Published in final edited form as:

J Appl Stat. 2011 ; 38(5): 987–1005. doi:10.1080/02664761003692449.

Bayesian Frequentist hybrid Model wth Application to the Analysis of Gene Copy Number Changes

Ao Yuan¹, Guanjie Chen², Juan Xiong³, Wenqing He³, and Charles Rotimi²

¹National Human Genome Center, Howard University, Washington D.C. USA

²Center for Research on Genomics and Global Health, NHGRI, NIH, USA

³Department of Statistical and Actuarial Sciences, University of Western Ontario, Canada

Abstract

Gene copy number (GCN) changes are common characteristics of many genetic diseases. Comparative genomic hybridization (CGH) is a new technology widely used today to screen the GCN changes in mutant cells with high resolution genome-wide. Statistical methods for analyzing such CGH data have been evolving. Existing methods are either frequentist's, or full Bayesian. The former often has computational advantage, while the latter can incorporate prior information into the model, but could be misleading when one does not have sound prior information. In an attempt to take full advantages of both approaches, we develop a Bayesian-frequentist hybrid approach, in which a subset of the model parameters is inferred by the Bayesian method, while the rest parameters by the frequentist's. This new hybrid approach provides advantages over those of the Bayesian or frequentist's method used alone. This is especially the case when sound prior information is available on part of the parameters, and the sample size is relatively small. Spatial dependence and false discovery rate are also discussed, and the parameter estimation is efficient. As an illustration, we used the proposed hybrid approach to analyze a real CGH data.

Keywords

Bayesian; Gene copy number; Frequentist; Hybrid model; prior information

1. Introduction

Recent evidences show that gene copy number (GCN) amplifications and deletions are common characteristics of many diseases caused by genetic mutations. For example, GCN can be elevated in cancer cells as demonstrated in the epidermal growth factor receptor (EGFR) gene in patients with non-small cell lung cancer (Cappuzzo *et al.* 2005) and also higher copy number of CCL3L1 has been associated with susceptibility to human HIV infection (Gonzalez *et al.* 2005). Thus, identifying these genetic gains and losses provides useful information about the genesis of specific diseases. GCN analysis among normal people within the human genome is also of interest. However, these genetic characteristics are usually not directly observable. Recent technology development on comparative genomic hybridization (CGH) array allows to screen putative mutant cells with high resolution for genome-wide analysis of copy number changes (Solinas *et al.*, 1997; Pinkel *et al.*, 1998; Snijders *et al.*, 2001; Hupe *et al.*, 2004; Myers *et al.*, 2004; Olshen *et al.*, 2004; Huang *et al.*, 2005; Lingjaerde *et al.*, 2005; Wang *et al.*, 2005; Guha *et al.*, 2008). The development of CGH has proved to be an efficient tool to study the genetic mutations responsible for disease development. The assay was first developed by Kallioniemi *et al.*, (1992). The basic strategy of the technique is to use genomic DNA from cancer cells labeled with one fluorochrome and genomic DNA from normal reference cells labeled with a second

fluorochrome and co-hybridize the labeled samples to a metaphase spread from a normal reference cell. The ratio, of matched pair on age, gender etc, of the two fluorochrome intensities is then calculated and regions where the cancer DNA are amplified or deleted are readily detected on the metaphase spread. After hybridization, emission from each of the two uorescent dyes is measured, and the signal intensity ratios are indicative of the relative copy number of the two samples. This technique not only gives us information about copy number gains and losses in the disease genomic DNA but also allows us to identify the specific chromosomes and the regions of the chromosomes where these changes have occurred.

However, the CGH data does not provide direct measurements of the GCN changes. Hence, various statistical approaches for analyzing and describing results from these experiments have been developed.

