Opportunistic Detection Rules

Wenyi Zhang (USTC, China), joint work with G. V. Moustakides (Patras, Greece) and H. V. Poor (Princeton, USA)

July 2014
ISIT, Honolulu, HI, USA

Acknowledgement:
NSFC 61379003 and SRFDP-RGC ERG JRS 20133402140001 (WZ);
US NSF DMS-1118605 and ECCS-1343210 (HVP)
A simple binary hypothesis test:

null : \( \mathcal{H}_0 \sim p_0(x) \)

alternative : \( \mathcal{H}_1 \sim p_1(x) \)

- A fixed sample size (FSS) setting: Consider \( N \) i.i.d. samples \( X_i, \ i = 1, 2, \ldots, N \).
- Optimal rule (Neyman-Pearson): A threshold test comparing the likelihood ratio \( \Lambda_N = \prod_{i=1}^{N} \frac{p_1(X_i)}{p_0(X_i)} \) against a prescribed threshold.
Prelude

- Suppose that the samples are collected sequentially.
- We are usually tempted to decide before collecting all the \( N \) samples, since the first few of them may be already informative enough.
- The main theme of this work hence is about adding some sequential analysis ingredients into the classic FSS framework.
- A key distinction with sequential probability ratio test (SPRT) is the limited availability of samples (at most \( N \)).
Opportunistic Detection Rules

Asymptotic Regime

Finite-length Regime

Wrap-up Remarks
Opportunistic detection rules

An ODR consists of

▸ a stopping time $T$ adapted to the filtration generated by $X_i, i = 1, \ldots, N$;

▸ a terminal decision rule $D \in \{\mathcal{H}_0, \mathcal{H}_1\}$.

**Condition:** When $T < N$, $D = \mathcal{H}_1$. This is because we

▸ desire an early stopping under $\mathcal{H}_1$ (e.g., an abnormal condition that requires immediate attention).

▸ tolerate delay in decision under $\mathcal{H}_0$ (e.g., a normal condition).
Performance metrics

- False alarm probability: \( P_{FA} = P_0[D = H_1] \)
- Miss probability: \( P_M = P_1[D = H_0] \)
- Mean delay under \( H_1 \): \( T = E_1[T] \)
Opportunistic Detection Rules

Asymptotic Regime

Finite-length Regime

Wrap-up Remarks
Asymptotic regime

Let $N \to \infty$.

Asymptotic performance tuple:

$$\left( \liminf_{N \to \infty} - \log P_{FA}, \liminf_{N \to \infty} - \log P_M, \limsup_{N \to \infty} \frac{\mathbb{E}_1[T]}{N} \right)$$

ODR performance region: the closure of the union of achievable tuples under all possible ODRs.

$$\mathcal{R}(\rho_0, \rho_1) \subset [0, \infty) \times [0, \infty) \times [0, 1] \subset \mathbb{R}^3$$
Characterization of $\mathcal{R}(p_0, p_1)$

Denote $d_0 = D(p_0 \parallel p_1)$ and $d_1 = D(p_1 \parallel p_0)$. 

The boundary $(\Delta_{FA}, \Delta_M, \eta)$ of $\mathcal{R}(p_0, p_1)$ satisfies, for each $0 \leq \eta \leq 1$,

$$
\Delta_{FA} = \min \left\{ \eta d_1, \sup_{\alpha > 0} \left\{ \alpha [d_1 - \nu (d_0 + d_1)] - \log \mathbb{E}_0 \left[ e^{\alpha \log p_1(X)/p_0(X)} \right] \right\} \right\}
$$

$$
\Delta_M = \sup_{\beta < 0} \left\{ \beta [d_1 - \nu (d_0 + d_1)] - \log \mathbb{E}_1 \left[ e^{\beta \log p_1(X)/p_0(X)} \right] \right\}
$$

for $0 \leq \nu \leq 1$. 
Illustration of $\mathcal{R}(p_0, p_1)$ (1)

For FSS tests, the optimal $(P_{\text{FA}}, P_{\text{M}})$ exponents tradeoff is given by

\[
\Delta_{\text{FA}} = \sup_{\alpha > 0} \left\{ \alpha [d_1 - \nu(d_0 + d_1)] - \log \mathbb{E}_0 \left[ e^{\alpha \log p_1(X)/p_0(X)} \right] \right\}
\]

\[
\Delta_{\text{M}} = \sup_{\beta < 0} \left\{ \beta [d_1 - \nu(d_0 + d_1)] - \log \mathbb{E}_1 \left[ e^{\beta \log p_1(X)/p_0(X)} \right] \right\}
\]

for $0 \leq \nu \leq 1$. 
\( \Delta M/(A^2/2) \sim \mathcal{N}(0, 1) \), \( \Delta FA/(A^2/2) \sim \mathcal{N}(A, 1) \): \( d_0 = d_1 = (1/2) \cdot A^2 \)
Illustration of $\mathcal{R}(p_0, p_1)$ (2)

