Estimating the Proportion of True Null Hypotheses in Identifiable Nonparametric Models

June 19, 2009

Statistical Genomics Programs
IMS at National University of Singapore

Based on joint work with X. Wu

Multiple-hypothesis testing problem

Consider simultaneous testing of m null hypotheses where we test the simple H_{0i} versus H_{1i}, $i = 1, \cdots, m$, on the basis of the p-values:

$$p_1, p_2, \cdots, p_m$$

Multiple-hypothesis testing problem

Consider simultaneous testing of m null hypotheses where we test the simple H_{0i} versus H_{1i}, $i = 1, \cdots, m$, on the basis of the p-values:

$$p_1, p_2, \cdots, p_m$$

Basic assumptions:

- All p_i's are independent.
- All p_i's are continuous-type.
- The same test statistic is used for all null hypotheses.

Multiple-hypothesis testing problem

Consider simultaneous testing of m null hypotheses where we test the simple H_{0i} versus H_{1i}, $i = 1, \cdots, m$, on the basis of the p-values:

$$p_1, p_2, \cdots, p_m$$

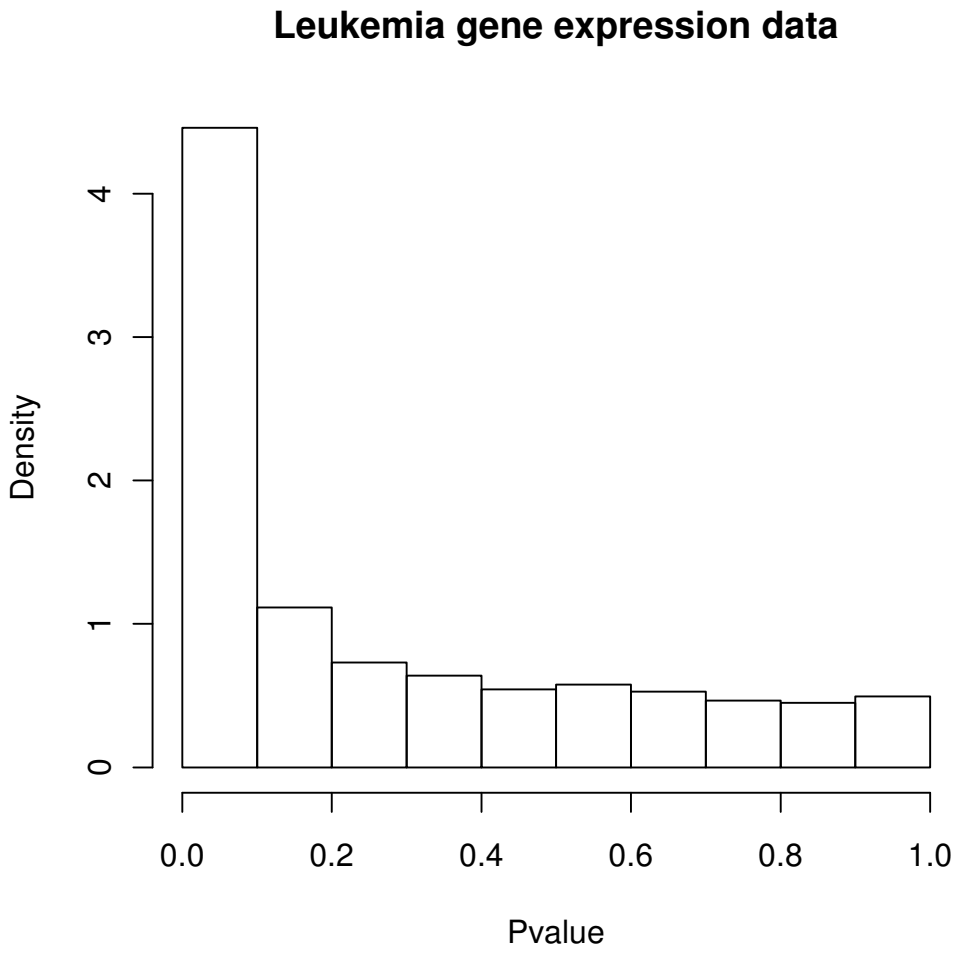
Basic assumptions:

- All p_i's are independent.
- All p_i's are continuous-type.
- The same test statistic is used for all null hypotheses.

So under H_{0i}, p_i follows the uniform distribution on (0,1), for $i = 1, \cdots, m$. In genome studies with microarray data, m is usually the number of genes and typically large.

The Leukemia gene expression data

In Golube et al (1999), the leukemia gene microarray data are produced from 38 bone marrow samples [27 acute lymphoblastic (ALL) and 11 acute myeloid leukemia (AML)]. Each patient's marrow cells are analyzed on a separate microarray plate that measures expression levels for 6817 genes. After pre-processing, only 3051 genes are accurately readable so that the microarray data are altogether given in a 3051×38 matrix in expression levels.