The existing methods are either frequentist's (Hodgson *et al.*, 2001; Pollack *et al.*, 2002; Cheng *et al.*, 2003; Wang *et al.*, 2004; Yuan *et al.*, 2008), or full Bayesian (Barash and Friedman, 2002; Daruwala *et al.*, 2004; Broet and Richardson, 2005). The former often has computational advantage, while the latter can incorporate useful prior information into the model, but can also be misleading when such prior information is not reliable. For finite dimensional parametric models, it is known that under very general conditions the frequentist's MLE and Bayesian estimates are asymptotically equivalent, in that the convergence to the truth of one implies that of the other, and both of them have the inverse Fisher information as asymptotic variance(matrix), this is the Bernstein-von Mises theorem. But in practice, sample size is always finite, and the two methods differ in many aspects. For Bayesian inference, the choice of prior is a source of debating, and the computation is much more demanding, that's the main reason why the real practice is dominated by the frequentist's methods. However, if the prior represents a good summary of the experience about the underlying parameters, a Bayesian analysis do have small sample advantages over the frequentist's. For example, if $X \sim \text{Binomial}(m, \theta)$, $\pi(\cdot) = U(0, 1)$, then given data x , the posterior is $\pi(\theta|x) = \text{Beta}(x+1, m-x+1)$. The frequentist estimate of θ is the MLE $\hat{\theta} = x$. Under the squared error loss, the Bayes estimate of θ is the posterior mean $\tilde{\theta} = (x+1)/(m+2)$. If $x = 0$ or 1 , the MLE of θ is 0 or 1 , while the Bayes estimate is only close to 0 or 1 , which is more reasonable. But if we do not have good prior experience, or the prior density does not characterize the parameter distribution reasonably, then the result may be misleading, as almost all Bayesian estimates are biased (Blackwell and Grischick, 1954; Blackwell and Bickel, 1967), and computationally more demanding than the frequentist method, so in this case the MLE is more preferred. In this case one may put a non-informative prior on the part of parameters which we do not have confident prior information, to formulate a full Bayesian procedure. But the hybrid model has more flexibility, for example, as we can see in Section 2, that parameter estimates under the hybrid model can be formulated as a EM algorithm, which significantly simplify the computation for the mixture model, in contrast to that of a full Bayesian model.

Often in multi-dimensional parametric analysis, the parameters of interest can be partitioned into two subsets, and in practice we may have good knowledge in one subset of the parameters, and not enough experience on the other subset. In this case, a Bayesian analysis on one subset will benefit from the prior information, but for the other subset a frequentist method is preferred. Thus a hybrid Bayesian-frequentist method for the analysis of full parameters will be advantageous and is proposed in Yuan (2009). In CGH data analysis, each data sequence has three unobserved status to be inferred: copy number deletion, amplification, or normal (no deletion nor amplification). Often, for specified disease and chromosome, the experimenter can accumulate knowledge about each status, but relatively lack of experience on the other unknown nature which constitute part of the model, thus the

above method fits the case well. In section 2 we describe the method in detail, then illustrate its application to a real CGH data in section 3. Finally a brief discussion is given in section 4.

2. The Method

2.1 The Method in General Setting

We first give a general description of the hybrid method. The parameters is partitioned as $\theta = (\alpha, \beta)$, on part of the parameter α , the Bayesian method is preferred, and on the other part β , the MLE is more favored. Let $\mathbf{x} = (x_1, \dots, x_n)$ be the sample i.i.d. with density function $f(x|\alpha, \beta)$, the likelihood is $f(\mathbf{x}_n|\alpha, \beta) = \prod_{i=1}^n f(x_i|\alpha, \beta)$. Let $\pi(\alpha)$ be the prior density for α , and $\pi(\alpha|\mathbf{x}_n, \beta) = f(\mathbf{x}_n|\alpha, \beta)\pi(\alpha)/m(\mathbf{x}_n|\beta)$ the posterior density of α for given β , where $m(\mathbf{x}_n|\beta) = \int f(\mathbf{x}_n|\alpha, \beta) \times \pi(\alpha) d\alpha$ is the mixture over α at β . Let $\mathcal{D}$ be the decision space, $\mathbf{d}(\mathbf{x}_n) \in \mathcal{D}$ be a decision rule for inferring α , $W(\mathbf{d}(\mathbf{x}_n), \alpha)$ be the loss function for α , $R(\mathbf{d}, \alpha, \beta) = E_{(\alpha, \beta)} W(\mathbf{d}(\mathbf{x}_n), \alpha)$ be the risk, $R(\mathbf{d}|\beta) = \int R(\mathbf{d}, \alpha, \beta) \pi(\alpha) d\alpha$ be the Bayes risk, and $R(\mathbf{d}|\mathbf{x}_n, \beta) = \int W(\mathbf{d}(\mathbf{x}_n), \alpha) \pi(\alpha|\mathbf{x}_n, \beta) d\alpha$ the posterior risk for α at β . Then $R(\mathbf{d}|\beta) = \int R(\mathbf{d}|\mathbf{x}_n, \beta) m(\mathbf{x}_n|\beta) d\mathbf{x}_n$. The Bayes decision for α given β is $\mathbf{d}^*(\cdot) = \mathbf{d}^*(\cdot|\beta) = \arg \inf_{\mathbf{d} \in \mathcal{D}} R(\mathbf{d}|\beta)$, and it is known that, $\mathbf{d}^*(\mathbf{x}_n) = \arg \inf_{\mathbf{d} \in \mathcal{D}} R(\mathbf{d}|\mathbf{x}_n, \beta)$ (a.s.), the generalized Bayesian estimator $\hat{\alpha}_n$ given β .

