

Sequential Hypothesis Tests: Historical Overview and Recent Results

PLENARY LECTURE

Alexander Tartakovsky

Department of Mathematics
and
Center for Applied Mathematical Sciences



USC University of
Southern California

Los Angeles, CA 90089-2532

tartakov@usc.edu

<http://cams.usc.edu/usr/facmemb/tartakov/>

Fourth International Workshop in Sequential Methodologies
IWSM 2013
Athens, Georgia
July 18, 2013

- **Tests of Composite Hypotheses for iid Data:** Some history

- **Tests of Composite Hypotheses for iid Data:** Some history
- **Generalized SPRT, Weighted SPRT (mixtures) and Adaptive SPRT:**
Asymptotic optimality (Bayesian and frequentist problems)

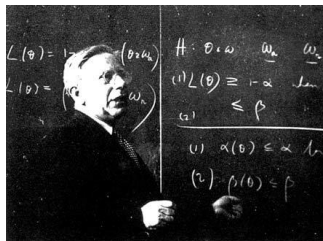
- **Tests of Composite Hypotheses for iid Data:** Some history
- **Generalized SPRT, Weighted SPRT (mixtures) and Adaptive SPRT:**
Asymptotic optimality (Bayesian and frequentist problems)
- **Minimax Problems:** Asymptotic optimality for Kullback–Leibler information cost measure

- **Tests of Composite Hypotheses for iid Data:** Some history
- **Generalized SPRT, Weighted SPRT (mixtures) and Adaptive SPRT:** Asymptotic optimality (Bayesian and frequentist problems)
- **Minimax Problems:** Asymptotic optimality for Kullback–Leibler information cost measure
- **Multihypothesis Tests:** Asymptotic optimality of a Matrix SPRT for iid and general non-iid models in multiple decision problems

- **Tests of Composite Hypotheses for iid Data:** Some history
- **Generalized SPRT, Weighted SPRT (mixtures) and Adaptive SPRT:** Asymptotic optimality (Bayesian and frequentist problems)
- **Minimax Problems:** Asymptotic optimality for Kullback–Leibler information cost measure
- **Multihypothesis Tests:** Asymptotic optimality of a Matrix SPRT for iid and general non-iid models in multiple decision problems
- **Examples and Applications**
- **Acknowledgements**

- 1 **Abraham Wald (1902–1950) – Founder of Sequential Analysis**
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems
- 7 Acknowledgements

Abraham Wald (1902–1950) – Founder of Sequential Analysis



- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)**
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems
- 7 Acknowledgements

Testing Two Simple Hypotheses (iid Case)

- $X_1, X_2, \dots$ is a sequence of iid observations, and $p_\theta(x)$ is a density (parametrized by a parameter θ)
- **Wald's SPRT (1943-47)** $\delta^* = (T^*, d^*)$ for testing a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1$:

$$T^*(A_0, A_1) = \inf \{n \geq 1 : \Lambda_n(\theta_0, \theta_1) \notin (A_0, A_1)\}, \quad 0 < A_0 < 1 < A_1 \text{ (two thresholds)}$$

$$d^* = \begin{cases} 1 \text{ (} H_1 \text{ is accepted)} & \text{if } \Lambda_{T^*}(\theta_0, \theta_1) \geq A_1 \\ 0 \text{ (} H_0 \text{ is accepted)} & \text{if } \Lambda_{T^*}(\theta_0, \theta_1) \leq A_0 \end{cases}$$

$$\Lambda_n(\theta_0, \theta_1) = \prod_{k=1}^n \frac{p_{\theta_1}(X_k)}{p_{\theta_0}(X_k)} - \text{Likelihood Ratio (LR)}$$

- **SPRT's Remarkable Optimality Property:** Minimizes both ESSs $E_{\theta_0}T$ and $E_{\theta_1}T$ in the class of tests

$$\mathbb{C}(\alpha_0, \alpha_1) = \{\delta = (T, d) : P_{\theta_0}(d = 1) \leq \alpha_0, P_{\theta_1}(d = 0) \leq \alpha_1\}, \quad \alpha_0 + \alpha_1 < 1$$

whenever the thresholds $A_i(\alpha_0, \alpha_1)$ can be selected so that $\alpha_i^* = \alpha_i$.

- While the SPRT has this remarkable optimality property in the *iid* case there are several problems and extensions that are dictated by various applications:
- 1 How to select thresholds $A_i(\alpha_0, \alpha_1)$ to guarantee exact or almost exact equalities $\alpha_i^* = \alpha_i$? The overshoot problem: renewal-theoretic considerations and Siegmund's Brownian motion approximations.
- 2 The true parameter value is usually unknown and putative values θ_0 and θ_1 are rarely representative in applications, so one needs to address composite hypotheses, in which case the SPRT is not optimal anymore. Associated problems: Kiefer–Weiss, Invariant tests, Uniformly asymptotically optimal tests (Generalized LR, Mixtures).
- 3 What if the observations are non-iid? SPRT's optimality still holds when the LLR $\lambda_n = \log \Lambda_n$ is a random walk, which is always the case for iid models, but rarely for non-iid (some examples will be considered later on).
- 4 Generalization to multiple decision problems (multi-hypotheses tests) for iid and non-iid models – simple hypotheses.
- 5 Generalization to multiple decision problems (multi-hypotheses tests) for iid and non-iid models – composite hypotheses.

Critique of the SPRT: Composite Hypotheses

- Even though Wald's SPRT has the *remarkable* optimality property of minimizing the ESS at the putative values θ_i of the parameter θ to which it is tuned, it loses the optimality property for the values of θ other than θ_i
- **Example:** Testing for the mean of the normal population $X_n \sim \mathcal{N}(\theta, 1)$, $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta \geq \theta_1$ with the indifference interval $I_{\text{in}} = (\theta_0, \theta_1)$
- The SPRT still has some useful optimality properties since $\sup_{\Theta_i} E_{\theta} T^* = E_{\theta_i} T^*$, $\sup_{\Theta_i} \alpha_i^*(\theta) = \alpha_i^*(\theta_i)$, so that it is minimax.
- However, in the indifference zone the situation changes dramatically:

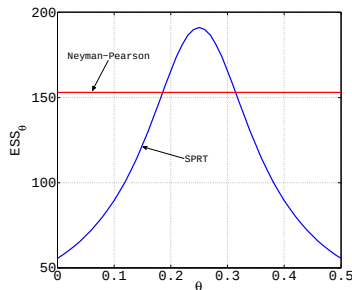


Figure: SPRT's ESS versus θ for $\alpha = 0.001$, $\theta_0 = 0$, $\theta_1 = 0.5$. The horizontal line corresponds to the fixed sample size of the NP test.

Testing Two Composite Hypotheses

- **Wald'43,47** proposed two approaches for modifying the SPRT to test composite hypotheses $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$:

① **Weighted SLRT (WSLRT)** often referred to as **mixtures**

$$\bar{T}(A_0, A_1) = \inf \{n : \bar{\Lambda}_n \notin (A_0, A_1)\}, \quad \bar{\Lambda}_n = \frac{\int_{\Theta_1} w_1(\theta) \prod_{k=1}^n p_\theta(X_k) d\theta}{\int_{\Theta_0} w_0(\theta) \prod_{k=1}^n p_\theta(X_k) d\theta} - \text{WLR}$$

② **Generalized SLRT** (adopted from classical fixed-sample size theory)

$$\hat{T}(A_0, A_1) = \inf \{n : \hat{\Lambda}_n \notin (A_0, A_1)\}, \quad \hat{\Lambda}_n = \frac{\sup_{\theta \in \Theta_1} \prod_{k=1}^n p_\theta(X_k)}{\sup_{\theta \in \Theta_0} \prod_{k=1}^n p_\theta(X_k)} - \text{GLR}$$

- The upper bounds on the average error probabilities of the WSLRT:

$$\int_{\Theta_0} \mathbf{P}_\theta(\bar{d} = 1) w_0(\theta) d\theta \leq 1/A_1, \quad \int_{\Theta_1} \mathbf{P}_\theta(\bar{d} = 0) w_1(\theta) d\theta \leq A_0$$

- *Frequentists* would strongly prefer to upper-bound the maximal error probabilities:

$$\mathbb{C}(\alpha_0, \alpha_1) = \{\delta : \sup_{\theta \in \Theta_0} \mathbf{P}_\theta(d = 1) \leq \alpha_0, \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(d = 0) \leq \alpha_1\}$$

- It is not clear how to obtain the upper bounds on maximal error probabilities of WSLRT and GSLRT

- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone**
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems
- 7 Acknowledgements

Hypothesis Testing with Indifference Zone: Frequentist Formulation

- The ℓ -dimensional vector parameter $\theta = (\theta_1, \dots, \theta_\ell) \in \Theta \subseteq \mathbb{R}^\ell$, and the hypotheses are $H_i : \theta \in \Theta_i, i = 0, 1$, where Θ_i are disjoint subsets of Θ
- **Indifference zone** I_{in} (i.e., $\Theta = \Theta_0 + \Theta_1 + I_{\text{in}}$): No constraints on the probabilities of errors imposed in I_{in} .
- The indifference zone, where any decision is acceptable, is usually introduced keeping in mind that the correct action is not critical and often not even possible when the hypotheses are too close, which is perhaps the case in most, if not all, practical applications
 - The loss $L(\theta)$ associated with wrong decisions is 0 if $\theta \in I_{\text{in}}$ and positive and nondecreasing for $\theta \in \Theta_0 + \Theta_1$
- “Simple” (0 – 1) loss function: $L(\theta) = \mathbb{1}_{\{\theta \in \Theta_0 + \Theta_1\}}$
 - Then we may ask if there are tests that would minimize the ESS $E_\theta T$ uniformly for all $\theta \in \Theta$ in the class $\mathbb{C}(\alpha_0, \alpha_1)$ (of course approximately/asymptotically since strictly uniformly optimal tests do not exist): Find a test T_0 such that asymptotically

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_\theta [T] / E_\theta [T_0] \sim 1 \quad \text{as } \alpha_0, \alpha_1 \rightarrow 0 \text{ for all } \theta \in \Theta$$

or more generally minimizes asymptotically higher moments

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_\theta [T^r] / E_\theta [T_0^r] \sim 1 \quad \text{as } \alpha_0, \alpha_1 \rightarrow 0 \text{ for all } \theta \in \Theta, r > 1$$