Leukemia gene expression data
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Pvalue

The Leukemia gene example

For each of the 3051 genes, one wishes to know if the ALL group is significantly different from the AML group. So for each row of the microarray data matrix, a p-value is computed based on the two-sample t-test statistic with degree 36 of freedom (two-sided):

- $m = 3051$
- $H_{0i} : \mu_{ALL} = \mu_{AML}$
- $p_i = P(|T_{36}| \geq |t_i|)$

Overall false rejection rate

Question: Suppose that H_{0i} is rejected if and only if $p_i \leq \alpha$. How to measure the overall false rejection rate and how to choose α so that the overall false rejection rate is under controlled?

Overall false rejection rate

Question: Suppose that H_{0i} is rejected if and only if $p_i \leq \alpha$. How to measure the overall false rejection rate and how to choose α so that the overall false rejection rate is under controlled?

Note that in the traditional testing theory, if one wishes to have the overall false rejection rate ≤ 0.05, then α has to be at most

$$1 - (0.95)^{1/3051} = 0.0000168$$

Overall false rejection rate

Question: Suppose that H_{0i} is rejected if and only if $p_i \leq \alpha$. How to measure the overall false rejection rate and how to choose α so that the overall false rejection rate is under controlled?

Note that in the traditional testing theory, if one wishes to have the overall false rejection rate ≤ 0.05, then α has to be at most

$$1 - (0.95)^{1/3051} = 0.0000168$$

In a seminal paper, Benjamini and Hochberg (1995) proposed a new approach to control the mean false rejection rate, i.e., the false discovery rate (FDR) approach.

FDR and pFDR

Given a decision rule, let R be the number of rejections and V be the number of false rejections. Define

- $FDR = E\{(V/R)I(R > 0)\}$
- $pFDR = E\{(V/R)|R > 0\}$

Note that V is unobservable. Therefore FDR or pFDR needs to be estimated in its application. This is a difficult task.

FDR and pFDR

Given a decision rule, let R be the number of rejections and V be the number of false rejections. Define

- $FDR = E\{(V/R)I(R > 0)\}$
- $pFDR = E\{(V/R)|R > 0\}$

Note that V is unobservable. Therefore FDR or pFDR needs to be estimated in its application. This is a difficult task.

Benjamini and Hochberg (1995) shows that FDR of Simes' rule is not greater than $\pi\alpha$, where π is the proportion of the true null-hypotheses among m null-hypotheses.

A Bayes approach

Define iid Bernoulli random variables $H_1, \cdots, H_m$, where $H_i = 0$ iff H_{0i} is true, so that $P(H_1 = 0) = \pi$. Let p_i have the pdf $h(x)$ given $H_i = 1$. Then the marginal distribution of p_i is as follows:

$$f(x) = \pi + (1 - \pi)h(x).$$

Storey (2002) shows that if the rejection region for testing H_{0i} is $\{p_i \leq \alpha\}$, then

$$pFDR = \frac{\pi\alpha}{P(p \leq \alpha)}.$$

Note that $P(p \leq \alpha)$ can be estimated by the empirical distribution of $p_1, \cdots, p_m$. Estimation of pFDR is then accomplished by estimating π.

The λ-estimator of π

Schweder and Spjotvoll (1982) and Storey (2002) suggest the λ-estimator of π as follows. For any $\lambda \in (0, 1)$,

$$P(p > \lambda) = \pi(1 - \lambda) + (1 - \pi) \int_{\lambda}^{1} h(x)dx.$$

For λ not too small, the integral $\int_{\lambda}^{1} h(x)dx$ should be small. Consequently, π can be estimated, albeit conservative, by

$$\hat{\pi}(\lambda) = \frac{\#\{p_i > \lambda\}}{m(1 - \lambda)} = \frac{1 - \hat{F}(\lambda)}{1 - \lambda}.$$

The λ-estimator of π

Schweder and Spjotvoll (1982) and Storey (2002) suggest the λ-estimator of π as follows. For any $\lambda \in (0, 1)$,

$$P(p > \lambda) = \pi(1 - \lambda) + (1 - \pi) \int_{\lambda}^{1} h(x)dx.$$

For λ not too small, the integral $\int_{\lambda}^{1} h(x)dx$ should be small. Consequently, π can be estimated, albeit conservative, by

$$\hat{\pi}(\lambda) = \frac{\#\{p_i > \lambda\}}{m(1 - \lambda)} = \frac{1 - \hat{F}(\lambda)}{1 - \lambda}.$$

Obviously, the choice of λ is crucial. Storey (2002) suggests a bootstrap method to choose λ to minimize MSE of $\hat{\pi}(\lambda)$. Conservativeness stays, though, and is hard to clean up.