In this hybrid inference, we infer α by the above generalized Bayesian rule for each fixed β , and at the same time infer β by the frequentist MLE, i.e. we are to find $\theta_n = (\hat{\alpha}_n, \hat{\beta}_n) = (\hat{\alpha}(\mathbf{x}_n), \hat{\beta}(\mathbf{x}_n))$ such that

$$(\hat{\alpha}_n, \hat{\beta}_n) = \arg \inf_{(\mathbf{d}, \beta)} \int W(\mathbf{d}(\mathbf{x}_n), \alpha) f(\mathbf{x}_n|\alpha, \beta) \pi(\alpha) d\alpha \quad (1)$$

is the joint optimization over $(\mathbf{d}, \beta)$. Note that by imposing a 0-1 loss and constant prior on β , (1) can be formulated as a double minimization over $(\mathbf{d}, \beta)$ in a full Bayesian sense. Thus $(\hat{\alpha}_n, \hat{\beta}_n)$ always exists and is generally locally unique. Here in $(\hat{\alpha}_n, \hat{\beta}_n)$, $\hat{\alpha}_n$ is a generalized Bayes estimator of α and $\hat{\beta}_n$ the MLE of β . This hybrid estimator is consistent and efficient while keeping the small sample advantage for the components of parameters with good prior knowledge. Some basic asymptotic properties are studied in Yuan (2009), here we discuss some computational aspect of it and address its application. Here, in a way, the use of MLE for parameter estimates are like using non-informative prior distributions for those parameters in the Bayesian sense, but there is a difference: under the commonly used squared loss, the estimate of β is the posterior mean (with respect to the non-informative prior), while in the hybrid method, estimate of β is a MLE. However, if we use the 0-1 loss for β , and a constant prior, a non-informative prior (but not Jeffreys' non-informative prior), then the hybrid estimate is full Bayesian procedure. This is also true also for the full MLE, which is a special Bayesian estimate under 0-1 loss and constant prior. In this sense, many frequentist's methods can be regarded as special Bayesian (Wald's complete class theory states that any admissible procedure can be formulated as a Bayesian or limit of Bayesian procedures). But in the literature they are not called (special) Bayesian for historical and technical reasons, as they differ from standard Bayesian procedures. Similarly we do not call the hybrid method a Bayesian, although it can be formulated as a full Bayesian under special loss and prior. Also, we haven't seen in the literature a Bayesian model as our hybrid one by choosing such special loss and prior on part of the parameters. They may use a non-informative prior on part of the parameters, but typically a squared error or absolute error loss will be used for all parameters.

