Almost optimal sequential detection in multiple data streams

Georgios Fellouris

Department of Statistics

University of Illinois

Joint work with Alexander Tartakovsky

University of Michigan
Ann Arbor, May 13th, 2015
Outline

1. Simple null against simple alternative
2. A simple null against a finite number of alternatives
3. The continuous-parameter case
Sequentially testing of two simple hypotheses

Sequentially acquired observations

\[ X_1, \ldots, X_t, \ldots \text{ iid} \sim f. \]

Stop sampling as soon as possible and distinguish between

\[ H_0 : f = f_0 \quad \text{and} \quad H_1 : f = f_1. \]

Let \( \mathcal{F}_t \) be the history of observations up to time \( t \),

\[ \mathcal{F}_t = \sigma(X_s : 1 \leq s \leq t). \]
Wald’s formulation

Find an $\mathcal{F}_t$-stopping time, $T$, at which to stop sampling and an $\mathcal{F}_T$-measurable r.v., $d_T$, so that

$$\{d_T = 1\} = \{\text{Accept } \mathbb{H}_1, T < \infty\}$$
$$\{d_T = 0\} = \{\text{Accept } \mathbb{H}_0, T < \infty\}.$$  

A sequential test is such a pair $(T, d_T)$.

Goal: Minimize $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ in

$$C_{\alpha,\beta} = \{(T, d_T) : \mathbb{P}_0(d_T = 1) \leq \alpha \text{ and } \mathbb{P}_0(d_T = 1) \leq \beta\}.$$
Let $Z_t$ the log-likelihood ratio of the first $t$ observations:

$$Z_t := \sum_{s=1}^{t} \log \frac{f_1(X_s)}{f_0(X_s)}, \quad Z_0 := 0.$$ 

**Sequential Probability Ratio Test (SPRT)**

Let $\alpha, \beta$ such that $\alpha + \beta < 1$ and $A, B > 0$ be fixed thresholds. Define

$$S = \inf\left\{ t \geq 1 : Z_t \notin (-A, B) \right\}$$

$$d_S = \begin{cases} 
0, & \text{if } Z_S \leq -A \\
1, & \text{if } Z_S \geq B 
\end{cases}$$
Suppose that $A, B$ are selected so that

$$\mathbb{P}_0(d_S = 1) = \alpha \quad \text{and} \quad \mathbb{P}_1(d_S = 0) = \beta.$$ 

Then,

$$\mathbb{E}_0[S] = \inf_{(T, d_T) \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] \quad \text{and} \quad \mathbb{E}_1[S] = \inf_{(T, d_T) \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_1[T].$$
Suppose that $\beta | \log \alpha | + \alpha | \log \beta | \to 0$ and $\mathbb{E}_i[Z_i^2] < \infty$, $i = 0, 1$.

Then, as $\alpha, \beta \to 0$,

$$
\mathbb{E}_1[S] = \frac{1}{I_1} \left[ | \log \alpha | + \rho_1 + \log \delta_1 + o(1) \right]
$$

$$
\mathbb{E}_0[S] = \frac{1}{I_0} \left[ | \log \beta | + \rho_0 + \log \delta_0 + o(1) \right].
$$

$I_1 \:= D(f_1 || f_0)$ and $I_0 \:= D(f_0 || f_1)$ are the K-L information numbers.

Let $H_1$ be the asymptotic distribution of the overshoot of $Z^k$ under $\mathbb{P}_1$. Then

$$
\rho_1 := \int x H_1(dx) \quad \text{and} \quad \delta_1 := \log \int e^{-x} H_1(dx).
$$

Let $H_0$ be the asymptotic distribution of the overshoot of $-Z$ under $\mathbb{P}_0$. Then

$$
\rho_0 := \int x H_0(dx) \quad \text{and} \quad \delta_0 := \log \int e^{-x} H_0(dx).
$$
Sequentially testing a simple null against a finite number of alternatives

Sequentially acquired observations

\[ X_1, \ldots, X_t, \ldots \overset{iid}{\sim} f. \]

Stop sampling as soon as possible and distinguish between

\[ \mathbb{H}_0 : f = f_0 \quad \text{vs} \quad \mathbb{H}_1 : f \in \{f_1, \ldots, f_M\}. \]

Let \( \mathcal{F}_t \) be the history of observations up to time \( t \),

\[ \mathcal{F}_t = \sigma(X_s : 1 \leq s \leq t). \]