In order to reduce the (normalized) decision delay to $\eta$, the tradeoff region is cut back by the straight line $\Delta_{FA} = \eta d_1$:
Illustration of $\mathcal{R}(p_0, p_1)$ (3)

Running over $\eta \in [0, 1]$, the ODR performance region $\mathcal{R}(p_0, p_1)$ is like
The FSS-optimal exponent of $P_M$ under a fixed $P_{FA}$, $D(p_0\|p_1)$, as given by the Stein-Chernoff lemma, is still achieved by ODRs.

This holds true even with $\eta \to 0$!
The Bayesian-optimal exponent as given by the Chernoff information

\[ C(p_0, p_1) = -\inf_{\alpha \in (0,1)} \log \int_{\mathcal{X}} p_0^\alpha(x) p_1^{1-\alpha}(x) dx \]

is achieved by ODRs if and only if \( \eta \geq C(p_0, p_1)/D(p_1 \| p_0) \).
Sketch of proof: direct part

- Basic idea: a simple two-stage decision procedure.
- Stage 1: upon observing $X_1, \ldots, X_M$, where $M \sim \eta N$.
  - At time $M$, perform a threshold test comparing $\sum_{i=1}^{M} \log \frac{p_1(X_i)}{p_0(X_i)}$ against a threshold $\tau_M$.
  - Set $\tau_M$ slightly below $d_1 M$.
- Stage 2: if not stopped at time $M$, wait until the end to decide.
  - Again perform a threshold test comparing $\sum_{i=1}^{N} \log \frac{p_1(X_i)}{p_0(X_i)}$ against a threshold $\tau_N$.
  - Set $\tau_N$ approximately as $[d_1 - \nu(d_0 + d_1)] N$. 
Miss:
- Occurs under $p_1$ when neither $\sum_{i=1}^{M} \log \frac{p_1(X_i)}{p_0(X_i)}$ nor $\sum_{i=1}^{N} \log \frac{p_1(X_i)}{p_0(X_i)}$ exceeds $\tau_M$ nor $\tau_N$, respectively.
- Dominated by the latter, leading to the exponent $\Delta_M$.

False alarm:
- Occurs under $p_0$ when either $\sum_{i=1}^{M} \log \frac{p_1(X_i)}{p_0(X_i)}$ or $\sum_{i=1}^{N} \log \frac{p_1(X_i)}{p_0(X_i)}$ exceeds $\tau_M$ or $\tau_N$, respectively.
- Bounded by union bound, and can be shown to lead to the exponent $\Delta_{FA}$.

Expected decision delay:
- With high probability we stop at time $M$, thus leading to $\mathbb{E}_1[T]/N \sim M/N \sim \eta$. 
Sketch of proof: converse part

- Basic idea: an information-theoretic proof invoking a channel coding converse.

- The key of the converse is to show that the corner point \((\Delta_M = 0, \Delta_{FA} = \eta d_1)\) cannot be exceeded.

- An argument by contradiction:
  - Suppose that \((\Delta_M = 0, \Delta_{FA} = \eta d_1)\) could be exceeded by a certain sequence of ODRs indexed by \(N\).
  - Then we can construct a (variable-length) channel code, for a memoryless binary-input channel \(U \rightarrow X\) with \(p(x|u = 0) = p_0(x)\) and \(p(x|u = 1) = p_1(x)\), in the presence of feedback, to achieve a rate per unit cost higher than its capacity per unit cost.
  - But this is impossible due to the channel coding converse.


Opportunistic Detection Rules

Asymptotic Regime

Finite-length Regime

Wrap-up Remarks
Finite-length regime: problem formulation

Now turn to the finite-length regime with a fixed $N$.

- Consider Bayesian formulations.
- False alarm: \( \{ T < N \} \cup \{ T = N, D = \mathcal{H}_1 \} \) w.r.t. \( p_0 \)
- Miss: \( \{ T = N, D = \mathcal{H}_0 \} \) w.r.t. \( p_1 \)
- Bayesian risk:

\[
J = (1 - \pi)c_0 P_{FA} + \pi c_1 P_M + c \mathbb{E}_1[T],
\]

\( 0 \leq \pi \leq 1 \): prior probability of \( \mathcal{H}_0 \), and \( c_0, c_1, c > 0 \): cost assignments.
Solution structure

The Bayesian optimal ODR minimization problem can be cast and solved as a Markov optimal stopping problem in standard form.

