When enough is enough: early stopping of biometrics error rate testing

Michael E. Schuckers, Emily Sheldon and Hilary Hartson

St. Lawrence University, and

Center for Identification Technology Research (CITeR)

Funding for this work comes from NSF grants CNS-0325640 and CNS-0520990. CNS-0325640 is cooperatively funded by the National Science Foundation and the United States Department of Homeland Security. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Overall Goal

Statistical methodology for making **interim** decisions regarding whether or not a device's error rate meets a threshold. These interim decisions will result in one of three outcomes

1. Accept device
2. Reject device
3. Continue testing

STOP TESTING EARLY
SAVE $$$$$$

Inspiration

Not usual academic/vendor test for improving performance

Testing of a biometric device for passing a test, achieve a threshold

Example

US TSA Qualified Products List (QPL)

Notation

n = number of comparison pairs to be tested

m_i = number of times i^{th} comparison pair is tested (here m_i =m)

δ = error rate under study, $\hat{\delta}$ will be its estimate

δ_0 = specific value to be tested

δ_1 be value below which clearly acceptable, with $\delta_0 > \delta_1$

Y_{ij} is (error (1) - no error (0)) binary response for j^{th} attempt by i^{th} comparison pair.

$X_i = \sum_{j=1}^{m_i} Y_{ij}$ $\boldsymbol{\alpha}$ is significance level, probability of Type I error

$\mathbf{1} - \boldsymbol{\beta}$ is power, one minus probability of Type II error

Correlation Structure Assumed

$$Corr(Y_{ij}, Y_{i'j'}) = \begin{cases} 1 & \text{if} \quad i = i', j = j' \\ \rho & \text{if} \quad i = i', j \neq j' \\ 0 & \quad otherwise. \end{cases} \qquad (1)$$

where i is for a comparison pair and j is for the decision

- Correlation on binary scale, different from Pearson's ρ
- Correlation structure of Beta-binomial distribution
- Schuckers (2003) showed Beta-binomial fit both FMR and FNMR
- Assume (here) that no correlation between AB decisions and BC decisions for FMR
- Only correlation between AB decisions and other AB decisions for FMR
- Implicit in 'user-specific' bootstrap (Poh, to appear?) or 'subsets bootstrap' (Bolle *et al* 2004)

Relevant Statistical Theoretical Background

Non-random sampling

Wald(1947), Bartlett(1946), Whitehead(1979), Lan and DeMets (1983), Sen and Ghosh(1991), Todd and Whitehead(1997), Everson and Bradlow (2002), Sooriyarachchi et al. (2004), Govindarajulu (2004)

Motivation

Military testing (WWII)

Clinical trials

Wald Approach

$H_0 : \delta = \delta_0$

$H_a : \delta = \delta_1$

For testing error rate, δ against some threshold $\delta_0 > \delta_1$.

Have to choose value for δ_1 to evaluate data

Set bounds for the entire trial, check as many times as you want.

Compare $P(Data|\delta = \delta_0)$ vs. $P(Data|\delta = \delta_1)$

Specifically ratio of former to latter

After collecting some data, calculate ratio

Large $\mapsto$ Accept H_0

Small $\mapsto$ Accept H_1

Neither keep testing

Sequential Probability Ratio Test(SPRT)

The test at time t is then

$$LR_t = \frac{L(\delta_0, \hat{\rho}_0 \mid \mathbf{Y}_t)}{L(\delta_1, \hat{\rho}_1 \mid \mathbf{Y}_t)} \tag{2}$$

where and $L(\delta, \rho \mid \mathbf{X}_t)$ is the likelihood function at time t
and $\hat{\rho}$ is the maximum likelihood estimate based on $\hat{\rho}_i = \arg\max_\rho L(\delta_i, \rho)$.

The Beta-binomial likelihood is then

$$\begin{aligned} L(\delta, \rho \mid \mathbf{X}_t) \quad &= \quad \prod_{i=1}^{n} \left\{ \binom{m_i^{(t)}}{X_i^{(t)}} \frac{\Gamma((1-\rho)\rho^{-1})}{\Gamma(\delta(1-\rho)\rho^{-1})} \; \frac{\Gamma(\delta(1-\rho)\rho^{-1} + X_i^{(t)})}{\Gamma((1-\delta)(1-\rho)\rho^{-1})} \right. \\ &\times \quad \left. \frac{\Gamma((1-\delta)(1-\rho)\rho^{-1} + m_i^{(t)} - X_i^{(t)})}{\Gamma((1-\rho)\rho^{-1} + m_i^{(t)})} \right\} \end{aligned}$$

where $m_i^{(t)}$ and $X_i^{(t)}$ are number of decisions and the number of errors, respectively, for the i^{th} individual at time t.