Find \( (T, d_T) \), where \( T \) is an \( \mathcal{F}_t \)-stopping time and \( d_T \) an \( \mathcal{F}_T \)-measurable r.v.

\[ \{d_T = i\} = \{\text{Accept } \mathbb{H}_i, T < \infty\}, \quad i = 0, 1. \]
1st Motivation: The multichannel problem
1st Motivation: The multichannel problem

Observations are collected from $K$ independent sources so that

$$X_t = (X_t^1, \ldots, X_t^K).$$

For every sensor $k$, the true density is $f_k$ and

$$X_1^k, \ldots, X_t^k, \ldots \sim f^k = \begin{cases} f_0^k, & \text{noise} \\ f_1^k, & \text{signal} \end{cases}$$

We want to test the simple hypothesis

$$H_0 : f^k = f_0^k \quad \forall k \in \{1, \ldots, K\}$$

against

$$H_1 : f^k = \begin{cases} f_0^k, & k \notin A \\ f_1^k, & k \in A \end{cases}$$

where $A$ is an unknown subset of $\{1, \ldots, K\}$. 
\( \mathcal{A} \) is known to belong to some class of subsets of \( \{1, \ldots, K\} \), \( \mathcal{P} \).

Then, \( \mathbb{H}_1 \) contains \( M = |\mathcal{P}| \) possibilities, where \( |\mathcal{P}| \) is the size of class \( \mathcal{P} \).

When signal can be present in only one sensor, then \( |\mathcal{P}| = K \).

When signal can be present in at most \( L \) sensors, then \( |\mathcal{P}| = \sum_{k=1}^{L} \binom{K}{k} \).
2nd motivation: Discretization of a continuous alternative

Sequentially acquired observations

\[ X_1, \ldots, X_t, \ldots \overset{iid}{\sim} f \in \{f_\theta, \theta \in \Theta\} \]

Stop sampling as soon as possible and distinguish between

\[ H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta_1, \]

where \( \theta_0 \notin \Theta_1 \subset \Theta \).

Then, we would like to minimize \( \mathbb{E}_{\theta_0}[T] \) and \( \mathbb{E}_\theta[T] \) for every \( \theta \in \Theta_1 \) in

\[ C_{\alpha,\beta} = \left\{ (T, d_T) : \mathbb{P}_{\theta_0}(d_T = 1) \leq \alpha \quad \text{and} \quad \sup_{\theta \in \Theta_1} \mathbb{P}_\theta(d_T = 0) \leq \beta \right\}. \]

Approximating \( \Theta_1 \) by \( \{\theta_1, \ldots, \theta_M\} \subset \Theta_1 \) may have (computational) benefits.
A simple null against a finite number of alternatives

Sequentially acquired observations

\[ X_1, \ldots, X_t, \ldots \stackrel{iid}{\sim} f. \]

Let \( \mathcal{F}_t \) be the history of observations up to time \( t \),

\[ \mathcal{F}_t = \sigma(X_s : 1 \leq s \leq t). \]

Stop sampling as soon as possible and distinguish between

\[ \mathbb{H}_0 : f = f_0 \quad \text{vs} \quad \mathbb{H}_1 : f \in \{f_1, \ldots, f_M\}. \]

Find \((T, d_T)\), where \( T \) is an \( \mathcal{F}_t \)-stopping time and \( d_T \) an \( \mathcal{F}_T \)-measurable r.v.

\[ \{d_T = i\} = \{\text{Accept } \mathbb{H}_i, T < \infty\}, \quad i = 0, 1. \]
$\mathbb{P}_i$ is the probability measure and $\mathbb{E}_i$ the expectation when

$$f = f_i, \quad i = 0, 1, \ldots, M.$$ 

Goal: Minimize

$$\mathbb{E}_0[T] \quad \text{and} \quad \mathbb{E}_i[T], \quad i = 1, \ldots, K$$

among sequential tests in

$$\mathcal{C}_{\alpha, \beta} = \{(T, d_T) : \mathbb{P}_0(d_T = 1) \leq \alpha \quad \text{and} \quad \max_{1 \leq i \leq K} \mathbb{P}_i(d_T = 0) \leq \beta \}.$$ 

This can be done only in an asymptotic sense, i.e., as $\alpha, \beta \to 0$. 
Generalized Sequential Likelihood Ratio Test (GSLRT)

