Old and New Results in Robust Hypothesis Testing

Bernard C. Levy
Department of Electrical and Computer Engineering
University of California, Davis

January 12, 2008
Outline

- Binary Hypothesis Testing
- Robust Hypothesis Testing
- Huber’s Clipped LR Test
- Robustness with a KL Divergence Tolerance
- Simulations
Binary Hypothesis Testing

- Consider observation $Y \in \mathbb{R}$ where under hypothesis $H_0$, $Y$ has probability density $f_0(y)$ and under $H_1$, it has density $f_1(y)$.

- Given $Y$, we need to decide between $H_1$ or $H_0$. We use a randomized decision rule $\delta \in \mathcal{D}$, where given $Y = y$, we select $H_1$ with probability $\delta(y)$ and $H_0$ with probability $1 - \delta(y)$, where $0 \leq \delta(y) \leq 1$ for all $y \in \mathbb{R}$. Note that set $\mathcal{D}$ is convex.

- Bayesian hypothesis testing assumes a priori probabilities

$$
\pi_0 = P[H_0] , \quad \pi_1 = 1 - \pi_0 = P[H_1]
$$

and costs $C_M$ and $C_F$ for a miss (deciding $H_0$ when $H_1$ holds) and a false alarm (deciding $H_1$ when $H_0$ holds), respectively.
Binary hypothesis testing (cont’d)

Let

\[ P_F(\delta, f_0) = \int_{-\infty}^{\infty} \delta(y) f_0(y) dy \]

\[ P_M(\delta, f_1) = \int_{-\infty}^{\infty} (1 - \delta(y)) f_1(y) dy \]

denote the probability of false alarm and of a miss under \( H_0 \) and \( H_1 \), respectively. The optimal Bayesian test minimizes the risk

\[ R(\delta, f_0, f_1) = C_F P_F(\delta, f_0) \pi_0 + C_M P_M(\delta, f_1) \pi_1 \]

\[ = C_M \pi_1 + \int_{-\infty}^{\infty} \delta(y) [C_F \pi_0 f_0(y) - C_M \pi_1 f_1(y)] dy. \]
Optimal Bayesian test: Let $L(y) = \frac{f_1(y)}{f_0(y)}$ = likelihood ratio (LR) and $\tau_B = C_F \pi_0 / (C_M \pi_1)$. The test minimizing the Bayesian risk is given by

$$\delta(y) = \begin{cases} 
1 & L(y) > \tau_B \\
0 & L(y) < \tau_B \\
\text{arbitrary} & L(y) = \tau_B
\end{cases},$$

and randomization is not needed.

Neyman-Pearson test (of type I): Minimizes $P_M(\delta, f_1)$ under the constraint $P_F(\delta, f_0) \leq \alpha$. Solution:

$$\delta(y) = \begin{cases} 
1 & L(y) > \tau \\
0 & L(y) < \tau \\
p & L(y) = \tau
\end{cases}.$$
Binary hypothesis testing (cont’d)

- The threshold $\tau$ and randomization probability $p$ are selected as follows. Let $F_L(\ell|H_0) = P[L \leq \ell|H_0]$ denote the cumulative probability distribution of likelihood ratio $L$ under $H_0$. Then $F_L(\tau|H_0) = 1 - \alpha$ and $p = 0$ if $1 - \alpha$ is in the range of $F_L(\ell|H_0)$, and if

  $$F_L(\tau_-|H_0) < 1 - \alpha < F_L(\tau|H_0)$$

  then

  $$p = \frac{F_L(\tau|H_0) - (1 - \alpha)}{F_L(\tau|H_0) - F_L(\tau_-|H_0)}$$

- Both the Bayesian and NP tests rely on the LR function $L(y)$. Only the threshold selection changes.
Outline

- Binary Hypothesis Testing
- Robust Hypothesis Testing
- Huber’s Clipped LR Test
- Robustness with a KL Divergence Tolerance
- Simulations
Robust Hypothesis Testing

- The actual probability densities $g_0$ and $g_1$ of observation $Y$ under $H_0$ and $H_1$ may differ slightly from the nominal densities $f_0$ and $f_1$. Assume $g_j \in \mathcal{F}_j$, where $\mathcal{F}_j$ denotes a convex neighborhood of $f_j$ for $j = 0, 1$.