More SPRT details

Following Wald (1947), at time t the following decisions are made:

1. Accept H_0 if $LR > B$,

2. Accept H_1 if $LR < A$, and

3. Continue collecting data if $A < LR < B$

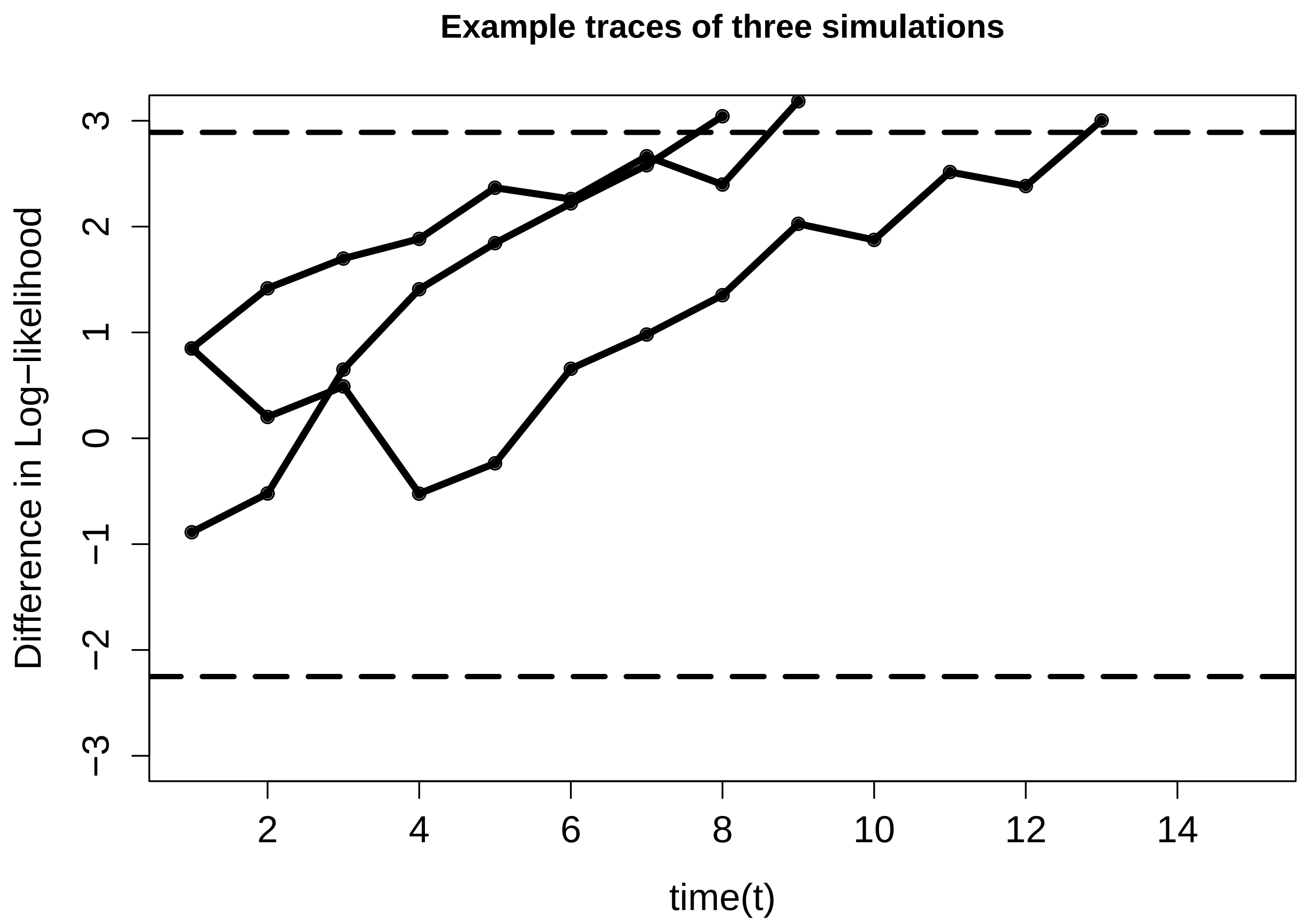
where $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$ with

$$P(\text{Accept } H_0 \mid \delta = \delta_1) = \beta \text{ and}$$
$$P(\text{Reject } H_0 \mid \delta = \delta_0) = \alpha.$$

A and B are derived from Wald's original formulation of the SPRT. Alternatively, we can make decision and comparisons on a log-scale.