For $i = 1, \ldots, M$ let

$$\Lambda_t^i := \prod_{s=1}^t \frac{f_i(X_s)}{f_0(X_s)}, \quad Z_t^i := \log \Lambda_t^i, \quad t \in \mathbb{N}.$$ 

GSLRT

Following a maximum likelihood approach, we obtain

$$\hat{S} = \inf \left\{ t \geq 1 : \max_{1 \leq i \leq M} Z_t^i \notin (-A, B) \right\},$$

$$\{d_{\hat{S}} = 1\} = \left\{ \max_{1 \leq i \leq M} Z_{\hat{S}}^i \geq B \right\}, \quad \{d_{\hat{S}} = 0\} = \left\{ \max_i Z_{\hat{S}}^i \leq -A \right\}.$$

where $A, B > 0$ are fixed thresholds.

Weighted Sequential Likelihood Ratio Test (WSLRT)

- Recall that
  \[ \Lambda_t^i := \prod_{s=1}^{t} \frac{f_i(X_s)}{f_0(X_s)}, \quad Z_t^i := \log \Lambda_t^i, \quad t \in \mathbb{N}. \]

- We will call \( q = (q_1, \ldots, q_M) \) a weight if \( q_i > 0 \) for every \( i \).

- We will write:
  \[ \Lambda_t(q) := \sum_{i=1}^{K} q_i \Lambda_t^i \quad \text{and} \quad Z_t(q) := \log \Lambda_t(q) \]

WSLRT

\[ S = \inf \left\{ t \geq 1 : Z_t(q) \notin (-A, B) \right\} \]
\[ \{d_S = 1\} = \left\{ Z_S(q) \geq B \right\}, \quad \{d_S = 0\} = \left\{ Z_S(q) \leq -A \right\} \]

An idea that goes back to Wald (1945) in the case of a continuous parameter.
In order to treat the maximizing and the averaging approach similarly, let \( q = (q_1, \ldots, q_M) \) a weight.

We will write:

\[
\hat{\Lambda}_t(q) := \max_{1 \leq i \leq K} (q_i \Lambda^i_t) \quad \text{and} \quad \hat{Z}_t(q) = \log \hat{\Lambda}_t(q).
\]

**Weighted (WGSLRT)**

\[
\hat{S} = \inf \left\{ t \geq 1 : \hat{Z}_t(q) \notin (-A, B) \right\},
\]
\[
\{d_{\hat{S}} = 1\} = \left\{ \hat{Z}_{\hat{S}} \geq B \right\}, \quad \{d_{\hat{S}} = 0\} = \left\{ \hat{Z}_{\hat{S}} \leq -A \right\}.
\]
Controlling the error probabilities

- For any given $\alpha, \beta \in (0, 1)$, $S, \hat{S} \in C_{\alpha, \beta}$ when $A, B$ are chosen so that

$$A = |\log \beta| + \log \left( \max_{1 \leq k \leq K} q_k \right) \quad \text{and} \quad B = |\log \alpha| + \log \left( \sum_{k=1}^{K} q_k \right).$$

- Suppose also that each $Z^i$ has a non-arithmetic distribution. Then, $\mathbb{P}_0(S = 1) \sim \alpha$ when

$$B = |\log \alpha| + \log \left( \sum_{k=1}^{K} q_i \delta_i \right).$$

- Let $H_i$ the limiting distribution of the overshoot of the random walk $Z^i$ under $\mathbb{P}_i$. That is, if we set

$$T^i_a := \inf \{ t : Z^i_t \geq a \},$$

then $H_i$ is the limiting distribution of $Z^i_{T^i_a} - a$ as $a \to \infty$. Then

$$\delta_i := \log \int e^{-x} H_i(dx).$$
Asymptotic Expansions under $\mathbb{H}_1$

Suppose further

- that each $Z^i$ has a finite second moment under $\mathbb{P}_i$.
- $|\log \alpha|/|\log \beta|$ goes to some constant as $\alpha, \beta \to 0$.
- $A, B \to \infty$ so that

$$k_0 \alpha(1 + o(1)) \leq \mathbb{P}_0(d_S = 1) \leq \alpha(1 + o(1))$$
$$k_1 \beta(1 + o(1)) \leq \max_{1 \leq i \leq M} \mathbb{P}_i(d_S = 0) \leq \beta(1 + o(1))$$

for some $k_0, k_1 \in (0, 1)$.