- Let $\mathcal{F} = \mathcal{F}_0 \times \mathcal{F}_1$. The robust Bayesian hypothesis problem can be expressed as

$$\min_{\delta \in \mathcal{D}} \max_{(g_0, g_1) \in \mathcal{F}} R(\delta, g_0, g_1).$$

Since $R(\delta, g_0, g_1)$ is separately linear with respect to $\delta$, and $(g_0, g_1)$, the min-max problem has a convex-concave structure. For appropriate choices of metrics, $\mathcal{D}$ and $\mathcal{F}$ are compact, so by Von-Neumann’s minimax theorem, there exists a saddle point $(\delta_R, g_0^L, g_1^L)$ satisfying

$$R(\delta_R, g_0, g_1) \leq R(\delta_R, g_0^L, g_1^L) \leq R(\delta, g_0^L, g_1^L). \quad (1)$$
Robust Hypothesis Testing (cont’d)

- Here $\delta_R = \text{robust test}$, and $(g_0^L, g_1^L) = \text{least-favorable densities}$. The second inequality in (1) implies $\delta_R$ is the optimum Bayesian test for the pair $(g_0^L, g_1^L)$, so $\delta_R$ can be expressed as the LR test

$$L_L(y) = \frac{g_1^L(y)}{g_0^L(y)} \begin{cases} \geq & H_1 \\ \leq & H_0 \end{cases} \tau_B.$$

- Since $R(\delta, g_0, g_1)$ is a fixed linear combination of $P_M(\delta, g_1)$ and $P_F(\delta, g_0)$, the first inequality in (1) is equivalent to

$$P_F(\delta_R, g_0) \leq P_F(\delta_R, g_0^L)$$

for all $g_0 \in \mathcal{F}_0$ and $g_1 \in \mathcal{F}_1$. 

\[ P_M(\delta_R, g_1) \leq P_M(\delta_R, g_1^L) \quad (2) \]
Robust Hypothesis Testing (cont’d)

- The **robust NP test** solves

\[
\min_{\delta \in \mathcal{D}_\alpha} \max_{g_1 \in \mathcal{F}_1} P_M(\delta, g_1),
\]

where

\[
\mathcal{D}_\alpha = \{ \delta \in \mathcal{D} : \max_{g_0 \in \mathcal{F}_0} P_F(\delta, g_0) \}
\]

is the set of decision rules of size less than \( \alpha \). Since \( P_F(\delta, g_0) \) is a convex function of \( \delta \) for each \( g_0 \in \mathcal{F}_0 \), so is

\[
\max_{g_0 \in \mathcal{F}_0} P_F(\delta, g_0),
\]

hence \( \mathcal{D}_\alpha \) is convex.

- The cost function \( P_F(\delta, g_1) \) has a convex concave structure, so a saddle point exist, and \( \delta_R \) is the optimal NP test for least favorable observation densities \( (g_0^L, g_1^L) \).
Outline

- Binary Hypothesis Testing
- Robust Hypothesis Testing
- Huber’s Clipped LR Test
- Robustness with a KL Divergence Tolerance
- Simulations
Huber’s Clipped LR Test

- Different choices of neighborhoods $\mathcal{F}_j$ yield different robust tests. Let $G_j(y)$ and $F_j(y)$ denote the cumulative probability distribution functions corresponding to the actual and nominal densities $g_j(y)$ and $f_j(y)$ for $j = 0, 1$. For some numbers $0 \leq \epsilon_0, \epsilon_1, \nu_0, \nu_1 < 1$, Huber considered neighborhoods

$$\mathcal{F}_0 = \{ g_0 : G_0(y) \geq (1 - \epsilon_0)F_0(y) - \nu_0 \text{ for all } y \in \mathbb{R} \}$$
$$\mathcal{F}_1 = \{ g_1 : 1 - G_1(y) \geq (1 - \epsilon_1)(1 - F_1(y)) - \nu_1 \text{ for all } y \in \mathbb{R} \} .$$

- The constraints specifying $\mathcal{F}_j$ are linear in $g_j$, so the neighborhoods are convex.

