \min_U \mathcal{D}(U,X), \mathcal{D}(0,X) \Big\} dX$$

We decide in favor of $\mathbb{H}_1$ when

$$\mathcal{D}(0,X) > \lambda f(X|0) + \min_U \mathcal{D}(U,X)$$

The final test is: $$\frac{\mathcal{D}(0,X) - \min_U \mathcal{D}(U,X)}{f(X|0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

$$\mathcal{C}(U,X) = \int \mathcal{C}(U,\theta) f(X|\theta)\pi(\theta) d\theta$$

Examples

MAP estimator:

$$\mathcal{C}(\hat{\theta}, \theta) = \begin{cases} 0 & \text{for } \|\hat{\theta} - \theta\| \leq \Delta \ll 1 \\ 1 & \text{otherwise} \end{cases}$$

Optimum test:

$$\frac{\max_{\theta} \pi(\theta) f(X|\theta)}{f(X|0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

If prior $\pi(\theta)$ is uniform we recover the classical GLRT

$$\frac{\max_{\theta} f(X|\theta)}{f(X|0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

MMSE estimator:

$$\mathcal{C}(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|^2$$

Optimum test:

$$\|\hat{\theta}_{\mathrm{MMSE}}\|^2 \frac{\int \pi(\theta) f(X|\theta)\, d\theta}{f(X|0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

$$\hat{\theta}_{\mathrm{MMSE}} = \frac{\int \theta \pi(\theta) f(X|\theta)\, d\theta}{\int \pi(\theta) f(X|\theta)\, d\theta} = \mathbb{E}[\theta|X]$$

MEDIAN estimator:

$$\mathcal{C}(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

Optimum test:

$$\frac{\displaystyle\int_0^{\hat{\theta}_{\mathrm{MEDIAN}}} \theta \pi(\theta) f(X|\theta)\, d\theta}{f(X|0)} \underset{\mathbb{H}_0}{\overset{\mathbb{H}_1}{\gtreqless}} \lambda$$

$$\hat{\theta}_{\mathrm{MEDIAN}} = \arg\left\{ \hat{\theta} : \ \frac{\displaystyle\int_{-\infty}^{\hat{\theta}} \pi(\theta) f(X|\theta)\, d\theta}{\displaystyle\int_{-\infty}^{\infty} \pi(\theta) f(X|\theta)\, d\theta} = 0.5 \right\}$$

$$= \arg\{\hat{\theta} : \mathbb{P}[\theta \leq \hat{\theta}|X] = 0.5\}$$