Then, as $A, B \to \infty$ we have

$$\mathbb{E}_i[S] = \frac{1}{L_i} [B + \rho_i - \log q_i] + o(1) = \mathbb{E}_i[\hat{S}],$$

where $\rho_i$ is the limiting expected overshoot of $Z^i$ under $\mathbb{P}_i$, i.e.,

$$\rho_i := \int x H_i(dx) = \lim_{a \to \infty} \mathbb{E}_i[Z^i_{T^i_a} - a], \quad T^i_a := \inf\{n : Z^i_n \geq a\}.$$
Uniform Second-Order Asymptotic Optimality under $\mathbb{H}_1$

Suppose that the previous assumptions hold.

- If $A, B$ are selected so that $S, \hat{S} \in C_{\alpha, \beta}$, then for every $i$ we have
  \[
  E_i[S] = \inf_{(T, d_T) \in C_{\alpha, \beta}} E_i[T] + O(1) = E_i[\hat{S}]
  \]

- However, even first-order asymptotic optimality is lost when $f \not\in \{f_1, \ldots, f_M\}$.
- If $A, B$ are selected so that $P_0(S = 1) \sim \alpha \sim P_0(\hat{S} = 1)$, then
  \[
  E_i[S] = \frac{1}{I_i} \left[ |\log \alpha| + \log \left( \sum_{k=1}^{M} q_k \delta_k \right) + \rho_i - \log q_i \right] + o(1) \geq E_i[\hat{S}].
  \]
Accuracy of asymptotic approximations

First Channel

$M = 3$, exponential distribution, $\theta_1 = 0.5$, $\theta_2 = 1$, $\theta_3 = 2$
Accuracy of asymptotic approximations

\[ K = 3, \text{ exponential distribution, } \theta_1 = 0.5, \theta_2 = 1, \theta_3 = 2 \]
$M = 3$, exponential distribution, $\theta_1 = 0.5$, $\theta_2 = 1$, $\theta_3 = 2$
Asymptotic Optimality under $\mathbb{H}_0$

- Let $I^i_0 = D(f_0||f_i)$ for every $1 \leq i \leq M$ and
  
  $$I_0 = \min_{1 \leq i \leq M} I^i_0.$$  

- If there is a unique $i$ that attains $I_0$, then
  
  $$\mathbb{E}_0[S] = \frac{1}{I_0} \left[ |\log \beta| + O(1) \right]$$

- If not,
  
  $$\mathbb{E}_0[S] = \frac{1}{I_0} \left[ |\log \beta| + \Theta(\sqrt{\log B}) \right]$$

- The second-order term is not always constant.

- If $A, B$ are selected so that $S, \hat{S} \in C_{\alpha, \beta}$, then

  $$\mathbb{E}_0[S] \sim \inf_{(T,d)\in C_{\alpha, \beta}} \mathbb{E}_0[T] \sim \mathbb{E}_0[\hat{S}].$$
Remarks

- These asymptotic results are based on non-linear renewal theory (Lai and Siegmund ‘77,’79, Woodrooffe ’82, Zhang’88) and Dragalin et al. (’99,’00).

- Were known for the GSLRT (Tartakovsky, 2003).

- Here, we have shown that they hold for arbitrary weights (and both tests).

- How should one choose these weights?

- For this choice, we will show that a particular choice of weights satisfies an even stronger asymptotic optimality property.
Almost minimax?

What if we select $q$ so that

$$\max_{1 \leq i \leq M} \mathbb{E}_i[S] = \inf_{(T,d_T) \in C_{\alpha,\beta}} \max_{1 \leq i \leq M} \mathbb{E}_i[S] + o(1)?$$

This would require that $\mathbb{E}_i[S] = \mathbb{E}_j[S] + o(1)$ for every $1 \leq i, j \leq M$.

However, to have $\mathbb{E}_i[S] \sim \mathbb{E}_j[S]$ for every $1 \leq i, j \leq M$, we need

$$\mathbb{E}_i[S] \sim \frac{|\log \alpha|}{I_i} \sim \frac{|\log \alpha|}{I_j} \sim \mathbb{E}_j[S].$$

This is not possible unless $I_1 = \ldots = I_M$. 
Almost optimality with respect to a weighted expected sample size

- Let $p = (p_1, \ldots, p_K)$ a vector of positive numbers that add up to 1.
- We will try to design the proposed tests so that
  
  $$
  \sum_{k=1}^{K} p_i \mathbb{E}_i[S] = \inf_{(T,d_T) \in C_{\alpha, \beta}} \sum_{k=1}^{K} p_i \mathbb{E}_i[T] + o(1) = \sum_{k=1}^{K} p_i \mathbb{E}_i[\hat{S}].
  $$

- (Later how to choose the $p_i$’s).
- For this, we need to generalize the class of sequential tests.
Let \( q_0, q_1 \) \( M \)-dimensional vectors of positive numbers.

