

A Bayesian decision theoretic solution to the problem related to microarray data with consideration of intra-correlation among the affected genes

By

Naveen K. Bansal

Marquette University

Milwaukee, Wisconsin(USA)

Email: naveen.bansal@mu.edu

<http://www.mscs.mu.edu/~naveen/>

With

Dr. Klaus Miescke

University of Illinois-Chicago

Illinois (USA)



When the genes are altered under adverse condition, such as cancer, the affected genes show under or over expression in a microarray.

X_i – *Expression Level*

$X_i \sim P(\theta_i, \alpha)$

$H_0^i : \theta_i = 0$ vs $H_{-1}^i : \theta_i < 0$ or $H_1^i : \theta_i > 0$

The objective is to find the genes with under expressions and genes with over expressions.



Probability Model: $P(\theta, \alpha)$

$H_0: \theta = \theta_0$ (say, 0) vs $H_1: \theta < \theta_0$, $H_2: \theta > \theta_0$

Directional Error (Type III error):

Sarkar and Zhou (2008, JSPI)

Finner (1999, AS)

Shaffer (2002, Psychological Methods)

Lehmann (1952, AMS; 1957, AMS)

Main points of these work is that if the objective is to find the true Direction of the alternative after rejecting the null, then a Type III error must be controlled instead of Type I error.

Type III error is defined as $P(\text{false directional error if the null is rejected})$. The traditional method does not control the directional error. For example,

$|t| > t_{\alpha/2}$, and $t > t_{\alpha/2}$, an error occurs if $\theta < 0$.



Bayesian Decision Theoretic Framework

Probability Model: $P(\theta, \alpha)$

$H_0 : \theta = \theta_0$ (say, 0) , $H_1 : \theta < \theta_0$, $H_2 : \theta > \theta_0$

$\pi(\theta) = p_{-1}\pi_{-1}(\theta) + p_0\pi_0(\theta) + p_1\pi_1(\theta)$

where

$\pi_{-1}(\theta) = g_{-1}(\theta)I(\theta < 0)$, $\pi_0(\theta) = I(\theta = 0)$, $g_1(\theta)I(\theta > 0)$

$g_{-1}(\cdot)$ – density with support contained in $(-\infty, 0)$

$g_1(\cdot)$ – density with support contained in $(0, \infty)$

$A = \{-1, 0, 1\}$, $L(\theta, a)$ – loss for taking action $a \in A$.



g_{-1} and g_1 could be truncated densities of a density on θ .

The skewness in the prior is introduced by (p_{-1}, p_0, p_1) .

$p_{-1} < p_1$ reflects that the right tail is more likely than the left tail.

$p_{-1} = 0$ (or $p_1 = 0$) would yield a one - tail test.

$p_{-1} = p_1$ with g_{-1} and g_1 as truncated of a symmetric density would yield a two - tail test.

p_{-1} and p_1 can be assigned based on what tail is more important.



To introduce the correlation among under expressed and over expressed genes, we can modify the priors with $g_{-1}(\theta | \beta)$ and $g_{+}(\theta | \gamma)$ depending on the parameters β and γ .

A prior on β and γ introduces correlation among underexpressed genes and over expressed genes.



The Bayes risk for a decision rule δ is given by

$$r^\delta(\pi) = p_{-1}r_{-1}^\delta(\pi_{-1}) + p_0r_0^\delta(0) + p_1r_1^\delta(\pi_1)$$

where

$$r_{-1}^\delta(\pi_{-1}) = \int_{\theta < 0} R(\theta, \delta) \pi_{-1}(\theta) d\theta$$