- Many (especially early) works, starting with the seminal papers of **Schwarz'62** and **Kiefer&Sacks'63**, deal with Bayesian problems.
- Put prior $W(\theta)$ on Θ with the cost c per observation and a loss function $L(\theta)$ at the point θ associated with accepting the incorrect hypothesis, then the Bayes (integrated) risk of a sequential test $\delta = (T, d)$ is

$$\rho_c^W(\delta) = \int_{\Theta_0} L(\theta) \alpha_0(\delta, \theta) W(d\theta) + \int_{\Theta_1} L(\theta) \alpha_1(\delta, \theta) W(d\theta) + c \int_{\Theta} E_{\theta}[T] W(d\theta)$$

- Consider a one-parameter exponential family $p_{\theta}(X_n) = e^{X_n\theta - b(\theta)}$, hypotheses $H_0 : \theta \leq \theta_0$, $H_1 : \theta \geq \theta_1$ and zero-one loss. Then, using optimal stopping theory, it can be shown that the optimal decision-making strategy $\delta_0 = (T_0, d_0)$ is

$$T_0 = \inf \{n \geq 1 : (S_n, n) \in \mathcal{B}_c\}, \quad d_0 = 0 \text{ if } (S_n, n) \in \mathcal{B}_c^0, \quad d_0 = 1 \text{ if } (S_n, n) \in \mathcal{B}_c^1,$$

where $S_n = X_1 + \cdots + X_n$ and $\mathcal{B}_c = \mathcal{B}_c^0 \cup \mathcal{B}_c^1$ is a set that can be found numerically.

- Note that the *a posteriori* risk of stopping is

$$R_n^{\text{st}}(S_n) = \min_{i=0,1} \left\{ \frac{\int_{\Theta_i} \exp \{ \theta S_n - nb(\theta) \} W(d\theta)}{\int_{\Theta} \exp \{ \theta S_n - nb(\theta) \} W(d\theta)} \right\}$$

- **Schwarz'62** proposed a simple procedure: continue sampling until the posterior stopping risk $R_n^{\text{st}}(S_n)$ is greater than c , i.e., stop at $\tilde{T}(c) = \inf \{n : R_n^{\text{st}}(S_n) \leq c\}$
- Applying Laplace's asymptotic integration method to evaluate the integrals lead Schwarz to an approximation that prescribes to stop sampling at the time $\hat{T}(\hat{\theta}) = \min(\hat{T}_0(\hat{\theta}), \hat{T}_1(\hat{\theta}))$, where

$$\hat{T}_i(\hat{\theta}) = \inf \left\{ n \geq 1 : \sum_{k=1}^n \log \frac{p_{\hat{\theta}_n}(X_k)}{p_{\theta_i}(X_k)} \geq \log c^{-1} \right\}, \quad i = 0, 1$$

and $\hat{\theta}_n = \sup_{\theta \in \Theta} \prod_{k=1}^n p_{\theta}(X_k)$ is the MLE, i.e., to the GSLRT $\hat{\delta}$.

- **Kiefer&Sacks'63** showed that the quasi-Bayesian procedure $\tilde{\delta}(c) = (\tilde{T}(c), \tilde{d}(c))$ (initially proposed by Schwarz) is first-order asymptotically Bayes:

$$\rho_c^W(\tilde{\delta}(c)) \sim \inf_{\delta} \rho_c^W(\delta) \quad \text{as } c \rightarrow 0 \quad \text{for any prior } W.$$

- **Wong'68** proved that Schwarz's GSLRT $\hat{\delta}$ is also first-order asymptotically Bayes:

$$\rho_c^W(\hat{\delta}) \sim \inf_{\delta} \rho_c^W(\delta) \sim c |\log c| \int_{\Theta} \frac{W(d\theta)}{I_{\max}(\theta)}, \quad \mathbb{E}_{\theta}[\hat{T}] \sim \frac{|\log c|}{I_{\max}(\theta)} \quad \text{for every } \theta \in \Theta,$$

where $I_{\max}(\theta) = \max \{I(\theta, \theta_0), I(\theta, \theta_1)\}$, $I(\theta, \theta_i) = \mathbb{E}_{\theta}[\log \frac{p_{\theta}(X_1)}{p_{\theta_i}(X_1)}]$ – KL numbers.

- **Lorden'72** and **Lorden'77 (unpublished)** refined previous Bayesian results first to second order ($O(c)$) and then to third order ($o(c)$) for testing separated hypotheses $H_0 : \underline{\theta} \leq \theta \leq \theta_0$, $H_1 : \bar{\theta} \geq \theta \geq \theta_1$, showing that the family of GSLRTs as well as WSLRTs (mixtures) can be designed so as to attain the Bayes risk to within $o(c)$ (i.e., third order asymptotic optimality).
- Lorden's GSLRT stops at $\hat{T} = \min \{ \hat{T}_0, \hat{T}_1 \}$,

$$\hat{T}_0(\hat{\theta}) = \inf \left\{ n \geq 1 : \sum_{k=1}^n \log \left[\frac{p_{\hat{\theta}_n}(X_k)}{p_{\theta_0}(X_k)} \right] \geq \log c^{-1} - \frac{1}{2} \log \log c^{-1} - \log h_i(\hat{\theta}_n) \right\},$$

$$h_i(\hat{\theta}_n) = \sqrt{\frac{2\pi I(\hat{\theta}_n, \theta_i)}{\ddot{b}(\hat{\theta}_n)}} \frac{w(\hat{\theta}_n) |\dot{b}(\hat{\theta}_n) - \dot{b}(\theta_i)|}{w(\theta_i) \zeta(\hat{\theta}_n, \theta_i)}, \quad i = 0, 1,$$

where $\zeta(\theta, \theta_i)$ is a correction for the overshoot over the boundary, the factor which is the subject of renewal theory,

$$\zeta(\theta, \theta_i) = \frac{1}{I(\theta, \theta_i)} \exp \left\{ - \sum_{n=1}^{\infty} \frac{1}{n} [\mathbf{P}_{\theta}(\lambda_n(\theta, \theta_i) \leq 0) + \mathbf{P}_{\theta_i}(\lambda_n(\theta, \theta_i) > 0)] \right\}$$

($\lambda_n(\theta, \theta_i) = \sum_{k=1}^n \log \left[\frac{p_{\theta}(X_k)}{p_{\theta_i}(X_k)} \right]$ is the LLR).

Asymptotic Optimality Results:

- 1 The family of GSLRTs with adaptive weights h_i with correction for the overshoot is third-order asymptotically Bayes:

$$\rho_c^w(\hat{\delta}) = \inf_{\delta} \rho_c^w(\delta) + o(c) \quad \text{as } c \rightarrow 0.$$

- 2 If the correction for the overshoot is not performed, then the GSLRT is second-order asymptotically optimal:

$$\rho_c^w(\hat{\delta}) = \inf_{\delta} \rho_c^w(\delta) + O(c) \quad \text{as } c \rightarrow 0$$

Crucial difference between Schwarz's and Lorden's GSLRTs:

- In Schwarz's test $h_i \equiv 1$ and the threshold $a = \log c^{-1}$:

$$\hat{T}_{\text{SW}}(\hat{\theta}) = \inf \left\{ n \geq 1 : \lambda_n(\hat{\theta}_n, \theta_i) \geq \log c^{-1} \right\}$$

- In Lorden's test there are two innovations: the boundaries are 1) reduced by $\frac{1}{2} \log \log c^{-1}$ and 2) *curved and adaptive* (depend on the behavior of the MLE $\hat{\theta}_n$),

$$\hat{T}_i(\hat{\theta}) = \inf \left\{ n \geq 1 : \lambda_n(\hat{\theta}_n, \theta_i) \geq \log c^{-1} - \frac{1}{2} \log \log c^{-1} - \log h_i(\hat{\theta}_n) \right\}, \quad i = 0, 1.$$

- Namely these both innovations make this modification of the GLR test nearly optimal.

- **Implementation Issues:** Implementation of Lorden's fully optimized tests is problematic since usually computing the overshoot correction numbers $\zeta(\theta, \theta_i)$ is not possible analytically except for some rare cases.
- For example, in the Gaussian case (testing the mean) these numbers can be computed only numerically.
- Therefore, for practical purposes only partially optimized solutions, which provide $O(c)$ -optimality, are feasible.
- A way around is a discretization of the parameter space and implementation of discrete versions (will be discussed later).

- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)**
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems
- 7 Acknowledgements

- **Limitations of the Schwarz–Lorden Asymptotic Theory:**

- Assumes a fixed indifference zone that does not allow for local alternatives, i.e., $\Delta = \theta_1 - \theta_0$ cannot approach 0 as $c \rightarrow 0$, i.e., is limited to the case where the width of the indifference zone Δ is considerably larger than $c^{1/2}$

- **Chernoff'65** considered the problem with no indifference zone $H_0 : \theta \leq \theta_0$ against $H_0 : \theta > \theta_0$ and with the loss $L(\theta) = |\theta|$ for the problem of testing the mean of a normal distribution, $X_n \sim \mathcal{N}(\theta, 1)$
- Derived a different and more complicated approximation to the Bayes test, which exploits a curved, time-varying boundary $a(cn)$ such that

$$a(t) = \log t^{-1} + \frac{1}{2} \log \log t^{-1} + O(1) \quad \text{as } t \rightarrow 0$$

- This is opposite to Schwarz's test where $a = \log c^{-1}$ is constant and Lorden's test where the boundaries $a_i(c, \hat{\theta}_n) = \log c^{-1} - \frac{1}{2} \log \log c^{-1} - \log h_i(\hat{\theta}_n)$ are adaptive but stabilize when n becomes large
- As a result, setting $\theta_0 = \theta_1$ in Schwartz's and Lorden's tests does not yield Chernoff's test
- This is frustrating since intuitively the Bayesian test with indifference zone should approach the Bayesian test without indifference zone as $\theta_0 \rightarrow \theta_1$