Decreasing $h(x)$

Assume $h(x)$ is decreasing with $h(1) = 0$ so that $f(x)$ is decreasing with $f(1) = \pi$. Thus

$$1 = F(1) = F(\lambda) + (1 - \lambda) f(\lambda) + o(\lambda).$$

$$\pi(\lambda) = \frac{1 - F(\lambda)}{1 - \lambda} \approx f(\lambda).$$

Decreasing $h(x)$

Assume $h(x)$ is decreasing with $h(1) = 0$ so that $f(x)$ is decreasing with $f(1) = \pi$. Thus

$$1 = F(1) = F(\lambda) + (1 - \lambda) f(\lambda) + o(\lambda).$$

$$\pi(\lambda) = \frac{1 - F(\lambda)}{1 - \lambda} \approx f(\lambda).$$

Let $\hat{f}$ is the isotonic nonparametric estimate for f. Define

$$\hat{\pi} = \min_{\lambda < p_{(m)}} \left\{ \frac{1 - \hat{F}(\lambda)}{1 - \lambda} \right\}.$$

Convex decreasing model

To improve the estimate, Langaas and Lindqvist (2005) suggests to use a convex decreasing estimate $\hat{f}(x)$. If $f(x)$ is convex and decreasing, there is a mixing distribution μ such that

$$f(x) = \int_0^1 f_\theta(x)\mu(d\theta),$$

where $f_\theta(x) = 2(\theta - x)I(\theta > x)/\theta^2$. Let $\hat{f}(x)$ be the isotonic estimate subject to convex decreasing $f(x)$. Then

$$\hat{\pi} = \hat{f}(1).$$

Estimating π via Bernstein polynomials

Guan, Wu and Zhao (2008) proposes to estimate π based on Bernstein polynomials as follows: Let $B_{j,k}(x) = C_j^k x^j (1-x)^{k-j}$. Define

$$\hat{F}_k(x) = \sum_{j=0}^{k} \hat{F}(j/k) B_{j,k}(x)$$

$$\hat{f}_k(x) = \sum_{j=0}^{k-1} \hat{f}(j/(k-1)) B_{j,k-1}(x).$$

Propose

$$\hat{\pi} = \frac{1}{r} \sum_{i=1}^{r} \hat{f}_k(1 - i/k),$$

where r and k are chosen to minimize an upper bound of the MSE.

The mixture model structure

All the approaches above do not get use of the finite mixture model structure except in their last steps to construct the estimate of π. They can be summarized as follows:

The mixture model structure

All the approaches above do not get use of the finite mixture model structure except in their last steps to construct the estimate of π. They can be summarized as follows:

- Estimate $\hat{f}(x)$ of $f(x)$.
- Then solve π to give an estimate of π from $\hat{f}(x)$, attributing to the expression $f(x) = \pi + (1 - \pi)h(x)$.

The mixture model structure

All the approaches above do not get use of the finite mixture model structure except in their last steps to construct the estimate of π. They can be summarized as follows:

- Estimate $\hat{f}(x)$ of $f(x)$.
- Then solve π to give an estimate of π from $\hat{f}(x)$, attributing to the expression $f(x) = \pi + (1 - \pi)h(x)$.

The structure of finite mixture model is too important to ignore in the present problem, since π is identified only in the upper end of support interval of $f(x)$. One can't expect an estimate $\hat{f}(x)$ of $f(x)$ to perform well when x is nearing the boundaries of support set. Simply it is unfair to expect and demand so.

The latent model

Let random variables p and z be defined as follows:

- $P(z = 0) = \pi$
- $(p|z = 0)$ follows uniform distribution.
- $(p|z = 1)$ follows $h(x)$.

Then (p, z) has the pdf: $\pi^{1-z}\{(1-\pi)h(x)\}^z$, and the marginal distribution of p is exactly

$$f(x) = \pi + (1-\pi)h(x).$$

Note that in the Bayes approach, z is the Bernoulli variable to indicate the status of null hypothesis.

Log-likelihood of complete data

Given a sample $(p_i, z_i), i = 1, \cdots, m$ from (p, z), the log-likelihood is

$$l(\pi, h) = \sum_{i=1}^{m} \{(1 - z_i) \log \pi + z_i \log(1 - \pi) + z_i \log h(p_i)\}.$$

Log-likelihood of complete data

Given a sample $(p_i, z_i), i = 1, \cdots, m$ from (p, z), the log-likelihood is

$$l(\pi, h) = \sum_{i=1}^{m} \{(1 - z_i) \log \pi + z_i \log(1 - \pi) + z_i \log h(p_i)\}.$$

In the present problem, $p_1, \cdots, p_m$ (p-values) are observed, but $z_1, \cdots, z_m$ are unobservable. This defines a typical missing-value problem with missing values $z_1, \cdots, z_m$.