$$ln(A) < \ell(\delta_0, \hat{\rho}) - \ell(\delta_1, \hat{\rho}) < ln(B). \tag{3}$$

where $\ell(\delta, \rho) = ln(L(\delta, \rho))$


Example traces of three simulations
Difference in Log−likelihood
time(t)
−3
−2
−1
0
1
2
3
2
4
6
8
10
12
14

Applying SPRT to Biometrics Data

Biometrics Data from Ross and Jain (2003)

Three modalities:

Fingerprint

Face

Hand Geometry

FNMR data:

50 individuals (comparison pairs)

10 decisions/comparison pair

FMR data:

50×49 comparison pairs

5 decisions/comparison pair

Simulations

For both FNMR and FMR, we simulated from the data:

Selecting comparison pairs without replacement

Using $\delta_0 = 0.20, 0.10, 0.05, 0.01$

$\delta_1 = 0.5\delta_0$ i.e. $0.10, 0.05, 0.025, 0.005$, respectively

Found threshold, τ close to δ_0 and δ_1

$m_i = m$ and $m = 10$ for FNMR and $m = 5$ for FMR

All 3 modalities

1000 repetitions of each scenario

Recorded stopping times at which final decision made

One More Thing

Power calculation: Given α, β, m, δ_0, δ_1, and estimate of ρ we can determine the number of comparison pairs to be tested.

$$
\begin{aligned}
n^* &= \Big\lceil m^{-1}(\delta_0 - \delta_1)^{-2} \\
&\times \Big(z_{1-\alpha}\sqrt{\delta_0(1-\delta_0)(1+\rho(m-1))} \\
&+ z_{1-\beta}\sqrt{\delta_1(1-\delta_1)(1+\rho(m-1))} \Big)^2 \Big\rceil
\end{aligned} \tag{4}
$$

This is a *power* calculation which is a generalization of a *sample size* calculation.

Sample Results: Face FMR

Error Type	δ_0	δ_1	True error rate	ρ	Correct decision	n^*	$n_{0.75}$	$n_{0.975}$
FMR	0.100	0.050	0.0984	0.0000	H_0	48	31	81
FMR	0.050	0.025	0.0510	0.0000	H_0	100	57	144
FMR	0.010	0.005	0.0098	0.0000	H_0	517	321	742
FMR	0.200	0.100	0.0984	0.0000	H_1	22	23	44
FMR	0.100	0.050	0.0510	0.0000	H_1	48	43	92
FMR	0.020	0.010	0.0098	0.0000	H_1	257	197	394

Sample Results: Hand Geometry FNMR

Error Type	δ_0	δ_1	True error rate	ρ	Correct decision	n^*	$n_{0.75}$	$n_{0.975}$
FNMR	0.100	0.050	0.1020	0.0514	H_0	35	17	38
FNMR	0.050	0.025	0.0500	0.0222	H_0	60	33	91
FNMR	0.010	0.005	0.0100	0.0000	H_0	259	149	425
FNMR	0.200	0.100	0.1020	0.0514	H_1	16	15	31
FNMR	0.100	0.050	0.0500	0.0222	H_1	29	26	61
FNMR	0.020	0.010	0.0100	0.0000	H_1	129	106	231

Summary of Results

- SPRT performs very well all three modalities
- Type I and II error rates on SPRT are at nearly nominal levels
- On average (both median and mean) SPRT outperforms fixed sample size
- $n_{0.50}/n* \approx 0.50 \rightarrow$ Median savings 50%
- 75% of time savings more than 25%.
- Possible that stopping time $> n*$.
- Further testing (beyond this paper) echoes and extends these results
- ρ is nearly always **very** small

Next Steps

- Correlation Structure and consequences
- Alternative Sequential Approaches (Lan and DeMets)
- Practical Issues for Implementation: Choice of δ_1

Grazie!!

Domande o commenti?

`schuckers@stlawu.edu`

Correlation Structure: FMR Assymetric Matcher

$$
Corr(Y_{ik\ell}, Y_{i'k'\ell'}) = \begin{cases} 1 & \text{if} \quad i = i', k = k', \ell = \ell' \\ \eta & \text{if} \quad i = i', k = k', \ell \neq \ell' \\ \omega_1 & \text{if} \quad i = i', k \neq k', i \neq k, i \neq k' \\ \omega_2 & \text{if} \quad i \neq i', k = k', k \neq i, k \neq i' \\ \omega_3 & \text{if} \quad i = k', i' \neq k, i \neq i', i \neq k \\ \omega_3 & \text{if} \quad i' = k, i \neq k', i' \neq i, i' \neq k' \\ \xi_1 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell = \ell' \\ \xi_2 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell \neq \ell' \\ 0 & \quad\quad \text{otherwise.} \end{cases} \tag{5}
$$