- **Lai'88** suggested a unified solution that marries both problems (with and without the indifference zone) replacing the constant threshold $a = |\log c|$ in Schwarz's test by the time-varying boundary $a(cn)$
- **Lai's GSLRT** $\hat{\delta}_c(\hat{\theta}) = (\hat{T}_c(\hat{\theta}), \hat{d}_c(\hat{\theta}))$:

$$\hat{T}_c(\hat{\theta}) = \inf \left\{ n \geq 1 : n \max[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1)] \geq a_\gamma(cn) \right\},$$

where

$$a_\gamma(t) = [h(t) + \gamma t]^2 / 2t, \quad \gamma = (\theta_1 - \theta_0)[\ddot{b}(\theta_0)/c]^{1/2} / 2,$$

$$h(t) = \sqrt{2t \left(\log t^{-1} + \frac{1}{2} \log \log t^{-1} - \frac{1}{2} \log 4\pi + o(1) \right)} \quad \text{as } t \rightarrow 0$$

- It is seen that $a_\gamma(t) \sim \log t^{-1}$ as $t \rightarrow 0$, similar to Chernoff's bound
- Moreover, setting $\theta_0 = \theta_1$ yields the stopping time of Chernoff's test

$$\hat{T}_c = \inf \{ n \geq 1 : nI(\hat{\theta}_n, \theta_0) \geq a_0(cn) \}$$

with $a_0(cn) = h^2(cn)/2t = \log t^{-1} + \frac{1}{2} \log \log t^{-1} + O(1)$

- Lai's GSLRT is asymptotically **first-order Bayes**:

① **Indifference zone:** For fixed θ_0, θ_1 , as $c \rightarrow 0$

$$\rho_c^W(\hat{\delta}_c) \sim \inf_{\delta} \rho_c^W(\delta) \sim c |\log c| \int_{\Theta} \frac{W(d\theta)}{I_{\max}(\theta)}, \quad I_{\max}(\theta) = \max \{I(\theta, \theta_0), I(\theta, \theta_1)\}$$

$$\log \alpha_0(\hat{\delta}_c) \sim \log \alpha_1(\hat{\delta}_c) \sim \log c$$

② **No indifference zone:** If $c \rightarrow 0$ and $\theta_1 \rightarrow \theta_0$ such that $\Delta_c = c/(\theta_1 - \theta_0)^2 \rightarrow 0$ and $W(\theta)$ has positive continuous density $w(\theta)$ in a neighborhood of θ_0 , then

$$\rho_c^W(\hat{\delta}_c) \sim \inf_{\delta} \rho_c^W(\delta) \sim \frac{8w(\theta_0)}{\ddot{b}(\theta_0)} (\Delta_c)^{1/2} |\log \Delta_c|,$$

$$\log \alpha_0(\hat{\delta}_c) \sim \log \alpha_1(\hat{\delta}_c) \sim \log \Delta_c.$$

- All these results can be generalized to the multi-dimensional exponential family, $p_{\theta}(X_n) = e^{\theta^{\top} X_n - b(\theta)}$ (for testing $H_i : \theta \in \Theta_i$) and the GSLRT of the form

$$\hat{T} = \inf \left\{ n \geq 1 : \max_{i=0,1} \hat{\lambda}_n^i \geq a(cn) \right\},$$

$$\hat{\lambda}_n^i = \sum_{k=1}^n \log p_{\hat{\theta}_n}(X_k) - \sup_{\theta \in \Theta_i} \sum_{k=1}^n \log p_{\theta}(X_k) = n \inf_{\theta \in \Theta_i} I(\hat{\theta}_n, \theta),$$

$$I(\theta, \tilde{\theta}) = \mathbb{E}_{\theta} \left[\log \frac{p_{\theta}(X_1)}{p_{\tilde{\theta}}(X_1)} \right] = (\theta - \tilde{\theta})^{\top} \nabla b(\theta) - (b(\theta) - b(\tilde{\theta}))$$

- All previous versions of GSLRTs are first-order uniformly asymptotically optimal wrt the ESS in the class $\mathbb{C}(\alpha_0, \alpha_1)$ as $\alpha \rightarrow 0$:

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} \mathbb{E}_\theta[T] \sim \mathbb{E}_\theta[\widehat{T}(\hat{\theta})] \sim \frac{|\log \alpha|}{I_{\max}(\theta)} \quad \text{for all } \theta \text{ in a bounded subset } \tilde{\Theta} \text{ of } \Theta$$

- Note that this result assumes asymptotically symmetric case $\log \alpha_0 \sim \log \alpha_1 \sim \log \alpha$, i.e., $\log \alpha_0 / \log \alpha_1 \sim 1$
- This result can be generalized to the ℓ -dimensional exponential family and asymptotically asymmetric case (extending Schwarz's, Lorden's and Lai's tests to two different thresholds with $c_i = \alpha_i$) $\log \alpha_0 / \log \alpha_1 \sim \gamma$, $0 < \gamma < \infty$ as $\alpha_{\max} \rightarrow 0$:

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} \mathbb{E}_\theta T = \mathbb{E}_\theta \widehat{T}(1 + O(\log \log \alpha_{\max}^{-1} / \log \alpha_{\max}^{-1})) \quad \text{for all } \theta \in \tilde{\Theta}$$

and

$$\mathbb{E}_\theta[\widehat{T}(\hat{\theta})] \sim \min \left\{ \frac{|\log \alpha_0|}{\inf_{\tilde{\theta} \in \Theta_0} I(\theta, \tilde{\theta})}, \frac{|\log \alpha_1|}{\inf_{\tilde{\theta} \in \Theta_1} I(\theta, \tilde{\theta})} \right\} \quad \text{uniformly for all } \theta$$

- Asymptotic approximations can be obtained for error probabilities using large/moderate deviation theory

$$\sup_{\theta \in \Theta_i} \hat{\alpha}_i(\theta) \sim a^{\ell/2} e^{-a} C_i, \quad i = 0, 1 \quad \text{as } a \rightarrow \infty,$$

but still there are no upper bounds!

- The ℓ -dimensional parameter $\theta = (\theta_1, \dots, \theta_\ell) \in \Theta \subset \mathbb{R}^\ell$; hypotheses are

$$H_i : \theta \in \Theta_i, \quad i = 0, 1, \quad \Theta = \Theta_0 + \Theta_1 + I_{\text{in}}$$

- Weighted LR (mixture) Statistics**

$$\bar{\Lambda}_n^i = \frac{\int_{\Theta} \prod_{k=1}^n p_{\theta}(X_k) W(d\theta)}{\int_{\Theta_i} \prod_{k=1}^n p_{\theta}(X_k) W_i(d\theta)}, \quad i = 0, 1$$

where $W(\theta)$, $W_i(\theta)$ are weight functions not necessarily normalized to 1.

- Adaptive LR Statistics** ($\hat{\theta}_n$ – MLE or another reasonable estimate)

$$\hat{\Lambda}_n^i = \frac{\prod_{k=1}^n p_{\hat{\theta}_{k-1}}(X_k)}{\sup_{\theta \in \Theta_i} \prod_{k=1}^n p_{\theta}(X_k)}, \quad i = 0, 1$$

- Remark:** The adaptive LR exploits one-stage delayed estimates $\hat{\theta}_{k-1}(X_1, \dots, X_{k-1})$ – a non-anticipated approach of **Robbins and Siegmund'70**
- WSLRT** $\bar{\delta} = (\bar{T}, \bar{d})$ and **ASLRT** $\hat{\delta} = (\hat{T}, \hat{d})$:

$$\bar{T}_i = \inf \left\{ n \geq 1 : \log \bar{\Lambda}_n^i \geq a_i \right\}, \quad \bar{T} = \min(\bar{T}_0, \bar{T}_1), \quad \bar{d} = i \text{ if } \bar{T} = \bar{T}_i$$

$$\hat{T}_i = \inf \left\{ n \geq 1 : \log \hat{\Lambda}_n^i \geq a_i \right\}, \quad \hat{T} = \min(\hat{T}_0, \hat{T}_1), \quad \hat{d} = i \text{ if } \hat{T} = \hat{T}_i$$

- **Advantage of the adaptive LR over mixed LR and generalized LR:**

$$\hat{\Lambda}_n(\{\hat{\theta}_k\}, \theta) = \frac{\prod_{k=1}^n p_{\hat{\theta}_{k-1}}(X_k)}{\prod_{k=1}^n p_{\theta}(X_k)} = \hat{\Lambda}_{n-1}(\{\hat{\theta}_k\}, \theta) \times \frac{p_{\hat{\theta}_{n-1}}(X_n)}{p_{\theta}(X_n)}$$

is a martingale with unit expectation under P_{θ} (like usual LR), so it is a valid LR process, and as a result we may obtain simple upper bounds on the error probabilities using Wald's likelihood ratio identity $E_{\theta}[\hat{\Lambda}_T(\theta) \mathbb{1}_{\{T < \infty\}}] = 1$.

- **Upper Bounds on Maximal Error Probabilities of ASPRT:**

$$\sup_{\theta \in \Theta_i} P_{\theta}(\hat{d} \neq i) \leq e^{-a_i}, \text{ i.e., } a_i = \log(1/\alpha_i) \text{ implies } \hat{d} \in \mathbb{C}(\alpha_0, \alpha_1)$$

- **Asymptotic Approximations for Maximal Error Probabilities of WSLRT:**

$$\sup_{\theta \in \Theta_i} P_{\theta}(\hat{d} \neq i) \sim e^{-a_i} C_i, \quad i = 0, 1 \quad \text{as } a_i \rightarrow \infty,$$

but there are no upper bounds in general.

Asymptotic Optimality of WSLRT and ASLRT

- While asymptotic optimality holds under quite general conditions it is instructive to focus on the ℓ -dimensional exponential family, $p_\theta(X_n) = e^{\theta^\top X_n - b(\theta)}$, in which case both tests minimize *all positive* moments of the sample size
- Define $I_i(\theta) = \inf_{\tilde{\theta} \in \Theta_i} [(\theta - \tilde{\theta})^\top \nabla b(\theta) - (b(\theta) - b(\tilde{\theta}))]$, the minimal KL distance between the point $\theta \notin \Theta_i$ and the subset Θ_i

WSLRT/ASLRT Asymptotic Optimality

Theorem

If the thresholds a_i are so selected that $\sup_{\theta \in \Theta_i} P_\theta(d \neq i) \leq \alpha_i$ and $a_i \sim \log(1/\alpha_i)$, in particular $a_i = 1/\alpha_i$ for ASLRT, then for all $r > 0$ uniformly in $\theta \in \Theta$ as $\max(\alpha_0, \alpha_1) \rightarrow 0$

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_\theta T^r \sim E_\theta \bar{T}^r \sim E_\theta \hat{T}^r \sim \min \left\{ \frac{|\log \alpha_0|}{I_0(\theta)}, \frac{|\log \alpha_1|}{I_1(\theta)} \right\}^r \quad \text{for all } \theta \in I_{\text{in}}$$

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_\theta T^r \sim E_\theta \bar{T}^r \sim E_\theta \hat{T}^r \sim \left\{ \frac{|\log \alpha_i|}{I_i(\theta)} \right\}^r \quad \text{for all } \theta \in \Theta_i, \quad i = 0, 1$$

Consequently, the WSLRT and the ASLRT minimize asymptotically all moments of the sample size uniformly in $\theta \in \Theta$ in the class of tests $\mathbb{C}(\alpha_0, \alpha_1)$.