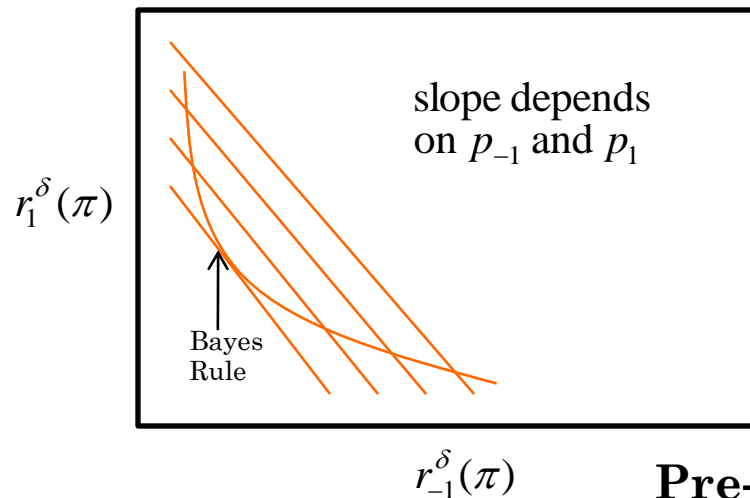
$$r_1^\delta(\pi_1) = \int_{\theta > 0} R(\theta, \delta) \pi_1(\theta) d\theta$$

$$r_0^\delta(0) = R(0, \delta)$$

For a fixed prior π , decision rules can be compared by comparing the space

$$S(\pi) = \{(r_{-1}^\delta(\pi_{-1}), r_0^\delta(0), r_1^\delta(\pi_1)) : \delta \in D^*\}$$

consider the class of all rules δ for which $R(0, \delta)$ are the same



Consider the following two priors

$$\pi(\theta) = p_{-1}\pi_{-1}(\theta) + p_0\pi_0(\theta) + p_1\pi_1(\theta),$$

$$\pi'(\theta) = p'_{-1}\pi_{-1}(\theta) + p_0\pi_0(\theta) + p'_1\pi_2(\theta)$$

$$p_1 > p'_1 \text{ (i.e. } H_1 \text{ is more likely under } \pi \text{ in comparison to } \pi')$$

Theorem 1: If $p_1 > p'_1$, and if δ_B and δ'_B are the Bayes rules under π and π' respectively, then

$$r_1^{\delta_B}(\pi) \leq r_1^{\delta'_B}(\pi')$$

and

$$r_{-1}^{\delta_B}(\pi) \geq r_{-1}^{\delta'_B}(\pi')$$

Remark : This theorem implies that if a priori it is known that H_1 is more likely than H'_{-1} ($p_1 > p'_{-1}$), then the average risk of the Bayes rule in the positive direction will be smaller than average risk in the negative direction.



An Application:

An experiment was conducted to see the effect of a sequence knockdown from non-coding genes. Hypothesis was that this knockdown will cause the overexpression of the mRNAs of the coding genes. The implication of this is that this will show how the non-coding genes interact with coding genes. In other words, non-coding genes also play a part in protein synthesis.

Here it can be assumed that that the effect of knockdown on the coding genes (if there is any) would be mostly overexpression of mRNAs than underexpression of mRNAs. **The objective would be to detect as many of overexpressed genes as possible.**



Relationship with the False Discovery Rates

Consider the "0-1" loss

$$L(\theta, -1) = \begin{cases} 0 & \theta < 0 \\ 1 & \theta \geq 0 \end{cases} \quad L(\theta, 0) = \begin{cases} 0 & \theta = 0 \\ 1 & \theta \neq 0 \end{cases} \quad L(\theta, 1) = \begin{cases} 0 & \theta > 0 \\ 1 & \theta \leq 0 \end{cases}$$

$$R(\theta, \delta) = \begin{cases} P_\theta(\delta \geq 0) & \theta < 0 \\ P_\theta(\delta = 0) & \theta = 0 \\ P_\theta(\delta \leq 0) & \theta > 0 \end{cases}$$

$$r_{-1}^{\delta_B}(\pi) = \int_{\theta < 0} P_\theta(\delta \geq 0) \pi_{-1}(\theta) d\theta = E[P(\theta < 0 | \delta \geq 0)]$$

which is expected false non - negative discovery rate in the Bayesian framework.

Similarly $r_1^{\delta_B}(\pi)$ can be interpreted as expected false non - positive discovery rate, and

$r_0^{\delta_B}(\pi)$ as expected false non - discovery rate. $r^{\delta_B}(\pi)$ is the weighted average of these false discoveries.