EM algorithm

In the EM algorithm, let π_0 and $f_0(x) = \pi_0 + (1-\pi_0)h_0(x)$ be the current approximation to π and $f(x)$. The next step is to maximize

$$E_0\{l(\pi, h)|p_1, \cdots, p_m\} = m\log(\pi) + \sum_{i=1}^{m}\left\{\hat{z}_i[\log\frac{1-\pi}{\pi} + \log h(p_i)]\right\},$$

where

$$\hat{z}_i = \frac{\pi_0}{f_0(p_i)}, \quad i = 1, \cdots, m.$$

Thus, $\hat{\pi} = \frac{1}{m}\sum \hat{z}_i$ and $\hat{h}$ maximizes

$$Q(h) = \sum_{i=1}^{m} \hat{z}_i \log\{h(p_i)\}.$$

Weighted log-likelihood

The object function Q is a weighted log-likelihood of h. In general, let $0 \leq w_i = w_i(\pi_0, f_0) \leq 1$ be a proposed weighting system and define

$$Q(h) = \sum_{i=1}^{m} w_i \log\{h(p_i)\}.$$

Weighted log-likelihood

The object function Q is a weighted log-likelihood of h. In general, let $0 \leq w_i = w_i(\pi_0, f_0) \leq 1$ be a proposed weighting system and define

$$Q(h) = \sum_{i=1}^{m} w_i \log\{h(p_i)\}.$$

With observation p_i, the weight w_i tells us how much we can trust upon $h(p_i)$ as a likelihood. If $z_i = 1$, i.e., p_i is actually sampled from h, then the ideal weight w_i should be $w_i = 100\%$. Thus, when $w_i = \hat{z}_i$ as an estimate of z_i, we expect it to do a reasonable job. The membership estimate $w_i = I(\hat{z}_i > 1/2)$ in the finite mixture model may be another good choice.

Constraints

The crucial issue in optimization of $Q(h)$ is then the constraint set-up on the nonparametric h.

- $h(x) = P(p = x|z = 1)$.
- Only need to consider discrete h: $\pi_i = h(p_i)$.
- $0 \leq \pi_i, \sum \pi_i = 1$.
- Common and reasonable restriction: h is decreasing. When p_i is sorted, require $\pi_1 \geq \cdots \geq \pi_m$.

Constraints

The crucial issue in optimization of $Q(h)$ is then the constraint set-up on the nonparametric h.

- $h(x) = P(p = x|z = 1)$.
- Only need to consider discrete h: $\pi_i = h(p_i)$.
- $0 \leq \pi_i, \sum \pi_i = 1$.
- Common and reasonable restriction: h is decreasing. When p_i is sorted, require $\pi_1 \geq \cdots \geq \pi_m$.

These constraints are not enough to identify the parameter π of interest.

Identifiability of π

Central to any approach, nonparametric or parametric, is the identifiability of π.

Lemma 1 *Given any family of h, the model $f(x) = \pi + (1-\pi)h(x)$ is identifiable in π if and only if any h in the family is not a strict mixture of $U(0,1)$ and another distribution in the family.*

Corollary 1 *Let $\mathcal{H}$ be the class of all decreasing pdf over $[0,1]$. Then the model $f(x) = \pi + (1-\pi)h(x)$ is identifiable in π if and only if $h(1) = 0$ for any $h \in \mathcal{H}$.*

Question: How to impose $h(1) = 0$ in maximizing $Q(h)$?

Finite Beta mixture models

Parker and Rothenberg (1998) points out that any distribution over (0,1) can be modeled as a beta mixture distribution. So Allison *et al* (2002) suggests to consider a finite parametric beta mixture model: Let

$$f(x) = \pi b(x|1,1) + \sum_{i=1}^{k} \pi_i b(x|a_i, b_i),$$

where $b(x|a,b)$ is the pdf of $Beta(a,b)$ distribution. Unknown parameters are: $\pi, \pi_i, a_i, b_i, i = 1, \cdots, k$ subject to $\pi + \sum \pi_i = 1$. Note $b(x|1,1) = 1$ is the uniform on (0,1).

Finite Beta mixture models

Parker and Rothenberg (1998) points out that any distribution over (0,1) can be modeled as a beta mixture distribution. So Allison *et al* (2002) suggests to consider a finite parametric beta mixture model: Let

$$f(x) = \pi b(x|1,1) + \sum_{i=1}^{k} \pi_i b(x|a_i, b_i),$$

where $b(x|a,b)$ is the pdf of $Beta(a,b)$ distribution. Unknown parameters are: $\pi, \pi_i, a_i, b_i, i = 1, \cdots, k$ subject to $\pi + \sum \pi_i = 1$. Note $b(x|1,1) = 1$ is the uniform on (0,1).

Wu, Guang and Zhao (2006) suggests to use a finite normal mixture model in the scale of original test statistic rather than p-value.

Nonparametric Beta mixture

Tang, Ghosal and Roy (2007) considers a nonparametric Bayes model: Given a distribution over $0 < a \leq 1, b \geq 1$, $p_1, \cdots, p_m$ is a random sample from

$$f(x) = \pi + (1 - \pi) \int b(x|a, b) dG(a, b).$$

Nonparametric Beta mixture

Tang, Ghosal and Roy (2007) considers a nonparametric Bayes model: Given a distribution over $0 < a \leq 1, b \geq 1$, $p_1, \cdots, p_m$ is a random sample from

$$f(x) = \pi + (1 - \pi) \int b(x|a, b) dG(a, b).$$

A Dirichlet process prior on G is suggested. Given p_i's, MCMC algorithm is used to sample from the posterior distribution. Let π^j be the value on π in the j-th posterior sample. The approximating Bayes estimate of π is given by

$$\hat{\pi} = \frac{1}{M} \sum_{j=1}^{M} \pi^j.$$

Identifiable nonparametric model

Assume the random sample $p_1, \cdots, p_m$ is from:

$$f(x) = \pi + (1-\pi) \int_0^1 \int_1^\infty b(x|a,b) dG(a,b).$$

That is, consider

$$\mathcal{H} = \left\{ h(x) = \int b(x|a,b) dG(a,b) : G\{(a,1) : 0 < a \leq 1\} < 1 \right\}.$$

The model is identifiable in π, and $h(1) = 0$ for any $h \in \mathcal{H}$.