- For the ASLRT and $r = 1$ this result can be derived from **Pavlov'90** and **Dragalin&Novikov'99** who considered multihypothesis tests.

Example 1: Testing for Normal Mean with Unknown Variance

- $X_n \sim \mathcal{N}(\mu, \sigma^2)$, $H_0 : \mu = 0$, $H_1 : \mu \geq \mu_1$, variance $\sigma^2 > 0$ is a nuisance parameter, and $\theta = (\mu, \sigma^2)$,

$$\lambda_n(\theta, \tilde{\theta}) = \frac{n}{2} \log \left(\frac{\tilde{\sigma}^2}{\sigma^2} \right) + \frac{\sigma^2 - \tilde{\sigma}^2}{2\tilde{\sigma}^2\sigma^2} \sum_{k=1}^n X_k^2 + \frac{\mu\tilde{\sigma}^2 - \tilde{\mu}\sigma^2}{\tilde{\sigma}^2\sigma^2} \sum_{k=1}^n X_k - \frac{\mu^2\tilde{\sigma}^2 - \tilde{\mu}^2\sigma^2}{2\tilde{\sigma}^2\sigma^2} n.$$

- It can be verified that conditions of Theorem hold with

$$I_1(q) = \frac{1}{2} \log[1 + (q_1 - q)^2], \quad 0 \leq q < q_1$$

$$I_0(q) = \frac{1}{2} \log(1 + q^2), \quad q > 0$$

where $q = \mu/\sigma$, $q_1 = \mu_1/\sigma$.

- By Theorem, WSLRT and ASLRT minimize all moments of the stopping time uniformly in θ and

$$\inf_{\delta \in \mathcal{C}(\alpha_0, \alpha_1)} \mathbb{E}_\theta T^r \sim \mathbb{E}_\theta \bar{T}^r \sim \mathbb{E}_\theta \hat{T}^r \sim \begin{cases} \{2|\log \alpha_1| / \log[1 + (q_1 - q)^2]\}^r & \text{if } 0 \leq q \leq q^* \\ \{2|\log \alpha_0| / \log[1 + q^2]\}^r & \text{if } q \geq q^* \end{cases}$$

where q^* is the solution of the equation $\frac{|\log \alpha_1|}{\log[1 + (q_1 - q)^2]} = \frac{|\log \alpha_0|}{\log(1 + q^2)}$

The Kiefer–Weiss Problem: Minimizing Maximal ESS

- **Kiefer&Weiss’57; Weiss’62:** Since SPRT has poor performance in the indifference zone, instead of minimizing ESS at the points θ_i minimize at the worst point θ^* ,

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} \sup_{\theta \in \Theta} E_{\theta} T = \inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_{\theta^*} T,$$

which is usually in the indifference zone.

- Kiefer and Weiss showed that this problem is equivalent to a Bayes problem that can be solved exactly (at least for the one-parameter exponential family) using Bellman’s backward induction.
- **Lorden’76** proposed an asymptotically optimal (very efficient) procedure, **2-SPRT**, which is based on a parallel running of two one-sided SPRTs tuned to θ^*, θ_0 and θ^*, θ_1 :

$$T_0 = \inf \{n : \lambda_n(\theta^*, \theta_1) \geq a_1\}, \quad T_1 = \inf \{n : \lambda_n(\theta^*, \theta_0) \geq a_0\}$$
$$T^* = \min(T_0, T_1), \quad d^* = i \quad \text{if } T^* = T_i$$

- **2-SPRT Asymptotic Optimality:** If $a_i = |\log \alpha_i|$, then for all $r > 0$

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_{\theta^*} T^r \sim E_{\theta^*} [T^*(\theta^*)]^r \sim [n^*(\theta^*)]^r \quad \text{as } \alpha_{\max} \rightarrow 0$$

where θ^* satisfies equation

$$\frac{|\log \alpha_1|}{I(\theta^*, \theta_0)} = \frac{|\log \alpha_0|}{I(\theta^*, \theta_1)} (\equiv n^*(\theta^*))$$

Jack Carl Kiefer 1924–1981

- International leader in math statistics and the foremost worker in optimal experiment design, as well an authority in *mycology*.
- Quite a few fundamental works in sequential analysis (strongly influenced by Wald and Wolfowitz)
- Advisor to many students who became famous scientists in statistics in general and sequential analysis in particular: Jerome Sacks, Gary Lorden, Larry Brown, Richard Schwartz, *etc.*



- **Question 1:** Does it make sense to optimize in the worst case scenario and as a result to obtain a very efficient procedure at a specific point if uniformly asymptotically optimal solutions such as GSLRT, WSLRT, and ASLRT exist?
- **Question 2:** Does it make sense to devise invariant solutions such as *Invariant SPRT* to minimize ESS again in two points?
 - In Example 1 (testing the mean of the normal population with unknown variance) the t -SPRT is asymptotically optimal at $q = q_0$ and $q = q_1$, where $q = \theta/\sigma$, but the SPRT's drawback remains – it is not efficient at other points and performs very poorly around the worst point $q^* \in (q_0, q_1)$.
- **Question 3:** Does it make sense to devise an invariant 2-SPRT that is approximately optimal around the worst point $q^* \in (q_0, q_1)$?
- My personal opinion is negative, and the next slide supports it.
- On the other hand, a lot of work has been done in these directions, which is good, but do we need to continue working in these directions? Perhaps not.

Example 1 (Cont): Testing for Normal Mean with Unknown Variance – Comparison of t -2-SPRT and ASLRT

- t -2-SPRT = parallel running of two one-sided t -tests – asymptotically optimal in the Kiefer–Weiss problem of minimizing the maximal ESS among all invariant tests
- **Scenario:** error probabilities $\alpha_0 = 10^{-3}, \alpha_1 = 10^{-1}, q = \theta/\sigma \geq 0$ (i.e., $|\log \alpha_0|/|\log \alpha_1| = 3$ and $q_1 = 1, q_0 = 0$ (indifference interval $(0, 1)$))
- ASPRT is uniformly better; the difference in the indifference zone is not big, but increases for $q \geq 1$

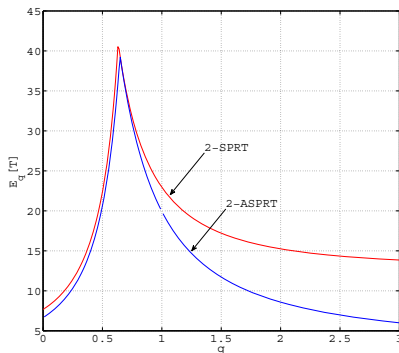


Figure: Asymptotic approximations for expected sample sizes of t -2-SPRT and ASPRT versus $q = \theta/\sigma$ ($\alpha_0 = 10^{-3}, \alpha_1 = 10^{-1}, q_1 = 1, q_0 = 0$).

- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure**
- 6 Multidecision Problems
- 7 Acknowledgements

- Consider testing simple null $H_0 : \theta = \theta_0$ vs composite alternative $H_1 : \theta \in \Theta_1$
- It follows from the previous discussion that the GSLRT, ASLRT and WSLRT asymptotically minimize the ESS $\mathbf{E}_\theta T$ to first order uniformly for all $\theta \in \Theta_1$ in the class $\mathbb{C}(\alpha_0, \alpha_1) = \{\delta : \mathbf{P}_0(d = 1) \leq \alpha_0, \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(d = 0) \leq \alpha_1\}$ as $\alpha_{\max} \rightarrow 0$. In particular, this is true for the WSPRT with an arbitrary weight (prior) $W(\theta)$
- Question:** How to choose $W(\theta)$ to further optimize performance in some sense, and in what sense?
- An appropriate minimax approach is to minimize the expected accumulated Kullback–Leibler (KL) information in the least favorable situation

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} \sup_{\theta \in \Theta_1} (I_\theta \mathbf{E}_\theta T),$$

where $I_\theta \mathbf{E}_\theta T = \mathbf{E}_\theta [\lambda_T(\theta, \theta_0)]$ is the accumulated KL distance between p_θ and p_{θ_0}

- This criterion, being natural in its own right, turns out to be “consistent” since for the WSLRT

$$I_\theta \mathbf{E}_\theta [\bar{T}(\alpha_0, \alpha_1)] = \log \alpha_0^{-1} + \frac{1}{2} \log \log \alpha_0^{-1} + C_\theta(W) + o(1) \quad \text{as } \alpha_{\max} \rightarrow 0,$$

so it is an equalizer rule up to a constant $C_\theta(W)$

- Idea:** Choose $W = W^*$ so that $C_\theta(W^*)$ does not depend on θ , making the WSLRT an equalizer rule up to the negligible term $o(1)$

- **Open Ended (Power 1) Tests:** Continue sampling indefinitely with prescribed probability $P_{\theta_0}(T < \infty) = \alpha$ when H_0 is true and accept H_1 at stopping, which has to be accepted as soon as possible if H_1 is true, i.e., minimize $E_{\theta}T$ for all $\theta \in \Theta_1$ to first order in the class $\mathbb{C}(\alpha) = \{T : P_0(T < \infty) \leq \alpha\}$ on one hand and attempt finding a third-order minimax test T^* on the other,

$$\inf_{\delta \in \mathbb{C}(\alpha)} \sup_{\theta \in \Theta_1} (I_{\theta} E_{\theta} T) = \sup_{\theta \in \Theta_1} (I_{\theta} E_{\theta} T^*) + o(1) \quad \text{as } \alpha \rightarrow 0$$