$\{\mathbf{X}_i, i=1,2,\dots,N\}$ independent data with

$$\mathbf{X}_i \sim f(\mathbf{x}_i; \theta_i, \alpha), i=1,2,\dots,N$$

Consider the following multiple hypotheses problem :

$$H_0^i: \theta_i = 0, H_{-1}^i: \theta_i < 0, H_1^i: \theta_i > 0$$

$$\theta_i \sim p_{-1}\pi_{-1}(\theta) + p_0\pi_0(\theta) + p_1\pi_1(\theta), i=1,2,\dots,N \text{ independent}$$

$$\delta_B^i \in \{-1,0,1\} - \text{non-randomized Bayes rule}$$



Table 1
Possible outcomes from m hypothesis tests

	Accept H_0	Accept H_{-1}	Accept H_1	Total
H_0 is true	U	V_1	V_2	m_0
H_{-1} is true	T_1	S_1	W_2	m_1
H_1 is true	T_2	W_1	S_2	m_2
Total	R_0	R_1	R_2	m

$$FDR_{-1} = E\left[\frac{V_1 + W_1}{R_1}\right] \quad FDR_{+1} = E\left[\frac{V_2 + W_2}{R_2}\right]$$

$$pFDR_{-1} = E\left[\frac{V_1 + W_1}{R_1} \mid R_1 > 0\right] \quad pFDR_{+1} = E\left[\frac{V_2 + W_2}{R_2} \mid R_2 > 0\right] \quad (\text{Storey, 2003 AS})$$



Theorem 2: Suppose that m hypothesis tests are based on test statistics T_1, T_2, \dots, T_m . Let Γ_{-1} and Γ_1 be the regions of accepting H_{-1}^i and H_1^i , respectively, for each $i=1,2,\dots,m$.

Assume that $T_i | H^i \stackrel{iid}{\sim} I(H_{-1}^i)F_{-1} + I(H_0^i)F_0 + I(H_1^i)F_1$, where F_{-1}, F_0, F_1 are the marginal distributions of T_i wrt π_{-1}, π_0 and π_1 respectively. If $(H_{-1}^i, H_0^i, H_1^i), i=1,2,\dots,m$ are independent and identical trinomial trials, then

$$pFDR_{-1}(\Gamma) = P(H_0 | T \in \Gamma_{-1}) + P(H_1 | T \in \Gamma_{-1})$$

and

$$pFDR_{+1}(\Gamma) = P(H_0 | T \in \Gamma_1) + P(H_{-1} | T \in \Gamma_1)$$



This theorem and our previous discussion on the Bayes decision rule implies that under the independent hypotheses testing setting, the Bayes rule obtained from a single test problem

$$H_0:\theta=0 \text{ vs. } H_{-1}:\theta<0, H_1:\theta>0,$$

Which minimizes the average expected false discoveries, can be used under multiple hypotheses problem. The false discovery rates of this procedure can be obtained by computing the posterior probabilities.

$$pFDR_{-1}(\delta_B) = P(H_0|\delta_B = -1) + P(H_1|\delta_B = -1)$$

$$pFDR_{+1}(\delta_B) = P(H_0|\delta_B = 1) + P(H_{-1}|\delta_B = 1)$$

The prior probabilities p_{-1} and p_1 will determine whether the positive hypothesis H_1 or the negative hypothesis H_{-1} will be more frequently correctly detected.



Posterior False Discovery Rate

$$PFDR_{-1} = E \left[\frac{\sum_{i=1}^m I(\theta_i \geq 0) I(d_i = -1)}{\sum_{i=1}^m I(d_i = -1) \vee 1} \mid \mathbf{X} \right]$$
$$= \frac{\sum_{i=1}^m P(\theta_i \geq 0 \mid \mathbf{X}) I(d_i = -1)}{\sum_{i=1}^m I(d_i = -1) \vee 1}$$

Thus if $P(\theta_i \geq 0 \mid \mathbf{X}) \leq \alpha_{-1}$, then $E[PFDR_{-1}] \leq \alpha_{-1}$.