- Since functions $P_M(\delta_R, g_1)$ and $P_F(\delta_R, g_0)$ are linear in $g_1$ and $g_0$, the maximization (2) for the least-favorable densities is a linear programming problem, so solutions will be located on the boundary of $\mathcal{F}_j$. 
Huber’s Clipped LR Test (cont’d)

Least-favorable densities: There exists $I = [y_L, y_U]$ such that over this interval

$$G^L_0(y) = (1 - \epsilon_0)F_0(y) - \nu_0$$
$$G^L_1(y) = (1 - \epsilon_1)F_1(y) + \epsilon_1 + \nu_1,$$

so the least-favorable densities are on the boundary of sets $\mathcal{F}_0$ and $\mathcal{F}_1$. For $j = 0, 1$, this implies

$$g^L_j(y) = (1 - \epsilon_j)f_j(y)$$

over $I$. Let

$$a(y) = v' f_0(y) + w' f_1(y)$$
$$b(y) = v'' f_0(y) + w'' f_1(y),$$
Huber’s Clipped LR Test (cont’d)

with

\[ v' = \frac{\epsilon_1 + \nu_1}{1 - \epsilon_1}, \quad v'' = \frac{\epsilon_0 + \nu_0}{1 - \epsilon_0} \]
\[ w' = \frac{\nu_0}{1 - \epsilon_0}, \quad w'' = \frac{\nu_1}{1 - \epsilon_1}. \]

Let \( l_L = L(y_L), l_U = L(y_U) \). Then

\[ g_j^L(y) = c_j a(y), \quad y \leq y_L \]
\[ g_j^L(y) = d_j b(y), \quad y \geq y_U, \]

with

\[ \frac{c_1}{c_0} = \frac{1 - \epsilon_1}{1 - \epsilon_0} l_L, \quad \frac{d_1}{d_0} = \frac{1 - \epsilon_1}{1 - \epsilon_0} l_U. \]
Clipping transformation: The least-favorable LR can be expressed as

\[ L_L(y) = \frac{g_1^L(y)}{g_0^L(y)} = \frac{1 - \epsilon_1}{1 - \epsilon_0} C(L(y)) \]

where the clipping nonlinearity \( C(\cdot) \) is shown below:

![Diagram of clipping nonlinearity](image-url)
Huber’s clipped LR test (cont’d)

**Robust test:** The decision rule

\[
L_L(y) \begin{cases} \frac{H_1}{H_0} \geq \tau_B \\
\frac{H_1}{H_0} < \tau_B 
\end{cases}
\]

can be rewritten as

\[
C(L(y)) \begin{cases} \frac{H_1}{H_0} \geq \eta \\
\frac{H_1}{H_0} < \eta \end{cases} = \frac{1 - \epsilon_0}{1 - \epsilon_1} \tau_B.
\]
Huber’s clipped LR test (cont’d)

- For $\nu_0 = \nu_1 = 0$, the LF distributions belong to the contamination class
  \[ \mathcal{N}_j^C = \{ g_j : G_j(y) = (1 - \epsilon_j)F_j(y) + \epsilon_j H(y) \text{ for all } y \} \]
  contained in $\mathcal{F}_j$, where $H(y) = \text{arbitrary probability distribution}$.

- For $\epsilon_0 = \epsilon_1 = 0$, the LF densities belong to the total variation class
  \[ \mathcal{N}_j^{TV} = \{ g_j : |g_j - f_j|_1 \leq 2\nu_j \} \]
  contained in $\mathcal{F}_j$. 
Outline

- Binary Hypothesis Testing
- Robust Hypothesis Testing
- Huber’s Clipped LR Test
  - Robustness with a KL Divergence Tolerance
- Simulations
Robustness with a KL Tolerance

- For $j = 0, 1$ consider neighborhoods

$$\mathcal{F}_j = \{g_j : D(g_j|f_j) \leq \epsilon\}$$

where

$$D(g|f) = \int_{-\infty}^{\infty} \ln(g(y)/f(y))g(y)dy$$

is the Kullback-Leibler divergence or relative entropy of density $g$ with respect to $f$.