**WSLRT**

\[
S = \inf \left\{ t \geq 1 : Z_t(q_1) \geq B \quad \text{or} \quad Z_t(q_0) \leq -A \right\}
\]

\[
\{d_S = 1\} = \left\{ Z_S(q_1) \geq B \right\}, \quad \{d_S = 0\} = \left\{ Z_S(q_0) \leq A \right\}
\]

**WG-SLRT**

\[
\hat{S} = \inf \left\{ t \geq 1 : \hat{Z}_t(q_1) \geq B \quad \text{or} \quad \hat{Z}_t(q_0) \leq A \right\},
\]

\[
\{d_{\hat{S}} = 1\} = \left\{ \hat{Z}_{\hat{S}}(q_1) \geq B \right\}, \quad \{d_{\hat{S}} = 0\} = \left\{ \hat{Z}_{\hat{S}}(q_0) \leq A \right\}.
\]
Almost optimality

Theorem (F. & Tartakovsky 2013)

If $A, B$ are chosen so that $(S, d_S) \in C_{\alpha, \beta}$ and

$$q_1^i = p_i/L_i, \quad q_0^i = p_iL_i,$$

then as $\alpha, \beta \rightarrow 0$ so that $|\log \alpha| \sim |\log \beta|$ we have

$$\sum_{i=1}^{M} p_iE_i[S] = \inf_{(T, d_T) \in C_{\alpha, \beta}} \sum_{i=1}^{M} p_iE_i[T] + o(1)$$

- The $L_i$’ were introduced by Lorden (1977)

$$L_i : = \exp\{-\sum_{n=1}^{\infty} n^{-1}[P_0(Z_n^i > 0) + P_i(Z_n^i \leq 0)]\}$$

$$= \delta_i I_i.$$

- $|\log \alpha| \sim |\log \beta|$ is more restrictive than what we had assumed before.
Ingredients of proof

- Formulate a Bayesian problem, in which there is a penalty for a wrong decision under each hypothesis and a cost of sampling, $c$, per observation.

- Show that the WSLRT with these particular weights that involve the $L$ numbers (and appropriate thresholds) attains the Bayes risk up to an $o(c)$ term (Lorden (1977)).

- A third-order asymptotic expansion for expected sample size of this rule:

$$E_i[S] = \frac{1}{I_i} \left[ |\log \alpha| + \rho_i + \log \delta_i + C_i(p) \right] + o(1),$$

where

$$C_i(p) = \log \left( \sum_{k=1}^{M} \frac{p_k}{I_k} \right) - \log \left( \frac{p_k}{I_k} \right).$$
How to select $p$?

- We have seen that an almost minimax rule does not make sense.

- We may design the rule so that

$$
\max_{1 \leq i \leq M} (I_i \mathbb{E}_i[S]) = \inf_{(T,d_T) \in \mathcal{C}_{\alpha,\beta}} \max_{1 \leq i \leq M} (I_i \mathbb{E}_i[T]) + o(1).
$$

- This is done when $p_i$ is selected $\propto \mathcal{L}_i e^{\rho_i}$.

- It is not clear with this is a good criterion.
Robustness

- Let $S^i$ the optimal SPRT for testing $f_0$ against $f_i$ and set

$$J_i[S] := \frac{\mathbb{E}_i[S] - \mathbb{E}_i[S^i]}{\mathbb{E}_i[S^i]}$$

when both tests satisfy, at least approximately, the error probability constraints.