- **Pollak'78** was the first who set and solved this problem to second order showing that, for a one-parameter exponential family with $\Theta_1 = [\underline{\theta}, \bar{\theta}]$, $\underline{\theta} > 0$, the one-sided WSLRT

$$T_A = \inf \{n : \bar{\Lambda}_n \geq A\}, \quad \bar{\Lambda}_n = \int_{\Theta} \exp\{\lambda_n^{\theta}\} W(d\theta) = \int_{\Theta} \exp\{\theta S_n - b(\theta)n\} W(d\theta)$$

with any prior $W(\theta)$ with continuous density $w(\theta)$ is second-order minimax

$$\inf_{\delta \in \mathbb{C}(\alpha)} \sup_{\theta \in \Theta_1} (I_{\theta} E_{\theta} T) = \sup_{\theta \in \Theta_1} (I_{\theta} E_{\theta} T_A) + O(1) \quad \text{as } \alpha \rightarrow 0$$

- **Nonlinear Renewal Theory** yields, as $A \rightarrow \infty$,

$$I_\theta \mathbf{E}_\theta T_A = \log A + \log \sqrt{\log A} - \frac{1 + \log(2\pi)}{2} + \log \left(e^{\varkappa_\theta} \frac{\sqrt{\ddot{b}(\theta)/I_\theta}}{w(\theta)} \right) + o(1),$$

$$\mathbf{P}_0(T_A < \infty) \sim A^{-1} \int_{\Theta} \zeta_\theta w(\theta) d\theta = \alpha$$

where $\varkappa_\theta = \lim_{a \rightarrow \infty} \mathbf{E}_\theta(\lambda_{\tau_a} - a)$ is the limiting average overshoot in the one-sided SPRT $\tau_a = \inf\{n : \lambda_n^\theta \geq a\}$ and $\zeta_\theta = \lim_{a \rightarrow \infty} \mathbf{E}_\theta[e^{-(\lambda_{\tau_a} - a)}]$

- Taking $A = \alpha^{-1} \int_{\Theta} \zeta_\theta w(\theta) d\theta$ implies $\mathbf{P}_0(T_A < \infty) \sim \alpha$ and setting

$$w(\theta) = w^*(\theta) = \frac{e^{\varkappa_\theta} \sqrt{\ddot{b}(\theta)/I_\theta}}{\int_{\Theta_1} e^{\varkappa_t} \sqrt{\ddot{b}(t)/I_t} dt}, \quad \theta \in \Theta_1,$$

we obtain that the WSLRT is an almost equalizer rule (up to $o(1)$):

$$I_\theta \mathbf{E}_\theta [T_A(w^*)] = |\log \alpha| + \log(\sqrt{|\log \alpha|}) + M + o(1) \quad \text{as } \alpha \rightarrow 0$$

$$M = \log \left(\int_{\Theta_1} \zeta_t e^{\varkappa_t} \sqrt{\ddot{b}(t)/I_t} dt \right) - \frac{1 + \log(2\pi)}{2}$$

Theorem

Assume that $\Theta_1 \subset \Theta$ is a finite subset bounded away from zero. Let $T_A(w^*)$ be the one-sided WSLRT whose mixing density is given by

$$w^*(\theta) = e^{\varkappa_\theta} \sqrt{\det[\nabla^2 b(\theta)]/I_\theta^\ell} \left(\int_{\Theta_1} e^{\varkappa_t} \sqrt{\det[\nabla^2 b(t)]/I_t^\ell} dt \right)^{-1}, \quad \theta \in \Theta_1,$$

and for which $\mathbf{P}_0(T_A(w^*) < \infty) = \alpha$. If the limiting average overshoot $\varkappa_\theta$ is a continuous function on Θ_1 , then as $\alpha \rightarrow 0$

$$\inf_{T \in \mathcal{C}(\alpha)} \sup_{\theta \in \Theta_1} (I_\theta \mathbf{E}_\theta T) = \sup_{\theta \in \Theta_1} \{I_\theta \mathbf{E}_\theta [T_A(w^*)]\} + o(1)$$

$$\sup_{\theta \in \Theta_1} \{I_\theta \mathbf{E}_\theta [T_A(w^*)]\} = \log \alpha^{-1} + \frac{\ell}{2} \log \log \alpha^{-1} + M + o(1)$$

where $M = \log \left(\int_{\Theta_1} \zeta_t e^{\varkappa_t} \sqrt{\det[\nabla^2 b(t)]/I_t^\ell} dt \right) - \frac{\ell}{2} [1 + \log(2\pi)]$. Therefore, the one-sided WSLRT $T_A(w^*)$ is third-order asymptotically minimax.

- **Implementation Problem:** Difficult to implement since in most cases κ_θ can be computed only numerically
- **Alternative:** Discretization of Θ_1 ; also N -sample slippage problems (multiple populations/channels) are discrete by nature:

$$p_0(X_n^1, \dots, X_n^N) = \prod_{j=1}^N f_0^j(X_n^j), \quad p_i(X_n^1, \dots, X_n^N) = f_1^i(X_n^i) \prod_{\substack{j=1 \\ j \neq i}}^N f_0^j(X_n^j), \quad i = 1, \dots, N.$$

- Discrete mixture LR $\bar{\Lambda}_n = \sum_{i=1}^N W_i e^{\lambda_n^i}$ or weighted GLR $\hat{\Lambda}_n = \max_{1 \leq i \leq N} W_i e^{\lambda_n^i}$

$$T_A(W) = \inf \{n : \bar{\Lambda}_n(W) \geq A\} \text{ - WSLRT, } \hat{T}_A(W) = \inf \{n : \hat{\Lambda}_n(W) \geq A\} \text{ - WGPLRT}$$

- **Error Probabilities:**

$$P_0(T_A < \infty) \leq P_0(\hat{T}_A < \infty) \leq \frac{\sum_{i=1}^N W_i}{A} \text{ for any } A > 0$$

$$P_0(T_A < \infty) = \frac{\sum_{i=1}^N W_i \zeta_i}{A} (1 + o(1)), \quad P_0(\hat{T}_A < \infty) \leq \frac{\sum_{i=1}^N W_i \zeta_i}{A} (1 + o(1)), \quad A \rightarrow \infty$$

- **ESS:**

$$I_i \mathbf{E}_i T_A = \log A + \kappa_i - \log W_i + o(1), \quad I_i \mathbf{E}_i \hat{T}_A = \log A + \kappa_i - \log W_i + o(1)$$

- **Almost Equalizer Rules** wrt KL information if $W_i = W_i^* = \frac{e^{\lambda_i}}{\sum_{j=1}^N e^{\lambda_j}}$:

$$I_i \mathbb{E}_i T_A(W^*) = \log A + \log \left(\sum_{j=1}^N e^{\lambda_j} \right) + o(1), \quad A \rightarrow \infty$$

Fellouris&Tartakovsky'12

Theorem

Suppose that $\mathbb{E}_i |\lambda_1^i|^2 < \infty$ and λ_1^i are $\mathbf{P}_i$ -nonarithmetic for $i = 1, \dots, N$. If the mixing distribution is chosen as $W_i^* = \frac{e^{\lambda_i}}{\sum_{j=1}^N e^{\lambda_j}}$ and the threshold as $A = \alpha^{-1} \sum_{i=1}^N W_i^* \zeta_i$, then

$\mathbf{P}_0(T_A < \infty) = \alpha(1 + o(1))$, $\mathbf{P}_0(\hat{T}_A < \infty) \leq \alpha(1 + o(1))$ as $\alpha \rightarrow 0$ and

$$\inf_{T \in \mathbb{C}(\alpha)} \max_{i=1, \dots, N} (I_i \mathbb{E}_i T) = \max_{i=1, \dots, N} \{I_i \mathbb{E}_i [T_A(W^*)]\} + o(1),$$

$$\max_{i=1, \dots, N} \{I_i \mathbb{E}_i [T_A(W^*)]\} = |\log \alpha| + \log \left(\sum_{j=1}^N \zeta_j e^{\lambda_j} \right) + o(1).$$

Therefore, both tests are third-order asymptotically minimax.

- **Gaussian Example:** $H_0 : p_0(x) = \varphi(x)$ and $H_1 : p_i(x) = \varphi(x - i)$ for $i = 1, 2, 3$, where $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is the standard normal pdf, i.e., $N = 3$

Table: The maximal expected accumulated Kullback–Leibler information $\max_i(I_i E_i[T_A(W)])$ for optimal and uniform mixing distributions W^* and W^u

(a) Optimal mixing distribution

α	Monte Carlo	Approximation
10^{-1}	4.99	4.31
10^{-2}	6.36	6.61
10^{-4}	10.99	11.21
10^{-6}	15.65	15.82
10^{-8}	20.33	20.42

(b) Uniform mixing distribution

α	Monte Carlo	Approximation
10^{-1}	5.04	5.52
10^{-2}	6.88	7.82
10^{-4}	11.87	12.42
10^{-6}	16.59	17.03
10^{-8}	21.29	21.63

- Average KL information is somehow smaller for the optimal weight, but can we really conclude something definite?
- So maybe an alternative approach is in order

- **Prior Distribution of Subhypotheses:** $\pi = (\pi_1, \dots, \pi_N)$, $\pi_i = P(p = p_i)$
- **Weights:** $W_i = \pi_i / (\zeta_i I_i)$ (to compensate for overshoots, $\zeta_i = \lim_{a \rightarrow \infty} e^{-(\lambda_{\tau_a}^i - a)}$)

Fellouris & Tartakovsky '13

Theorem

Suppose $E_i |\lambda_1^i|^2 < \infty$. If the mixing distribution is chosen as $W_i = \pi_i / (\zeta_i I_i)$ and $A = \alpha^{-1} \sum_{i=1}^N W_i \zeta_i$, then $P_0(T_A < \infty) = \alpha(1 + o(1))$, $P_0(\hat{T}_A < \infty) \leq \alpha(1 + o(1))$ and

$$\inf_{T \in \mathbb{C}(\alpha)} E^\pi T = \sum_{i=1}^N \frac{\pi_i}{I_i} \left[|\log \alpha| + \log(e^{\lambda_i} \zeta_i) + C_i(\pi) \right] + o(1),$$

$$E^\pi [T_A(W(\pi))] = \inf_{T \in \mathbb{C}(\alpha)} E^\pi T + o(1).$$

where $E^\pi = \sum_{i=1}^N \pi_i E_i$ and $C_i(\pi) = \log \left(\sum_{j=1}^N \frac{\pi_j}{I_j} \right) - \log \frac{\pi_i}{I_i}$. Therefore, both tests are third-order asymptotically π -Bayes.