Similarly, if $P(\theta_i \leq 0 \mid \mathbf{X}) \leq \alpha_1$, then $E[PFDR_1] \leq \alpha_1$.



Let

$$q_{\bar{i}}(\mathbf{X}) = P(\theta_i \geq 0 | \mathbf{X})$$

$$q_i^+(\mathbf{X}) = P(\theta_i \leq 0 | \mathbf{X})$$

$$q_{1,m}^- \leq q_{2,m}^- \leq \dots \leq q_{m,m}^-$$

$$q_{1,m}^+ \leq q_{2,m}^+ \leq \dots \leq q_{m,m}^+$$

$$Q_{\bar{j}}(X) = \frac{1}{j} \sum_{i=1}^j q_{i,1}^- \quad Q_j^+ = \frac{1}{j} \sum_{i=1}^j q_{i,1}^+$$



Theorem 3: Let

$$k_-(X) = \begin{cases} \max\{j: Q_j^-(X) \leq \alpha_{-1}\}, & \text{if maximum exists} \\ 0 & \text{otherwise} \end{cases}$$

$$k_+(X) = \begin{cases} \max\{j: Q_j^+(X) \leq \alpha_1\}, & \text{if maximum exists} \\ 0 & \text{otherwise} \end{cases}$$

Select all H_{-1}^j corresponding to $q_{1,1}^- \leq \dots \leq q_{k_-,m}^-$.

Select all H_1^j corresponding to $q_{1,1}^+ \leq \dots \leq q_{k_+,m}^+$.

Then $E[PFDR_-] \leq \alpha_{-1}$ and $E[PFDR_+] \leq \alpha_2$.



Computation of Bayes Rule:

$$\delta^B(\pi) = \{i \in (-1, 0, 1) : E[L(\theta, i) | \mathbf{X}] = \min_{j \in -1, 0, 1} E[L(\theta, j) | \mathbf{X}]\}$$

$$\pi(\theta | \mathbf{X}) = \pi(H_{-1} | \mathbf{X})\pi(\theta | H_{-1}, \mathbf{X}) + \pi(H_0 | \mathbf{X})I(\theta = 0) + \pi(H_1 | \mathbf{X})\pi(\theta | H_1, \mathbf{X})$$

Here $\pi(H_{-1} | \mathbf{X})$, $\pi(H_0 | \mathbf{X})$, and $\pi(H_1 | \mathbf{X})$ are the posterior probabilities of H_{-1} , H_0 , and H_1 respectively; and $\pi(\theta | H_{-1}, \mathbf{X})$ and $\pi(\theta | H_1, \mathbf{X})$ are the posterior densities of θ under $\pi_{-1}(\theta)$ and $\pi_1(\theta)$ respectively.

Computation parts involve computing $\pi(H_{-1} | \mathbf{X})$, $\pi(H_0 | \mathbf{X})$, and $\pi(H_1 | \mathbf{X})$ and computing $E[L(\theta, -1) | H_1, \mathbf{X}]$ and $E[L(\theta, 1) | H_{-1}, \mathbf{X}]$.



Choice of the loss function:

$$L(\theta,0) = \begin{cases} 0 & \theta=0 \\ q(\theta) & \theta \neq 0 \end{cases}, \quad L(\theta,-1) = \begin{cases} 0 & \theta < 0 \\ q_{-1}(\theta) & \theta \geq 0 \end{cases}, \quad L(\theta,+1) = \begin{cases} 0 & \theta > 0 \\ q_1(\theta) & \theta \leq 0 \end{cases}$$

For the "0-1" loss, $q_0(\theta) = q_{-1}(\theta) = q_1(\theta) = 1$.

A good choice would be

$$q_0(\theta) = l_0(1+q(\theta)), \quad q_{-1}(\theta) = l_{-1}(1+q(\theta)), \quad q_1(\theta) = l_1(1+q(\theta))$$

where l_0 , l_{-1} and l_1 are some non-negative constants.