- $D(g|f)$ is convex in $g$, so $\mathcal{F}_j$ is convex. $D(g|f)$ is not a true distance, since it is not symmetric ($D(f|g) \neq D(g|f)$) and does not satisfy the triangle inequality. But $D(g|f) \geq 0$ with equality if and only if $g = f$.

- $D(g|f)$ and its dual $D^*(g|f) = D(f|g)$ admit a non-Riemannian differential geometric interpretation in terms of dual connections.
Robustness with a KL Tolerance

Assumptions:

i) The nominal LR $L(y) = f_1(y)/f_0(y)$ is monotone increasing in $y$.

ii) $f_1(y) = f_0(-y)$.

iii) $0 < \epsilon < D(f_{1/2}|f_0)$, where $f_{1/2}(y)$ is the mid-way density on the geodesic

\[
    f_u(y) = \frac{f_0^{1-u}(y)f_1^u(y)}{Z(u)}
\]

linking $f_0$ and $f_1$. Here

\[
    Z(u) = \int_{-\infty}^{\infty} f_1^u(y)f_0^{1-u}(y)dy
\]

= normalization constant.
Robust test and LF densities: For a minimum probability of error criterion \((C_F = C_M = 1)\) and equally likely hypotheses, there exists \(y_U > 0\) such that

\[
\delta_R(y) = \begin{cases} 
1 & y > y_U \\
\frac{1}{2} \left[ 1 + \frac{\ln L(y)}{\ln \ell_U} \right] & -y_U \leq y \leq y_U \\
0 & y < -y_U ,
\end{cases}
\]

\[
g_0^L(y) = \begin{cases} 
\ell_U f_0(y)/Z(y_U) & y > y_U \\
\ell_U^{1/2} f_1^{1/2} f_0^{1/2}(y)/Z(y_U) & -y_U \leq y \leq y_U \\
f_0(y)/Z(y_U) & y < -y_U ,
\end{cases}
\]

\[g_1^L(y) = g_0^L(-y), \text{ with } \ell_U = L(y_U).\]
Nonlinear transformation: The least-favorable LR can be expressed as a nonlinear transformation $L_L = q(L)$ of the nominal LR.

![Graph showing a nonlinear transformation](image)
Robustness with a KL Tolerance (cont’d)

- $g_0^L(\cdot|y_U)$ is parametrized by $y_U$ with $g_0^L = f_0$ for $y_U = 0$ and $\lim g_0^L = f_{1/2}$ as $y_U \to \infty$.

- $y_U$ is selected such that $D(g_0^L(\cdot|y_U)|f_0) = \epsilon$. Relies on showing that

$$D(y_U) = D(g_0^L(\cdot|y_U)|f_0)$$

is a monotone increasing function of $y_U$. 
Outline

- Binary Hypothesis Testing
- Robust Hypothesis Testing
- Huber’s Clipped LR Test
- Robustness with a KL Divergence Tolerance
- Simulations
Simulations

Consider the nominal model

\[ H_0 : Y = -1 + V \quad H_1 : Y = 1 + V , \]

with \( V \sim N(0, \sigma^2) \), so \( f_0 \sim N(-1, \sigma^2) \). \( D(y_U) \) is plotted below for SNR = 0 dB (\( \sigma = 1 \)).
LF densities $g_0^L$ for $\epsilon = 0.1$ and SNR = 0, 10dB.
Simulations (cont’d)

Comparison of worst-case $P[E]$ for test $\delta_R$ with $\epsilon = 0.01, 0.1$ against $P[E]$ for the Bayesian test on nominal model.
References


Thank you!!