Basic properties of MLE

Lemma 2 *The log-likelihood $l(\pi, G)$ is strictly concave downward.*

The concavity nature ensures convergence of EM algorithm.

Basic properties of MLE

Lemma 3 *The log-likelihood $l(\pi, G)$ is strictly concave downward.*

The concavity nature ensures convergence of EM algorithm. Furthermore,

- $f(x)$ is decreasing in x with $f(1) = \pi$.
- The MLE is consistent in proper metric of (π, h).
- The efficiency of the MLE $\hat{\pi}$ can be studied in the semi-parametric model.

Why decreasing?

Consideration of decreasing $h(x)$ is appropriate.

Why decreasing?

Consideration of decreasing $h(x)$ is appropriate.

Wu, Guan and Zhao (2006) shows that if the rejection region of one-sided test is constructed by the likelihood ratio, then the p-value has a nonincreasing pdf.

Why decreasing?

Consideration of decreasing $h(x)$ is appropriate.

Wu, Guan and Zhao (2006) shows that if the rejection region of one-sided test is constructed by the likelihood ratio, then the p-value has a nonincreasing pdf.

Ghosal, Roy and Tang (2008) shows that if the testistic possesses the MLR property, then the p-value has decreasing pdf.

Why should be Beta mixture?

Ghosal, Roy and Tang (2008) proves the following results:

Why should be Beta mixture?

Ghosal, Roy and Tang (2008) proves the following results:

- $h(x) = \int_0^1 a x^{a-1} d\mu(a)$ for some probability measure μ if and only if $h(x)$ is decreasing and $H(e^{-x})$ is completely monotone, i.e., $(-1)^n d^n H(e^{-x})/dx^n \geq 0$ for $n = 1, 2, \cdots$

Why should be Beta mixture?

Ghosal, Roy and Tang (2008) proves the following results:

- $h(x) = \int_0^1 a x^{a-1} d\mu(a)$ for some probability measure μ if and only if $h(x)$ is decreasing and $H(e^{-x})$ is completely monotone, i.e., $(-1)^n d^n H(e^{-x})/dx^n \geq 0$ for $n = 1, 2, \cdots$

- $h(x) = \int_1^\infty b(1-x)^{b-1} d\mu(b)$ for some probability measure μ if and only if $h(x)$ is decreasing and $1 - H(1 - e^{-x})$ is completely monotone.

Further model reduction

Although $l(\pi, G)$ is concave, it is not clear how to compute the MLE in the nonparametric infinite mixture model, i.e., how to maximize $Q(h)$ in the M-step of EM algorithm.

Further model reduction

Although $l(\pi, G)$ is concave, it is not clear how to compute the MLE in the nonparametric infinite mixture model, i.e., how to maximize $Q(h)$ in the M-step of EM algorithm.

Reduce the nonparametric model further by considering independent mixing distribution: $G(a, b) = G_1(a)G_2(b)$.

- $G_1\{(0, 1]\} = 1$
- $G_2\{(1, C]\} = 1$ for some finite $C > 0$.

Further model reduction

Although $l(\pi, G)$ is concave, it is not clear how to compute the MLE in the nonparametric infinite mixture model, i.e., how to maximize $Q(h)$ in the M-step of EM algorithm.

Reduce the nonparametric model further by considering independent mixing distribution: $G(a, b) = G_1(a)G_2(b)$.

- $G_1\{(0, 1]\} = 1$
- $G_2\{(1, C]\} = 1$ for some finite $C > 0$.

Reimman sum approximation

Approximate the integral $\int b(x|a,b)dG_1(a)dG_2(b)$ by the Reimman sum as follows:

Let $\{a_1, \cdots, a_l\}$ be a net on $(0,1]$ and $\{b_1, \cdots, b_k\}$ be a net on $(1,C]$ such that

$$G_{1l}(a)G_{2k}(b) = \sum_{s=1}^{l}\sum_{t=1}^{k}\gamma_s\beta_t I(a_s \leq a)I(b_t \leq b)$$

approaches $G(a,b)$.