Here, $q(\theta)$ reflect the magnitude of departure from the null. One good choice of $q(\theta)$ is

$$q(\theta) = E_{\theta}[\log(f(X,0)/f(X,\theta))]$$



If $q(\theta) = q_{-1}(\theta) = q_1(\theta) = 1$, then we have a "0-1" loss.

Theorem 3: Under the "0-1" loss, the Bayes rule is given by

$$\delta^B = \{i: p_i f(\mathbf{X}|H_i) = \max(p_j f(\mathbf{X}|H_j), j = -1, 0, 1)\}$$

Here $f(\mathbf{X}|H_j)$ is the marginal density \mathbf{X} under H_j .

In other words, reject H_0 if

$$\frac{f(\mathbf{X}|H_0)}{f(\mathbf{X}|H_{-1})} < \frac{p_{-1}}{p_0} \quad \text{and} \quad \frac{f(\mathbf{X}|H_0)}{f(\mathbf{X}|H_1)} < \frac{p_1}{p_0}.$$

Otherwise, select $H_{-1}(H_1)$ if

$$p_{-1} f(\mathbf{X}|H_{-1}) > (<) p_1 f(\mathbf{X}|H_1)$$



Example:

Probability Model: $N(\theta, 1)$

$$H_0: \theta = 0 \text{ vs. } H_{-1}: \theta < 0, H_1: \theta > 0$$

Prior: $\pi(\theta) = p_{-1}\pi_{-1}(\theta) + p_0I(\theta=0) + p_1\pi_1(\theta)$

π_{-1} : left-half $N(0, 1)$ distribution

π_1 : right-half $N(0, 1)$ distribution



Theorem 4: Under the "0-1" loss, the Bayes rule reject H_0 if

$$T_{-1}(\bar{x}) > p_0/p_{-1} \quad \text{and} \quad T_1(\bar{x}) > p_0/p_1,$$

otherwise select $H_{-1}(H_1)$ if $p_{-1}T_{-1}(\bar{x}) >(<) p_1T_1(\bar{x})$, where

$$T_{-1}(x) = \frac{2}{\sqrt{n+1}} \exp\left(\frac{n^2x^2}{2(n+1)}\right) \Phi\left(-\frac{nx}{\sqrt{n+1}}\right)$$

and

$$T_1(x) = \frac{2}{\sqrt{n+1}} \exp\left(\frac{n^2x^2}{2(n+1)}\right) \Phi\left(\frac{nx}{\sqrt{n+1}}\right)$$

Note that T_{-1} and T_1 are monotonically decreasing and increasing functions, respectively. Thus the rejection region is equivalent to

$$\bar{x} < T_{-1}^{-1}(p_0/p_{-1}) \quad \text{and} \quad \bar{x} > T_1^{-1}(p_0/p_1)$$