- Based on the previous approximations,

$$J_i[S] \approx \frac{C_i(p)}{|\log \alpha| + \rho_i + \log \delta_i}, \quad \text{where} \quad C_i(p) = \log \left( \sum_{k=1}^{M} \frac{p_k}{I_k} \right) - \log \left( \frac{p_i}{I_i} \right).$$

- Setting $p_i \propto I_i$ guarantees that

$$J_i[S] \sim J_j[S] \quad \forall \ 1 \leq i \neq j \leq M.$$
Two channels with densities

\[ f_0^k(x) = h(x) \quad \text{and} \quad f_1^k(x) = e^{\theta_k x - \psi(\theta_k)} h(x), \; k = 1, 2. \]

Say, \( \theta_1 = 4 \) (fixed) and let \( \theta_2 = x \) vary.
Relative performance loss vs relative signal strength

\[ \mathcal{L}_i \leq I_i \leq e^{\rho k} \mathcal{L}_i \]
Sequentially testing of a continuous parameter

- Sequentially acquired observations

\[ X_1, \ldots, X_n, \ldots \overset{iid}{\sim} f \in \{f_{\theta}, \ \theta \in \Theta\} \]

- Stop sampling as soon as possible and distinguish between

\[ H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta_1, \]

where \( \theta_0 \notin \Theta_1 \subset \Theta \).

- Then, we would like to minimize \( \mathbb{E}_{\theta_0}[T] \) and \( \mathbb{E}_{\theta}[T] \) for every \( \theta \in \Theta_1 \) in

\[ C_{\alpha, \beta} = \{(T, d_T) : \mathbb{P}_{\theta_0}(d_T = 1) \leq \alpha \quad \text{and} \quad \sup_{\theta \in \Theta_1} \mathbb{P}_{\theta}(d_T = 0) \leq \beta\}. \]
A multi-parameter exponential family

Setup

- An exponential family

\[ f_{\theta}(x) := e^{\langle \theta, x \rangle - \psi(\theta)}, \ x \in \mathbb{R}^d, \ \theta \in \Theta \subset \mathbb{R}^d. \]

- \( \Theta = \{ \theta \in \mathbb{R}^d : \int e^{\langle \theta, x \rangle} \nu(dx) < \infty \} \) is the natural parameter space.

- \( \psi(\theta) = \log \int e^{\langle \theta, x \rangle} \nu(dx) \) is the log-moment generating function of \( X \).

- We denote by \( \dot{\psi}(\theta) \) the gradient and by \( \ddot{\psi}(\theta) \) the Hessian matrix of \( \psi(\theta) \).

- We assume that \( \ddot{\psi}(\theta) \) is non-singular for all \( \theta \in \Theta \).

- The Kullback–Leibler information number between \( f_{\theta_2} \) and \( f_{\theta_1} \) is

\[
I(\theta_2, \theta_1) := \mathbb{E}_{\theta_2} \left[ \log \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} \right] = \langle \theta_2 - \theta_1, \dot{\psi}(\theta_2) \rangle - [\psi(\theta_2) - \psi(\theta_1)].
\]

- \( I(\theta_0, \theta_1) > 0 \quad \forall \ \theta_0 \in \Theta_0, \ \theta_1 \in \Theta_1. \)
The one-sided setup

- Suppose that *sampling needs to stop only to reject* $\mathbb{H}_0$.
- Then, we need to minimize $\mathbb{E}_\theta[T]$ for every $\theta \in \Theta_1$ among stopping times in
  \[ C_\alpha = \{ T : \mathbb{P}_0(T < \infty) \leq \alpha \}. \]
- Let $\ell_n(\theta)$ the likelihood of the first $n$ observations under $\mathbb{P}_\theta$, i.e.,
  \[ \ell_n(\theta) = \prod_{k=1}^{n} f_\theta(X_k). \]

Open-ended WSPRT and GSLRT

Let $B > 1$ be a fixed threshold and $g$ a positive function on $\Theta_1$. Define

\[ S_B(g) = \inf\{ t \geq 1 : \Lambda_t \geq B \}, \quad \Lambda_t = \frac{1}{\ell_n(\theta_0)} \int_{\Theta_1} \ell_t(\theta) g(\theta) d\theta \]

\[ \hat{S}_B = \inf\{ t \geq 1 : \hat{\Lambda}_t \geq B \}, \quad \hat{\Lambda}_t = \frac{1}{\ell_t(\theta_0)} \sup_{\theta \in \Theta_1} \ell_t(\theta). \]
A minimax, second-order property

- The weighted idea goes back to Wald (1945).
- The GSLRT has been studied by Schwarz (1962), Wong (1968), Lorden (1977), Lai (1988, 2004), etc.