Reimman sum approximation

Approximate the integral $\int b(x|a,b)dG_1(a)dG_2(b)$ by the Reimman sum as follows:

Let $\{a_1, \cdots, a_l\}$ be a net on $(0,1]$ and $\{b_1, \cdots, b_k\}$ be a net on $(1, C]$ such that

$$G_{1l}(a)G_{2k}(b) = \sum_{s=1}^{l}\sum_{t=1}^{k}\gamma_s\beta_t I(a_s \leq a)I(b_t \leq b)$$

approaches $G(a,b)$. The Reimman sum approximation results in a finite Beta mixture:

$$h_{lk}(x) = \sum_{s=1}^{l}\sum_{t=1}^{k}\gamma_s\beta_t b(x|a_s, b_t).$$

The M-step

$$Q(h_{lk}) = \sum_{i}^{m} \hat{z}_i \log \left\{ \sum_{s=1}^{l} \sum_{t=1}^{k} \gamma_s \beta_t b(x|a_s, b_t) \right\}.$$

The M-step

$$Q(h_{lk}) = \sum_{i}^{m} \hat{z}_i \log \left\{ \sum_{s=1}^{l} \sum_{t=1}^{k} \gamma_s \beta_t b(x|a_s, b_t) \right\}.$$

In the M-step,

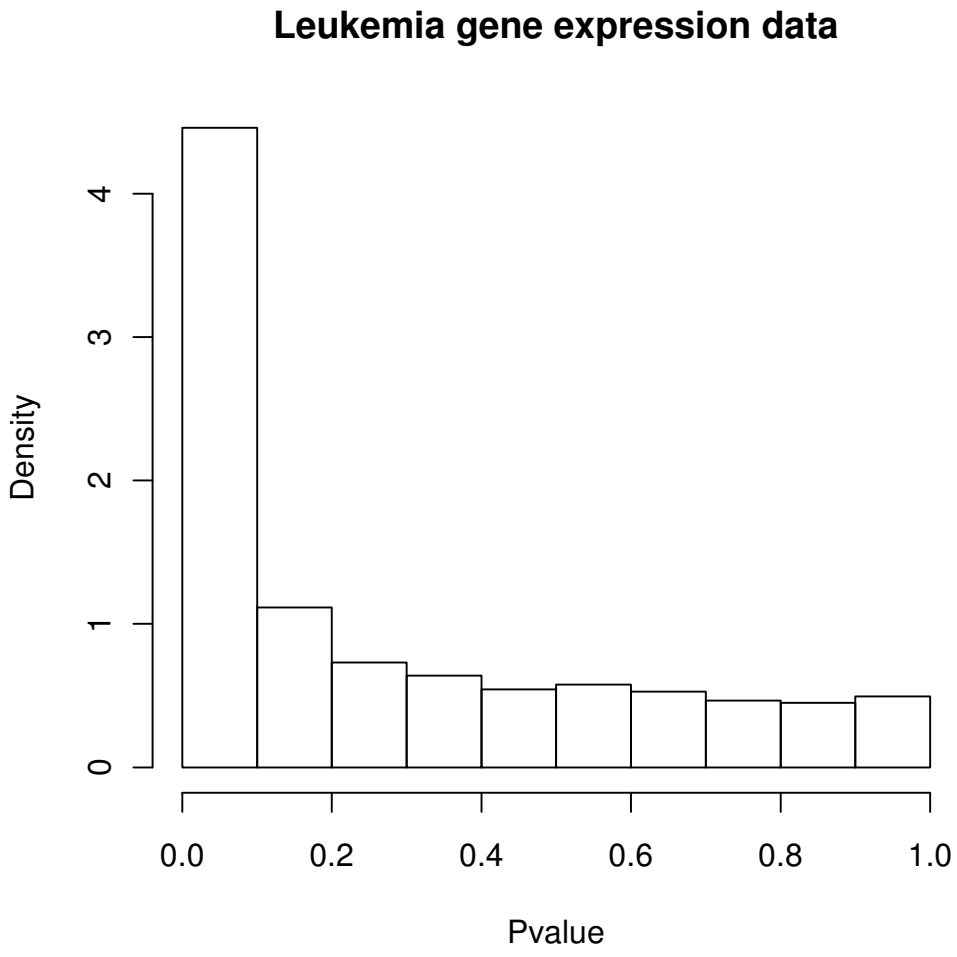
$$\begin{aligned}
\hat{\pi} &= \frac{1}{m} \sum_{i=1}^{m} \frac{\pi_0}{f_0(p_i)} \\
\hat{\gamma}_s &= \frac{1-\pi_0}{1-\hat{\pi}} \frac{\gamma_{0s}}{m} \sum_{i=1}^{m} \frac{\sum_{t=1}^{k} \beta_{0t} b(p_i|a_s, b_t)}{f_0(p_i)} \\
\hat{\beta}_t &= \frac{1-\pi_0}{1-\hat{\pi}} \frac{\beta_{0t}}{m} \sum_{i=1}^{m} \frac{\sum_{s=1}^{l} \gamma_{0s} b(p_i|a_s, b_t)}{f_0(p_i)}
\end{aligned}$$

Consequences of Reimman sum approximation

- Uniformly,

$$\frac{1}{m} l(\pi, G) \to E\{\log f(P|\pi, G)\}.$$

- $E\{\log f(P|\pi, G)\}$ is strictly concave downward.
- $(1/m)l(\pi, G) - (1/m)l(\pi, G_{lm,km}) = o(1)$.
- $(1/m)l(\pi, G_{lm,km})$ is strictly concave downward.
- The maximizer of $l(\pi, G_{lm,km})$ converges to the maximizer of $E\{\log f(P|\pi, G)\}$.
- The approximating MLE is consistent.