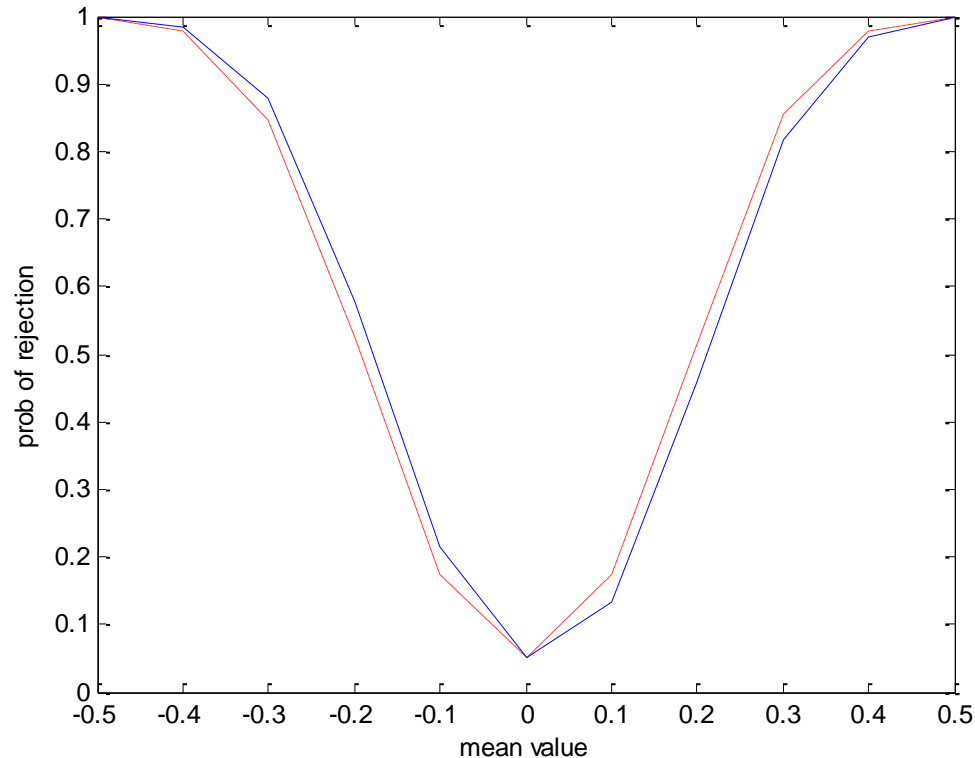


Power Comparison:

When $p_{-1} = p_1$, the rejection region is UMP unbiased.

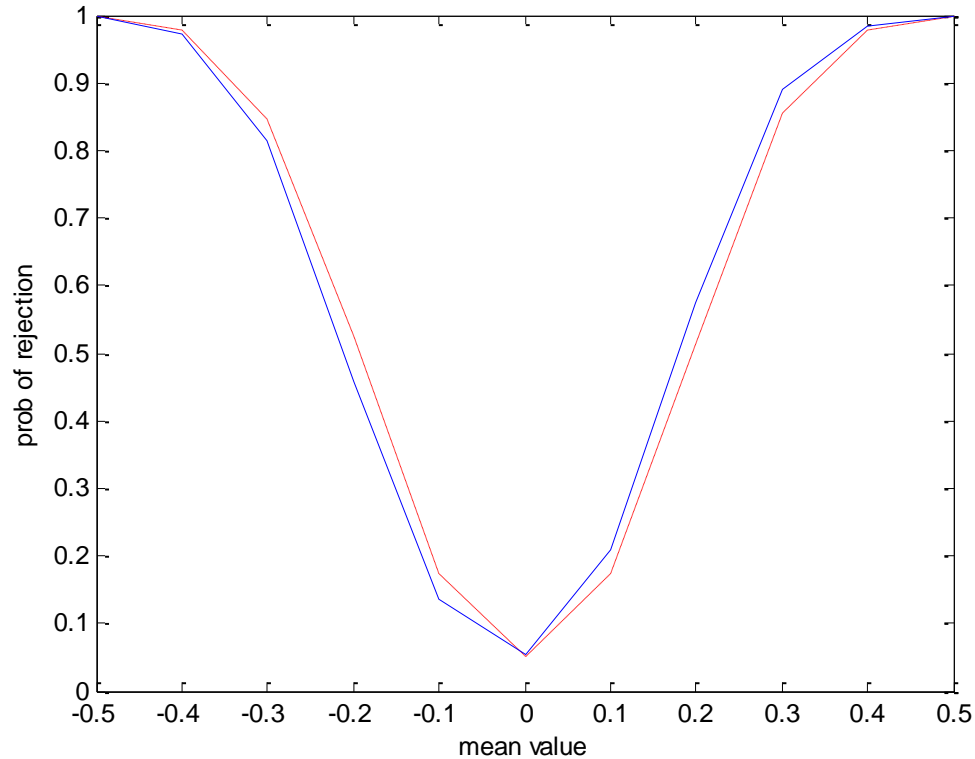
Bayes test with $p_{-1} = 0.39$, $p_0 = 0.39$, $p_1 = 0.22$: blue line)

UMP unbiased test: red dotted line



Bayes test with $p_{-1}=0.22$, $p_0=0.39$, $p_1=0.39$: blue line)

UMP unbiased test: red dotted line



What if p_{-1}, p_0 , and p_1 are unknown?