If $\Theta_1$ is a compact set bounded away from 0,

- both tests attain

$$\inf_{T \in \mathcal{C}_\alpha} \sup_{\theta \in \Theta_1} I(\theta, 0) \mathbb{E}_{\theta}[T]$$

within an $O(1)$ term as $\alpha \to 0$.

- Pollak (1978) proved this result for the WSPRT with any continuous mixing density whose support includes $\Theta_1$ (for a one-parameter exponential family).

- Lai (2004) proved this result for the GSLRT.
Almost Minimax WSPRT

Asymptotic average overshoot

Consider the one-sided SPRT for testing $f_\theta$ versus $f_0$,

$$T^\theta_a := \inf\{ t : Z^\theta_t \geq a \}, \quad \text{where} \quad Z^\theta_t := \log \frac{\ell_n(\theta)}{\ell_t(\theta_0)}$$

and define $\kappa^\theta := \lim_{a \to \infty} \mathbb{E}_\theta[Z^\theta_{T^\theta_a} - a]$.

Theorem (F. & Tartakovsky (2013))

Consider the WSPRT $S_B(g)$ with weight function

$$\tilde{g}(\theta) := e^{2\kappa^\theta} \sqrt{\det(\psi(\theta)) / I(\theta, 0)}$$

and suppose that $\mathbb{P}_0(S_B(\tilde{g}) < \infty) = \alpha$. Then, as $\alpha \to 0$,

$$\sup_{\theta \in \Theta_1} I(\theta, 0) \mathbb{E}_\theta[S_B(\tilde{g})] = \inf_{T \in \mathcal{C}_\alpha} \sup_{\theta \in \Theta_1} I(\theta, 0) \mathbb{E}_\theta[T] + o(1).$$
Idea of Proof

**Auxiliary Bayesian problem**

Consider the sequential decision problem with
- loss 1 when stopping under $\mathbb{P}_0$,
- sampling cost per observation equal to $cI_\theta$ under $\mathbb{P}_\theta$,
- conditional prior distribution on $\Theta_1$ given that $\theta \neq 0$ equal to $g$.

The WSPRT $S_B(g)$ is asymptotically Bayes as $c \to 0$ within an $o(c)$ term.

**Almost equalizer**

As $B \to \infty$

$$I(\theta, 0) \mathbb{E}_\theta[S_B(\tilde{g})] = \log B + \frac{d}{2} \log \log B + C + o(1),$$

where $C$ is a constant term that does not depend on $\theta$.

**Idea of proof**

“Almost Bayesian + Almost Equalizer = Almost Minimax”
Almost Minimax Weighted GSLRT

- A *weighted* version of the GSLRT turns out to have the same optimality property.
- Recall that

\[ \hat{\Lambda}_n = \sup_{\theta \in \Theta_1} \frac{\ell_n(\theta)}{\ell_n(\theta_0)} = \frac{\ell_n(\hat{\theta}_n)}{\ell_n(\theta_0)}, \]

where \( \hat{\theta}_n \) is the (constrained on \( \Theta_1 \)) MLE of \( \theta \) based on the first \( n \) observations.

We define:

\[ \hat{S}_B(g) = \inf\{n \geq 1 : \hat{\Lambda}_n g(\hat{\theta}_n) \geq B\} \]

where \( g \) is some positive function on \( \Theta_1 \) and \( B > 1 \) is a fixed threshold.

**Theorem**

Consider the WGSLRT with \( \hat{g}(\theta) = e^{x_\theta} \). If \( P_0(\hat{S}_B(\hat{g}) < \infty) = \alpha \), then as \( \alpha \to 0 \)

\[ \sup_{\theta \in \Theta_1} I(\theta, 0) \mathbb{E}_\theta[\hat{S}_B(\hat{g})] = \inf_{T \in C_\alpha} \sup_{\theta \in \Theta_1} I(\theta, 0) \mathbb{E}_\theta[T] + o(1). \]
Remarks

- The function $\theta \rightarrow \kappa_{\theta}$ usually does not admit a closed-form expression.

- As a result, the previous nearly minimax sequential tests can be implemented only approximately, as the corresponding mixture-based and generalized likelihood ratio statistics can only be computed numerically.

- For the two-sided testing problem, we need an almost Bayes rule for exponential families (Lorden (1977)) and we need to consider again the $\mathcal{L}$ numbers (Keener (2005)).
- Extension to a composite null hypothesis.
- Extension to multiple hypotheses.
THANK YOU!
References