Leukemia gene expression data
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Pvalue

The Leukemia gene expression data example

- $m = 3051$. Take $C = 5$.
- Take $l = k = 50$; $a_s = s/50$, $b_t = 5t/50$.

The Leukemia gene expression data example

- $m = 3051$. Take $C = 5$.
- Take $l = k = 50$; $a_s = s/50$, $b_t = 5t/50$.
- $\hat{\pi} = 0.4601$
- A lot of $\hat{\gamma}_s$ are almost zero.
- $\hat{\beta}_t$ are quite evenly distributed.
- The score functions are reasonably close to zero.

Leukemia gene expression example

```
> data(golub);
> TT<-teststat<-mt.teststat(golub,golub.cl,test="t")
> #  the two samples t test for equal variance
> m<-length(TT); # Sample size
> P<-2*(1-pt(abs(TT),df=36));
>
> am=50
> bm=50
> C=5
>
>
> print("PI0")
[1] "PI0"
> print(PI0)
[1] 0.460122
> print("GAMMA")
[1] "GAMMA"
> print(GAMMA)
 [1] 2.173417e-106  9.610702e-57  8.682050e-34  4.920911e-21  2.161781e-13
 [6]  1.198973e-08  1.151412e-05  8.473611e-04  1.176600e-02  5.450118e-02
[11]  1.220770e-01  1.692373e-01  1.714750e-01  1.421649e-01  1.041613e-01
[16]  7.107813e-02  4.681205e-02  3.047824e-02  1.993181e-02  1.322855e-02
[21]  8.967835e-03  6.233229e-03  4.450746e-03  3.267002e-03  2.464919e-03
[26]  1.910265e-03  1.519048e-03  1.237922e-03  1.032443e-03  8.799829e-04
[31]  7.654195e-04  6.784751e-04  6.120509e-04  5.611700e-04  5.222947e-04
[36]  4.928792e-04  4.710724e-04  4.555172e-04  4.452134e-04  4.394221e-04
[41]  4.375990e-04  4.393468e-04  4.443810e-04  4.525050e-04  4.635917e-04
[46]  4.775702e-04  4.944153e-04  5.141401e-04  5.367900e-04  5.624381e-04
> print("BETA")
[1] "BETA"
> print(BETA)
 [1] 0.03404166 0.02881158 0.02537966 0.02300246 0.02128951 0.02001810
 [7] 0.01905295 0.01830783 0.01772573 0.01726788 0.01690725 0.01662469
[13] 0.01640632 0.01624200 0.01612415 0.01604706 0.01600636 0.01599866
[19] 0.01602128 0.01607209 0.01614939 0.01625176 0.01637806 0.01652733
[25] 0.01669877 0.01689170 0.01710554 0.01733981 0.01759409 0.01786805
[31] 0.01816138 0.01847383 0.01880522 0.01915537 0.01952415 0.01991148
[37] 0.02031729 0.02074154 0.02118421 0.02164534 0.02212494 0.02262310
[43] 0.02313988 0.02367539 0.02422976 0.02480313 0.02539566 0.02600754
[49] 0.02663896 0.02729014
>
>
> print("SCORE")
[1] "SCORE"
> print(SCORE)
 [1]  1.375696e-01 -1.061758e+03 -6.857300e+02 -4.396215e+02 -2.780448e+02
 [6] -1.719303e+02 -1.024596e+02 -5.732600e+01 -2.840709e+01 -1.029933e+01
[11]  6.153460e-01  6.771187e+00  9.810821e+00  1.084382e+01  1.061881e+01
[16]  9.639057e+00  8.240628e+00  6.645583e+00  4.998310e+00  3.390424e+00
[21]  1.877852e+00  4.925647e-01 -7.493937e-01 -1.842404e+00 -2.787530e+00
[26] -3.589988e+00 -4.257438e+00 -4.798853e+00 -5.223777e+00 -5.541845e+00
[31] -5.762481e+00 -5.894722e+00 -5.947114e+00 -5.927666e+00 -5.843831e+00
[36] -5.702518e+00 -5.510103e+00 -5.272458e+00 -4.994982e+00 -4.682625e+00
[41] -4.339924e+00 -3.971032e+00 -3.579741e+00 -3.169516e+00 -2.743512e+00
[46] -2.304606e+00 -1.855412e+00 -1.398303e+00 -9.354307e-01 -4.687432e-01
[51]  2.910765e-01  1.331956e-01  8.025329e-02  8.355356e-02  1.151065e-01
[56]  1.583646e-01  2.034684e-01  2.446642e-01  2.788069e-01  3.044339e-01
[61]  3.211639e-01  3.292951e-01  3.295329e-01  3.228037e-01  3.101315e-01
[66]  2.925552e-01  2.710767e-01  2.466297e-01  2.200620e-01  1.921289e-01
[71]  1.634915e-01  1.347204e-01  1.063010e-01  7.864067e-02  5.207649e-02
[76]  2.688299e-02  3.279729e-03 -1.856152e-02 -3.851017e-02 -5.647074e-02
[81] -7.237751e-02 -8.618965e-02 -9.788699e-02 -1.074663e-01 -1.149382e-01
[86] -1.203240e-01 -1.236540e-01 -1.249648e-01 -1.242981e-01 -1.216988e-01
[91] -1.172144e-01 -1.108936e-01 -1.027856e-01 -9.293961e-02 -8.140413e-02
[96] -6.822674e-02 -5.345370e-02 -3.712979e-02 -1.929811e-02
>
>
```