One can consider a Dirichlet prior. It is easy to see that the only difference would be to replace p_{-1}, p_0 , and p_1 by $E(p_{-1}|X)$, $E(p_0|X)$, and $E(p_1|X)$ respectively.

For the normal problem, we get the following result.

Theorem 5: Under the "0-1" loss, the Bayes rule reject H_0 if

$$T_{-1}(\bar{x}) > \frac{E(p_0|\mathbf{X})}{E(p_{-1}|\mathbf{X})} \quad \text{and} \quad T_1(\bar{x}) > \frac{E(p_0|\mathbf{X})}{E(p_1|\mathbf{X})},$$

otherwise select $H_{-1}(H_1)$ if $E(p_{-1}|\mathbf{X})T_{-1}(\bar{x}) >(<) E(p_1|\mathbf{X})T_1(\bar{x})$, where



Another Approach:

Consider the prior

$$\pi(\theta) = p_0 I(\theta=0) + (1-p_0)\pi'(\theta),$$

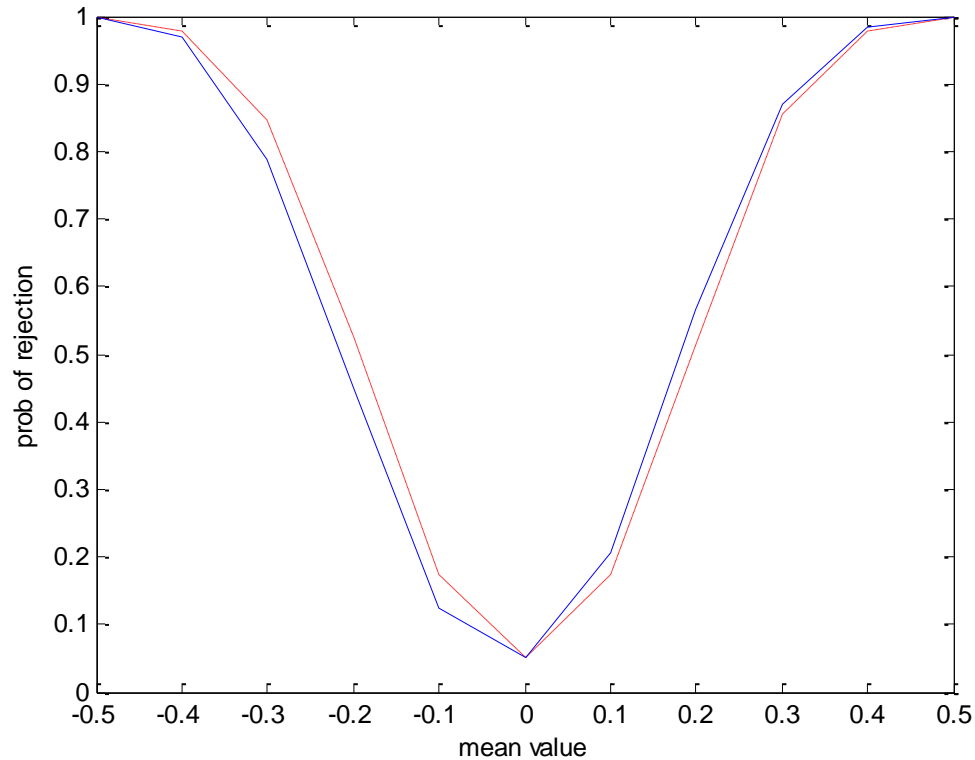
where $\pi'(\theta)$ is a skewed prior; for example, skew-normal

$$\pi'(\theta) = \frac{2}{\tau} \phi\left(\frac{\theta}{\tau}\right) \Phi\left(\lambda \frac{\theta}{\tau}\right) I(\theta \neq 0).$$

Here λ reflects the skewness in θ values, and τ the dispersion in θ values.



Power Comparison of a skew-normal Bayes with the UMP test



?





Pre-ICM Int. Convention, 2008