The solution is a sequence of likelihood ratio threshold tests:

- Thresholds are time-varying, given by the solutions of

$$c \lambda + \mathbb{E}_0[h_{k+1}(\lambda p_1(X)/p_0(X))] = (1 - \pi)c_0,$$

for \(k = 1, 2, \ldots, N - 1\), and \(\tau_N = \frac{(1-\pi)c_0}{\pi c_1}\).

- Backward recursion:

$$h_k(\lambda) = \min\{(1 - \pi)c_0, c \lambda + \mathbb{E}_0[h_{k+1}(\lambda p_1(X)/p_0(X))]\},$$

for \(k = N - 1, N - 2, \ldots, 1\), and

$$h_N(\lambda) = \min\{(1 - \pi)c_0, \pi c_1 \lambda\}.$$
Illustration

For $p_0 \sim \mathcal{N}(0, 1)$, $p_1 \sim \mathcal{N}(A, 1)$, set $A = 1$, $\pi = 1/2$, $c = 1$, $N = 50$.

- Thresholds become stationary within a few samples (returning from $n = N$).
- Trend may differ depending upon parameters: increasing, decreasing, and “overshooting”.
Extension for random maximum sample size

- Now, instead of a fixed $N$, suppose $N$ is a random variable following a geometric distribution with parameter $0 < \epsilon < 1$.
- The realization of $N$ is not revealed to the statistician until observing $X_N$.
  - If the statistician has reached $X_N$ without a detection yet, then he is required to make his decision immediately with $X_1, \ldots, X_N$.
- Imagine a scenario in which the observation process is subject to abrupt interruption, or in which an external controller, in a unanticipated manner, issues a command for prompt decision.
Again consider a Bayesian risk minimization setup.

\[ J = (1 - \pi)c_0 P_{FA} + \pi c_1 P_{M} + c E_1[T]. \]

Key idea:

1. View \( N \) as a stopping time defined as \( N = \min\{n : Z_n = 1\} \) where \( Z_n \) is an i.i.d. sequence of Bernoulli trials with success probability \( \epsilon \).
2. Optimize over stopping times \( T = \min\{T', N\} \) adapted to the product filtration generated by \((X_1, Z_1), (X_2, Z_2), \ldots\)
The Bayesian risk can be deduced into

\[ J = \mathbb{E}_0 \left[ (1 - \epsilon)^T g(\Lambda_T) + \sum_{n=0}^{T-1} (1 - \epsilon)^n c(\Lambda_n) \right], \]

where \( g(\lambda) = (1 - \pi)c_0 + \frac{\epsilon}{1 - \epsilon} \min \left\{ (1 - \pi)c_0, \pi c_1 \lambda \right\} \) and \( c(\lambda) = c\lambda, \) for \( \lambda \geq 0. \)

The risk considers both an instantaneous reward at the stopping time and (exponentially discounted) accumulated sampling costs.

Treated in, e.g., A. N. Shiryaev, *Optimal Stopping Rules*
Solution structure

- **Optimal stopping time:**
  
  $$T = \min\{n \geq 1 : V(\Lambda_n) = g(\Lambda_n)\},$$

  where $V(\cdot)$ is the solution of

  $$V(\lambda) = \min\{g(\lambda), (1 - \epsilon)E_0[V(\lambda p_1(X)/p_0(X))] + c(\lambda)\}.$$

- **Optimal decision rule:** A two-threshold scheme
  
  - A “running” threshold $\tau_r$ as the value of $\lambda$ at the intersection of $g(\lambda)$ and $(1 - \epsilon)E_0[V(\lambda p_1(X)/p_0(X))] + c(\lambda)$, used to compare with $\{\Lambda_n\}$ for $n < N$.
  
  - A “terminal” threshold $\tau_t$ which is simply $\frac{(1 - \pi)c_0}{\pi c_1}$, used when $X_N$ is reached.
For $p_0 \sim \mathcal{N}(0, 1)$, $p_1 \sim \mathcal{N}(A, 1)$, set $A = 1, \pi = 1/2, c = 1, \epsilon = 0.05, c_0 = c_1$.

- The dash-dot line indicates $\tau_t$.
- $\tau_r$ can be either greater or smaller than $\tau_t$ depending upon parameters.
Opportunistic Detection Rules

Asymptotic Regime

Finite-length Regime

Wrap-up Remarks
Wrap-up remarks

- ODRs mix sequential analysis ingredients into FSS decision framework.
- Characterization of exponential tradeoff in the asymptotic regime is established.
- Bayesian optimal solutions in the finite-length regime is provided; the two-threshold decision scheme for the random sample size case is unusual and interesting.
- A number of open issues unaddressed:
  - What is the behavior of the Bayesian optimal ODRs when $c \to 0$ and $N \to \infty$?
  - Is there a Wald-Wolfowitz theorem (for SPRT) type of result for ODRs?
  - How do optimal ODRs behave for continuous-time stochastic processes?
  - ...