>
>

The Leukemia example: Delete zero-gamma

Among $\hat{\gamma}_1, \cdots, \hat{\gamma}_{50}$, only $\hat{\gamma}_9, \cdots, \hat{\gamma}_{20}$ are greater than 0.009. Delete all other γ_s and rerun the EM algorithm. The result is almost the same:

$$\hat{\pi} = 0.4611$$

The Leukemia example: Delete zero-gamma

Among $\hat{\gamma}_1, \cdots, \hat{\gamma}_{50}$, only $\hat{\gamma}_9, \cdots, \hat{\gamma}_{20}$ are greater than 0.009. Delete all other γ_s and rerun the EM algorithm. The result is almost the same:

$$\hat{\pi} = 0.4611$$

We also run the EM algorithm with $l = 25, k = 35$ and $C = 3$:

$$\hat{\pi} = 0.4534$$

```
> data(golub);
> TT<-teststat<-mt.teststat(golub,golub.cl,test="t")
> #  the two samples t test for equal variance
> m<-length(TT); # Sample size
> P<-2*(1-pt(abs(TT),df=36));
>
> am=25
> bm=35
> M=3
>
> aa<-1:am
> bb<-1:bm
> aa=aa/(am+1)
> bb=bb*M/bm
> ca=c(6:9)
> aa=aa[ca]
> am=length(aa)
>

> print("PI0")
[1] "PI0"
> print(PI0)
[1] 0.453145
> print("GAMMA")
[1] "GAMMA"
> print(GAMMA)
[1] 0.41991467 0.43258538 0.12303543 0.02446452
> print("BETA")
[1] "BETA"
> print(BETA)
 [1] 0.02869092 0.02177004 0.01802176 0.01588381 0.01465945 0.01400251
 [7] 0.01372607 0.01372244 0.01392641 0.01429704 0.01480808 0.01544258
[13] 0.01618978 0.01704320 0.01799946 0.01905748 0.02021800 0.02148319
[19] 0.02285639 0.02434200 0.02594525 0.02767221 0.02952964 0.03152500
[25] 0.03366637 0.03596243 0.03842247 0.04105636 0.04387452 0.04688795
[31] 0.05010822 0.05354745 0.05721836 0.06113423 0.06530892
>
> print("SCORE")
[1] "SCORE"
> print(SCORE)
 [1]  0.2471124  2.7559327  4.5748350  2.8644245 -1.6870540 -2.0662217
 [7] -2.2988175 -2.4313474 -2.4942600 -2.5081998 -2.4875502 -2.4425185
[13] -2.3804249 -2.3065397 -2.2246516 -2.1374685 -2.0469063 -1.9543015
[19] -1.8605680 -1.7663139 -1.6719279 -1.5776441 -1.4835894 -1.3898180
[25] -1.2963370 -1.2031249 -1.1101441 -1.0173501 -0.9246978 -0.8321451
[31] -0.7396558 -0.6472005 -0.5547571 -0.4623114 -0.3698559 -0.2773898
[37] -0.1849178 -0.0924499
>
>
>
>
```

The example: Who is right?

The example is studied by many authors in various approaches:

New	λ-Boot	Convex Density	Bernstein Polynomial
0.461	0.499	0.376	0.449

The example: Who is right?

The example is studied by many authors in various approaches:

New	λ-Boot	Convex Density	Bernstein Polynomial
0.461	0.499	0.376	0.449

Who is right?

$\pi = 0.376$ or $\pi = 0.499$?

The example: Who is right?

The example is studied by many authors in various approaches:

New	λ-Boot	Convex Density	Bernstein Polynomial
0.461	0.499	0.376	0.449

Who is right?

$\pi = 0.376$ or $\pi = 0.499$?

Note that a 12% difference on π means about 360 genes.

Remark

What if the true $h(1) \neq 0$?

Remark

What if the true $h(1) \neq 0$?

The area of square $[0, a] \times [0, a]$ equals

$$A = a^2$$

Remark

What if the true $h(1) \neq 0$?

The area of square $[0, a] \times [0, a]$ equals

$$A = a^2$$

What if the object is not a square?

THANK YOU.