- **Previous KL Minimality:** $\pi_i^* = \frac{\mathcal{L}_i e^{\lambda_i}}{\sum_{j=1}^N \mathcal{L}_j e^{\lambda_j}}$, $\mathcal{L}_i = I_i \zeta_i$

- **Performance Loss under P_i :**

$$\mathcal{J}_i(\boldsymbol{\pi}) \approx \frac{\mathbb{E}_i[T_{A_\alpha}(\boldsymbol{\pi})] - \mathbb{E}_i[\tau_{a_\alpha}^i]}{\mathbb{E}_i[\tau_{a_\alpha}^i]} = \frac{\log \left[\sum_{j=1}^N (\pi_j / I_j) \right] + \log I_i - \log \pi_i}{|\log \alpha| + \kappa_i + \log \zeta_i}, \quad 1 \leq i \leq N,$$

where $\tau_{a_\alpha}^i$ is the one-sided SPRT for testing p_0 against p_i with error probability α .

- **Priors:** $\hat{\pi}_i \propto \mathcal{L}_i e^{\kappa_i}$, $\pi_i^I \propto I_i$, $\pi_i^{\mathcal{L}} \propto \mathcal{L}_i$, $\pi_i^u \propto 1$

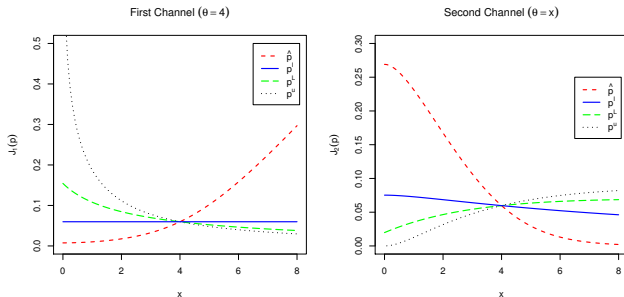


Figure: Performance loss for different prior distributions in a multichannel slippage problem with exponential data $f_0(X_n^i) = e^{-X_n^i}$, $f_1^i(X_n^i) = (1 + \theta_i)^{-1} e^{-X_n^i / (1 + \theta_i)}$

- **Conclusion:** $\boldsymbol{\pi}^I$ and $\boldsymbol{\pi}^{\mathcal{L}}$ lead to a more robust behavior, since the resulting performance loss is relatively low and stable for various signal strengths.

- **Exponential Slippage Model:** 3 channels with $\theta_1 = 0.5, \theta_2 = 1, \theta_3 = 2$; prior π^j (most robust)
- Dashed line = the asymptotic approximation for ESS of WSLRT/GSLRT; solid line = the asymptotic performance of the SPRT; circles = MC WSLRT; triangles = MC GSLRT

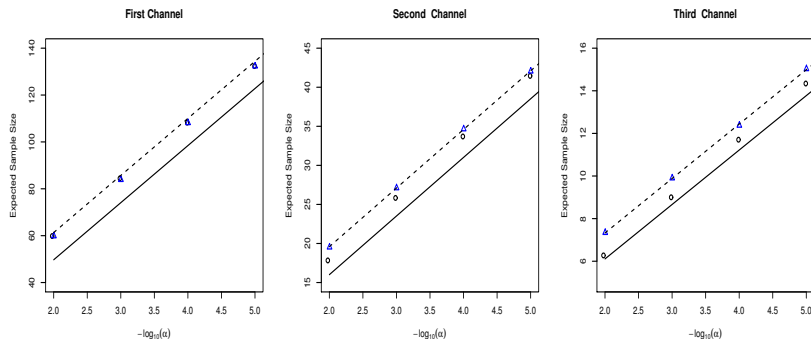


Figure: Simulated expected sample size of WSLRT (circles) and GSLRT (triangles) under P_i against error probability in logarithmic scale, $i = 1, 2, 3$.

- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems**
- 7 Acknowledgements

A General Multidecision Problem

- The ℓ -dimensional parameter $\theta = (\theta_1, \dots, \theta_\ell) \in \Theta \subset \mathbb{R}^\ell$; hypotheses are

$$H_i : \theta \in \Theta_i, \quad i = 0, 1, \dots, N, \quad \Theta = \sum_{i=0}^N \Theta_i + \mathbf{I}_{\text{in}}$$

- Frequentist Problem**

- Probabilities of Errors:** $\alpha_{ij}(\delta, \theta) = \mathbf{P}_\theta(d = j) \mathbf{1}_{\{\theta \in \Theta_i\}}$ (i.e., test terminates with a particular incorrect decision $d = j, j \neq i$) or $\beta_i(\delta, \theta) = \mathbf{P}_\theta(d \neq i) \mathbf{1}_{\{\theta \in \Theta_i\}}$ (i.e., test terminates with any incorrect decision) and the corresponding classes of tests

$$\mathbb{C}(|\alpha_{ij}|) = \left\{ \delta : \sup_{\theta \in \Theta_i} \alpha_{ij}(\delta, \theta) \leq \alpha_{ij}, \quad i, j = 0, 1, \dots, N, \quad i \neq j \right\}$$

$$\mathbb{C}(\beta) = \left\{ \delta : \sup_{\theta \in \Theta_i} \beta_i(\delta, \theta) \leq \beta_i, \quad i = 0, 1, \dots, N \right\}$$

- Find a test $\delta_* = (T_*, d_*)$ (or tests) that would minimize the ESS $\mathbf{E}_\theta T$ or more generally higher moments if the sample size $\mathbf{E}_\theta T^r, r > 1$ approximately (asymptotically) when $\alpha_{\max} = \max_{i,j} \alpha_{ij}, \beta_{\max} = \max_i \beta_i$ are small:

$$\inf_{\delta \in \mathbb{C}(|\alpha_{ij}|)} \mathbf{E}_\theta T^r \sim \mathbf{E}_\theta T_*^r \quad \text{for all } \theta \in \Theta \text{ as } \alpha_{\max} \rightarrow 0$$

- General Non-iid Stochastic Model:**

$$p_\theta(\mathbf{X}_1^n) = \prod_{k=1}^n p_\theta(X_k | \mathbf{X}_1^{k-1}), \quad \mathbf{X}_1^n = (X_1, \dots, X_n)$$

Multihypothesis (Matrix) SPRT for Simple Hypotheses

- **Multiple simple hypotheses:** $H_i : p(\mathbf{X}_1^n) = p_i(\mathbf{X}_1^n), i = 0, 1, \dots, N$

- **Classes of Tests:**

$$\mathbb{C}(|\alpha_{ij}|) = \{\delta : \alpha_{ij}(\delta) = \mathbf{P}_i(d = j) \leq \alpha_{ij}, i, j = 0, 1, \dots, N, i \neq j\},$$

$$\mathbb{C}(\alpha) = \{\delta : \alpha_i(\delta) = \mathbf{P}_i(d \neq i) \leq \alpha_i, i = 0, 1, \dots, N\}$$

- LRs and LLRs:

$$\Lambda_{ij}(n) = \frac{p_i(\mathbf{X}_1^n)}{p_j(\mathbf{X}_1^n)}, \quad \lambda_{ij}(n) = \log \Lambda_{ij}(n)$$

- **Matrix SPRT (MSPRT)** $\delta_N^* = (T_N^*, d_N^*)$: For a threshold/boundary matrix $\mathbf{A} = ||A_{ij}||$, MSPRT is defined as:

stop at the first $n \geq 1$ such that for some i $\Lambda_{ij}(n) \geq A_{ji}$ for all $j \neq i$

and accept the (unique) H_i that satisfies these inequalities.

- Alternatively, MSPRT can be written as

$$T_N^* = \min(T_0, T_1, \dots, T_N), \quad d_N^* = i \text{ if } T_N^* = T_i.$$

$$T_i = \inf \left\{ n \geq 1 : \lambda_{i0}(n) \geq \max_{\substack{1 \leq j \leq N \\ j \neq i}} [\lambda_{j0}(t) + a_{ji}] \right\}, \quad a_{ji} = \log A_{ji}$$

- Note that for $N = 1$ this test coincides with Wald's SPRT.

- **Upper Bounds for Error Probabilities of MSPRT (true in general):**

$$\alpha_{ij}^* = \alpha_{ij}(\delta^*) \leq e^{-a_{ij}}, \quad \alpha_i^* = \alpha_i(\delta^*) \leq \sum_{j \neq i} e^{-a_{ij}}$$

$$a_{ji} = \log(1/\alpha_{ji}) \text{ implies } \delta^* \in \mathbb{C}(\|\alpha_{ij}\|); \quad a_{ji} = a_j = \log(N/\alpha_j) \text{ implies } \delta^* \in \mathbb{C}(\alpha)$$

- **First-order Asymptotic Optimality in the iid Case [T'98]:** If $a_{ji} = \log(1/\alpha_{ji})$ and KL numbers $I_{ij} = \mathbb{E}_i[\lambda_{ij}(1)]$ are positive and finite, then for all $r > 0$

$$\inf_{\delta \in \mathbb{C}(\|\alpha_{ij}\|)} \mathbb{E}_i T^r \sim \mathbb{E}_i [T^*]^r \sim \max_{\substack{1 \leq j \leq N \\ j \neq i}} \left[\frac{|\log \alpha_{ji}|}{I_{ij}} \right]^r \quad \text{as } \alpha_{\max} \rightarrow 0 \quad \text{for all } i = 0, 1, \dots, N.$$

- **Third-order Optimality in the iid Case [Lorden'77]:** If $\mathbb{E}_i[\lambda_{ij}]^2 < \infty$ and if for any $\|B_{ij}\|$ ($B_{ij} > 0, i \neq j$), the thresholds $a_{ji} = \log(B_{ji}/c)$, then

$$\mathbb{E}_i T^*(c) = \inf_{\delta \in \mathbb{C}(\|\alpha_{ij}^*(c)\|)} \mathbb{E}_i T + o(1) \quad \text{as } c \rightarrow 0 \quad \text{for all } i = 0, 1, \dots, N,$$

$$\mathbb{E}_i T^*(c) = \inf_{\delta \in \mathbb{C}(\alpha^*(c))} \mathbb{E}_i T + o(1) \quad \text{as } c \rightarrow 0 \quad \text{for all } i = 0, 1, \dots, N,$$

i.e., the MSPRT $\delta^*(c)$ asymptotically minimizes the expected sample sizes for all hypotheses to within $o(1)$ as $c \rightarrow 0$ among all tests whose error probabilities $\alpha_{ij}(\delta)$ are less or equal to those of δ^* as well as whose error probabilities $\alpha_i(\delta)$ are less or equal to those of δ^* .

- Very accurate asymptotic approximations for ESSs can be obtained using nonlinear renewal theory **[Dragalin, T., Veeravalli'00]**:

$$\mathbb{E}_i T^* = \frac{1}{I_i} \left[a_i + h_{m,i} \sqrt{\frac{a_i}{I_i} + \frac{h_{m,i}^2}{4I_i^2}} + \frac{h_{m,i}^2}{2I_i} + \varkappa_i + C_{m,i} \right] + o(1) \quad \text{as } a_{\min} \rightarrow \infty,$$

where r is the number of symmetric hypothesis, $h_{m,i}$ is the expectation of maximum of m normal random variables, $\varkappa_i$ is an average limiting overshoot, and $C_{m,i}$ is a constant.

- Unfortunately, for error probabilities we can obtain only bounds also accounting for overshoots, and these bounds are not too accurate.

- **Notation:** $\mathcal{I}_{ij}(n) = \mathbb{E}_i[\lambda_{ij}(n)]$ – the accumulated KL information in the trajectory $\mathbf{X}_1^n$.
- **Assumption 1:** There exist a non-negative increasing function $\psi(n)$ ($\psi(\infty) = \infty$) and positive and finite numbers $I_i, i = 0, 1, \dots, N$ such that

$$\frac{\lambda_{ij}(n)}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_i\text{-a.s.}} I_{ij} \quad \text{for } i, j = 0, 1, \dots, N, i \neq j$$

- **Asymptotic Lower Bounds:** Under A 1 for all $r > 0$ and all $i = 0, 1, \dots, N$

$$\inf_{\delta \in \mathbb{C}(\|\alpha_{ij}\|)} \mathbb{E}_i T^r \geq \left[\Psi \left(\max_{j \neq i} \frac{|\log \alpha_{ji}|}{I_{ij}} \right) \right]^r (1 + o(1)) \quad \text{as } \max_{i,j} \alpha_{ij} \rightarrow 0, \quad (1)$$

$$\inf_{\delta \in \mathbb{C}(\alpha)} \mathbb{E}_i T^r \geq \left[\Psi \left(\max_{j \neq i} \frac{|\log \alpha_j|}{I_{ij}} \right) \right]^r (1 + o(1)) \quad \text{as } \max_i \alpha_i \rightarrow 0,$$

where $\Psi(t)$ is the inverse function for $\psi(t)$.

- **Upper Bounds and Optimality in the Non-iid Case [T'98]:** A proof of MSPRT's asymptotic optimality is based on showing that the above lower bounds are attained for the MSPRT with thresholds $a_{ji} = \log(1/\alpha_{ji})$, i.e.,

$$\inf_{\delta \in \mathbb{C}(\|\alpha_{ij}\|)} \mathbb{E}_i T^r \sim \mathbb{E}_i [T^*]^r \sim \max_{\substack{1 \leq j \leq N \\ j \neq i}} \left[\frac{|\log \alpha_{ji}|}{I_{ij}} \right]^r \quad \text{as } \alpha_{\max} \rightarrow 0 \quad \text{for all } i = 0, 1, \dots, N.$$

- But it requires much stronger conditions than the SLLN for the LLRs, which are discussed on the next slide

- **Assumption 2:** Strengthening *Assumption 1* into the r -quick version [Strassen'67, Lai'76]

$$\mathbf{E}_i[L_{ij}(\varepsilon)]^r < \infty \quad \text{for all } \varepsilon > 0, \text{ where } L_{ij}(\varepsilon) = \sup \left\{ n : \left| \frac{1}{\psi(n)} \lambda_{ij}(n) - I_{ij} \right| \right\} > \varepsilon$$

is the last time the normalized LLR $\lambda_{ij}(n)/\psi(n)$ leaves the region $[I_{ij} - \varepsilon, I_{ij} + \varepsilon]$.

- This r -quick convergence condition will be written as

$$\lambda_{ij}(n)/\psi(n) \xrightarrow[n \rightarrow \infty]{\mathbf{P}_i - r\text{-quickly}} I_{ij}$$

- Note that the a.s. convergence condition is equivalent to $\mathbf{P}_i \{L_{ij}(\varepsilon) < \infty\} = 1$
- The r -quick convergence condition is very close to the following condition:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_i \{ |\lambda_{ij}(n)/\psi(n) - I_{ij}| > \varepsilon \} < \infty \quad \text{for all } \varepsilon > 0,$$

which is sufficient for obtaining both lower and upper bounds, i.e., for asymptotic optimality wrt to moments of the stopping time distribution.

- Also, the following two *one-sided* (right and left tail) conditions are sufficient:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_i \left\{ \frac{1}{\psi(n)} \lambda_{ij}(n) < I_{ij}(1 - \varepsilon) \right\} < \infty \quad \text{for all } \varepsilon > 0$$

$$\lim_{L \rightarrow \infty} \mathbf{P}_i \left\{ \frac{1}{\psi(L)} \max_{1 \leq n \leq L} \lambda_{ij}(n) \geq (1 + \varepsilon) I_{ij} \right\} = 1 \quad \text{for all } \varepsilon > 0$$

MSPRT Asymptotic Optimality [Tartakovsky'98]

Theorem

Assume that there exist finite positive numbers I_{ij} , $i, j = 0, 1, \dots, N$, $i \neq j$ and an increasing nonnegative function $\psi(n)$ such that for some $r > 0$

$$\lambda_{ij}(n)/\psi(n) \xrightarrow[n \rightarrow \infty]{P_{i-r-\text{quickly}}} I_{ij} \quad \text{for all } i, j = 0, 1, \dots, N, i \neq j$$

(or any other relaxation given on the previous slide holds).

(i) For all $i = 0, 1, \dots, N$

$$E_i[T^*]^r \sim \left[\Psi \left(\max_{j \neq i} \frac{a_{ji}}{I_{ij}} \right) \right]^r \quad \text{as } \min_{j,i} a_{ji} \rightarrow \infty.$$

(ii) If the thresholds are selected so that $\alpha_{ij}(\delta^*) \leq \alpha_{ij}$ and $a_{ji} \sim \log(1/\alpha_{ji})$, in particular $a_{ji} = \log(1/\alpha_{ji})$, then for all $i = 0, 1, \dots, N$

$$\inf_{\delta \in \mathbb{C}(|\alpha_{ij}|)} E_i[T]^r \sim E_i[T^*]^r \sim \left[\Psi \left(\max_{j \neq i} \frac{|\log \alpha_{ji}|}{I_{ij}} \right) \right]^r \quad \text{as } \max_{i,j} \alpha_{ij} \rightarrow 0.$$

(iii) If the thresholds are selected so that $\alpha_i(\delta^*) \leq \alpha_i$ and $a_{ji} \sim \log(1/\alpha_j)$, in particular $a_{ji} = \log(N/\alpha_j)$, then for all $i = 0, 1, \dots, N$

$$\inf_{\delta \in \mathbb{C}(\alpha)} E_i[T]^r \sim E_i[T^*]^r \sim \left[\Psi \left(\max_{j \neq i} \frac{|\log \alpha_j|}{I_{ij}} \right) \right]^r \quad \text{as } \max_i \alpha_i \rightarrow 0.$$

Consequently, the MSPRT minimizes asymptotically moments of the stopping time distribution up to the order r for all hypotheses $H_0, \dots, H_N$ in the corresponding classes of tests.

- Let $\mathbf{X}_t = (X_{1,t}, \dots, X_{N,t})$, $N \geq 2$, be an N -component process observed either in discrete or continuous time. The component $X_{j,t}$ corresponds to the observation in the j -th population (or channel, sensor, data stream, etc.), and it is assumed components are *mutually independent* but may have a fairly *general* structure. Write $\mathbf{X}^t = \{\mathbf{X}_u, 0 \leq u \leq t\}$ and $\mathbf{X}_k^t = \{X_{k,u}, 0 \leq u \leq t\}$.

- Hypotheses:**

H_0 : all components have the same distribution (measure) P_0

H_i : the i -th representative has a different distribution (measure) P_i

- LLR:**

$$\lambda_{ij}(t) := \log \frac{dP_i^t}{dP_0^t}(\mathbf{X}^t) = \log \frac{dP_i^t}{dP_0^t}(\mathbf{X}_i^t) - \log \frac{dP_j^t}{dP_0^t}(\mathbf{X}_i^t)$$

depends on the observation process $\mathbf{X}_t$ through the components $X_{i,t}$ and $X_{j,t}$.

- Variety of Applications:**

- Ranking and selection problem in which the goal is to select the best population (either all populations are identical or one of them has slipped to the right of the rest, and if so, which one is the best?)
- Detection and identifications of objects in multi-channel or multi-sensor systems/networks (detect an object, if any, and identify a channel where it is located)

Example 1: Testing for a Nonhomogeneous AR Sequence in Multichannel System

- If H_i is true $X_{i,n} = S_i(n) + \xi_i(n)$; if wrong $X_{i,n} = \xi_i(n)$, where $S_i(n)$ are deterministic and noise processes $\{\xi_i(n)\}_{n \geq 1}$ are stable first-order AR Gaussian sequences $\xi_i(n) = \gamma_i \xi_i(n-1) + \zeta_i(n)$, $\zeta_i(n) \sim \mathcal{N}(0, \sigma_i^2)$
- LLRs

$$\lambda_{ij}(n) = \sum_{k=1}^n \tilde{S}_i(k) \tilde{X}_{i,k} - \sum_{k=1}^n \tilde{S}_j(k) \tilde{X}_{j,k} - \frac{1}{2} \sum_{k=1}^n [\tilde{S}_i^2(k) - \tilde{S}_j^2(k)],$$
$$\tilde{X}_i(k) = \sigma_i^{-1} X_i(k) - \gamma_i X_i(k-1), \quad \tilde{S}_i(k) = \sigma_i^{-1} [S_i(k) - \gamma_i S_i(k-1)]$$

- If $\lim_{n \rightarrow \infty} n^{-\ell} \sum_{k=1}^n \tilde{S}_i^2(k) = q_i$ for some $0 < q_i < \infty$ and $\ell > 0$, then

$$n^{-\ell} \lambda_{ij}(n) \rightarrow (q_i + q_j)/2 \quad \text{P}_i - r - \text{quickly for all } r > 0 \quad (q_0 = 0)$$

- By Theorem, the MSPRT is asymptotically optimal minimizing all positive moments of the sample size,

$$\inf_{\delta \in \mathbb{C}(|\alpha_{ij}|)} \mathbf{E}_i T^r \sim \mathbf{E}_i [T^*]^r \sim \begin{cases} (2|\log \alpha_0|/q_{\min})^{r/\ell} & \text{if } i = 1, \dots, N \\ (2|\log \alpha_1|/q_{\min})^{r/\ell} & \text{if } i = 0, \end{cases}$$

where $q_{\min} = \min_{1 \leq i \leq N} q_i$ is the minimal SNR in channels and where we assumed that the false alarm probabilities $\alpha_{0j} = \alpha_0$ and the misdetection probabilities $\alpha_{j0} = \alpha_1$ do not depend on channel numbers.

Example 2: Continuous-time Detection–Identification in a Multichannel System

- The i -th component admits a stochastic differential equation

$$dX_{i,t} = \begin{cases} S_i(t)dt + V_i(t)dt + \sigma_i dW_i(t) & \text{if } H_i \text{ is true} \\ V_i(t)dt + \sigma_i dW_i(t) & \text{if } H_i \text{ is wrong,} \end{cases}$$

where $S_i(t)$ is a deterministic signal, $W_i(t)$ is a standard Brownian motion, and $V_i(t)$ is an L_2 -continuous Gaussian process ($\sigma_i > 0$)

- The LLR (under the hypothesis H_j) can be written as

$$\lambda_{k0}^j(t) = \begin{cases} \frac{1}{\sigma_k} \int_0^t \tilde{S}_k(u) d\tilde{W}_k(u) + \frac{1}{2}\mu_k(t) & \text{if } k = j \\ \frac{1}{\sigma_k} \int_0^t \tilde{S}_k(u) d\tilde{W}_k(u) - \frac{1}{2}\mu_k(t) & \text{if } k \neq j, \end{cases}$$

where $\mu_k(t) = \sigma_k^{-2} \int_0^t \tilde{S}_k^2(u) du$ is the SNR in the k -th channel at the output of the whitening filter and $\tilde{W}_k(t)$ is the Brownian motion (innovative process) statistically indistinguishable from $W_k(t)$.

- Assuming that $\lim_{t \rightarrow \infty} \mu_k(t)/t^\ell = q_k$ for some $\ell > 0$, we obtain that for all $r > 0$

$$t^{-\ell} \lambda_{ij}(t) \xrightarrow[t \rightarrow \infty]{P_i - r - \text{quickly}} \frac{1}{2}(q_i + q_j), \quad i, j = 0, 1, \dots, N, i \neq j \quad (q_0 = 0)$$

- By Theorem, the MSPRT minimizes asymptotically all positive moments of the stopping time distribution.

Example 3: Continuous-time Detection of Stochastic Signal

- The i -th component has the stochastic differential

$$dX_{i,t} = \begin{cases} S_i(t)dt + \sigma_i dW_i(t) & \text{if } H_i \text{ is true} \\ \sigma_i dW_i(t) & \text{if } H_i \text{ is wrong,} \end{cases}$$

where $W_i(t)$, $i = 1, \dots, N$ are mutually independent standard Brownian motions, and the signals $S_1(t), \dots, S_N(t)$ are statistically independent Markov Gaussian processes,

$$\mathbb{E}[S_i(t)] = 0, \quad \mathbb{E}[S_i(t)S_i(t+u)] = d_i^2 \exp(-\rho_i |u|), \quad \mathbb{E}[S_i(t)S_j(t)] = 0, \quad i \neq j \quad (\rho_i, d_i > 0)$$

- The LLR can be written as

$$\lambda_{ij}(t) = \frac{1}{\sigma_i^2} \int_0^t \hat{S}_i(u) dX_i(u) - \frac{1}{\sigma_j^2} \int_0^t \hat{S}_j(u) dX_j(u) - \frac{1}{2} \int_0^t \left\{ [\hat{S}_i^2(u)/\sigma_i^2] - [\hat{S}_j^2(u)/\sigma_j^2] \right\} du,$$

where $\hat{S}_i(t) = \mathbb{E}_i[S_i(t)|\mathbf{X}_i^t]$ is the optimal mean-square filtering estimate of $S_i(t)$ that satisfies the Kalman equations.

- Cumbersome argument shows that in the symmetric case $\rho_i = \rho, d_i = d, \sigma_i = \sigma$ (only for simplicity!) for all $r > 0$

$$\frac{1}{t} \lambda_{ij}(t) \xrightarrow[t \rightarrow \infty]{P_{i-r} \text{--quickly}} I_{ij} = \begin{cases} \frac{Q}{\sqrt{1+Q}(1+\sqrt{1+Q})} \frac{d^2}{2\sigma^2} & \text{for } i \neq 0 \\ \frac{Q}{\sqrt{1+Q}(1+\sqrt{1+Q})^2} \frac{d^2}{2\sigma^2} & \text{for } i = 0, \end{cases}$$

where $Q = 2d^2/(\rho\sigma^2)$ is the SNR.

- By Theorem, the MSPRT minimizes asymptotically all positive moments of the stopping time distribution.

Example 4: Generalized Multisample t -Test

- $H_0 : X_{k,n} = \theta + \xi_k(n), k = 1, \dots, N$ and $H_i : X_{i,n} = \theta + \mu_i + \xi_i(n); X_{k,n} = \theta + \xi_k(n), k \neq i$, where $\xi_i(n) \sim \mathcal{N}(0, \sigma^2)$ and both θ and σ^2 are unknown nuisance parameters. We wish to test $H_0 : \mu_k/\sigma = 0 \forall k = 1, \dots, N$ and $H_i : \mu_k/\sigma = 0$ for $k \neq i$ and $\mu_i/\sigma = q_i, i = 1, \dots, N$
- The problem is invariant under change in location and scale, the maximal invariant is a vector of $Y_{k,n} = [X_{k,n} - X_{1,1}]/[X_{2,1} - X_{1,1}], Y_{1,1} = 0, Y_{2,1} = 1$
- LLRs:

$$\lambda_{ij}(n) = \log \frac{J_n^N(q_i t_{i,n}^N)}{J_n^N(q_j t_{j,n}^N)} - \frac{N-1}{2N} (q_i^2 - q_j^2) n, \quad J_n^N(z) = \int_0^\infty u^{-2} \exp[nNf(u, z)] du$$

$$\bar{X}_{nN} = \frac{1}{nN} \sum_{k=1}^n \sum_{j=1}^N X_{j,k}, \quad t_{i,n}^N = \frac{(Nn)^{-1} \sum_{k=1}^n [X_{i,k} - \bar{X}_{nN}]}{\left\{ (Nn)^{-1} \sum_{k=1}^n \sum_{j=1}^N [X_{j,k} - \bar{X}_{nN}]^2 \right\}^{1/2}},$$

- Quite tedious argument shows that $n^{-1} \lambda_{ij}(n)$ converges $P_i - r$ – quickly to

$$I_{ij} = N [\phi(Q_{ii}) - \phi(Q_{ij})] - \frac{N-1}{2N} (q_i^2 - q_j^2), \quad \text{where } Q_{ii} = \frac{(N-1)q_i^2}{N^2 \sqrt{1 + (N-1)q_i^2/N^2}},$$

$$Q_{ij} = -\frac{q_i q_j}{N^2 \sqrt{1 + (N-1)q_i^2/N^2}}, \quad \phi(y) = \frac{1}{4} t_n \left(y + \sqrt{4 + y^2} \right) + \log \left(y + \sqrt{4 + y^2} \right).$$

- By Theorem, the MSPRT minimizes asymptotically all positive moments of the stopping time distribution.

- **Deterministic signal plus white noise model:**

$$dX(t) = S(t) + dW(t)$$

where $W(t)$ is a standard Brownian motion and $S_t = [A^2(1+t)]^{-1/2}$, so that the energy of the signal grows very slowly,

$$\mu(t) = q\psi(t) = q \log(1+t), \text{ where } q = A^2/\sigma^2.$$

- The normalized LLR converges almost surely to $q/2$:

$$\frac{1}{\log(1+t)} \lambda_t \xrightarrow[t \rightarrow \infty]{P_1\text{-a.s.}} q/2, \quad \frac{1}{\log(1+t)} \lambda_t \xrightarrow[t \rightarrow \infty]{P_0\text{-a.s.}} -q/2$$

- However, the r -quick convergence does not hold for any $r \geq 1$ since

$$\int_0^\infty P_i(|W_t| > \varepsilon \log(1+t)) dt = 4 \int_0^\infty u (e^{u^2} - 1) \Phi(-\varepsilon u / \sqrt{q/2}) du = \infty.$$

- For symmetric SPRT with $-a_0 = a_1 = a$ and $q = 1$

$$E_i T^* = \frac{\cosh(a/2)}{\cos(\sqrt{7}a/2) - 1} \text{ if } 0 < a < 7^{-1/2}\pi, \quad E_i T^* = \infty \text{ if } a \geq 7^{-1/2}\pi,$$

so the ESS is **infinite** for the error probability $\alpha \leq (1 + e^{\pi/\sqrt{7}})^{-1} \approx 0.301$.

- 1 Abraham Wald (1902–1950) – Founder of Sequential Analysis
- 2 Testing Two Simple and Two Composite Hypotheses (iid Case)
- 3 Hypothesis Testing with Indifference Zone
- 4 Hypothesis Testing with and without Indifference Zone (Unified Theory)
- 5 Nearly Minimax Sequential Tests with Kullback–Leibler Information Cost Measure
- 6 Multidecision Problems
- 7 Acknowledgements**

- This work was supported at the University of Southern California by:
 - The U.S. Army Research Office grant W911NF-13-1-0073
 - The U.S. National Science Foundation under grants DMS-1221888, CCF-0830419 and EFRI-1025043
 - The Defense Threat Reduction Agency grant HDTRA1-10-1-0086
 - The U.S. Air Force Office of Scientific Research MURI grant FA9550-10-1-0569
 - The U.S. Defense Advanced Research Projects Agency under grant W911NF-12-1-0034

THANK YOU!