Although in some cases, $(\check{\alpha}_n, \hat{\beta}_n)$ has closed form expression, but the general solution of (1) may not easy. Denote $G_n(\mathbf{d}, \beta) = \int W(\mathbf{d}|\mathbf{x}_n, \alpha) f(\mathbf{x}_n|\alpha, \beta) \pi(\alpha) d\alpha$. Let $\sup_{\beta} \inf_{\mathbf{d}}$ denote first taking inf w.r.t. $\mathbf{d}$ then taking sup w.r.t. β ; $\inf_{(\mathbf{d}, \sup_{\beta})}$ the other way round, and $\inf_{(\mathbf{d}, \beta)}$ for taking inf w.r.t. $\mathbf{d}$ and the sup w.r.t. β simultaneously. Since $\sup_{\beta} \inf_{\mathbf{d}} G_n(\mathbf{d}, \beta) \leq \inf_{(\mathbf{d}, \beta)} G_n(\mathbf{d}, \beta) \leq \inf_{\mathbf{d}} \sup_{\beta} G_n(\mathbf{d}, \beta)$, and the “=” signs are not generally asserted, so generally $\arg \sup_{\beta} \inf_{\mathbf{d}} G_n(\mathbf{d}, \beta) (\check{\alpha}_n, \hat{\beta}_n) = \arg \inf_{(\mathbf{d}, \sup_{\beta})} G_n(\mathbf{d}, \beta) \arg \inf_{\mathbf{d}} \sup_{\beta} G_n(\mathbf{d}, \beta)$. However, if $\arg \inf_{\mathbf{d}} G_n(\mathbf{d}, \beta)$ does not depend on β , then $(\check{\alpha}_n, \hat{\beta}_n) = \arg \sup_{\beta} \inf_{\mathbf{d}} G_n(\mathbf{d}, \beta)$. Similarly, if $\arg \sup_{\beta} G_n(\mathbf{d}, \beta)$ does not depend on $\mathbf{d}$, then $(\check{\alpha}_n, \hat{\beta}_n) = \inf_{\mathbf{d}} \sup_{\beta} G_n(\mathbf{d}, \beta)$.

When $(\check{\alpha}_n, \hat{\beta}_n)$ is not directly computable, iterative procedures can be used. Below we give two such procedures although they are not directly used in this analysis. The first is a version of the multivariate Newton-Raphson method. Let $\ell(\alpha, \beta|\mathbf{x}_n)$ be the log-likelihood. We rewrite $\check{\theta} = (\check{\alpha}, \hat{\beta})$ as the solution of the equations

$$\begin{cases} \frac{\partial}{\partial \alpha} R(\alpha|\mathbf{x}_n, \beta) = 0 \\ \frac{\partial}{\partial \beta} l(\alpha, \beta|\mathbf{x}_n) = 0 \end{cases}.$$

For a vector $\mathbf{v}$ denote $\mathbf{v}'$ for its transpose. Let

$R^{(1)}(\theta) = ((\frac{\partial}{\partial \alpha} R(\alpha, \beta))', (\frac{\partial}{\partial \beta} l(\alpha, \beta|\mathbf{x}_n)))'$, $R^{(2)} = \frac{\partial}{\partial \theta} R^{(1)}(\theta)$. Choosing an arbitrary starting point $\theta^{(0)}$, this iterative procedure gives

$$\theta^{(k+1)} = \theta^{(k)} + (R^{(2)})^{-1}(\theta^{(k)})(R^{(1)}(\theta^{(k+1)}) - R^{(1)}(\theta^{(k)})), \quad k=0, 1, 2, \dots \quad (2)$$

It iterates between the two equations of $\frac{\partial}{\partial \alpha} R=0$ and $\frac{\partial}{\partial \beta} l=0$ in the above to get the sequence $\{(\alpha^{(k)}, \beta^{(k)})\}$. By the property of this well known procedure, without proof, we have

Proposition 1—Assume $|R^{(2)}(\theta)| \neq 0$ in a neighborhood of $\check{\theta}_n$, then as $k \rightarrow \infty$, $\{(\alpha^{(k)}, \beta^{(k)})\}$ converges to a stationary point such that (1) holds. If further, for fixed n , there is a unique point $(\check{\alpha}_n, \hat{\beta}_n)$ such that (1) holds, then as $k \rightarrow \infty$,

$$(\alpha^{(k)}, \beta^{(k)}) \rightarrow (\check{\alpha}_n, \hat{\beta}_n).$$

More generally, given β , a Bayesian procedure has the form $\check{\alpha}_n = G(\pi(\cdot|\mathbf{x}_n, \beta))$ for some known functional $G(\cdot)$. For squared error loss, $G(\cdot)$ is the posterior mean; for absolute error loss, $G(\cdot)$ is the posterior median; for 0-1 loss, $G(\cdot)$ is the posterior mode. The following algorithm is more intuitive and sometimes convenient. To be specific, start from $(\alpha^{(0)}, \beta^{(0)})$, ..., given the k -th iteration estimates $(\alpha^{(k)}, \beta^{(k)})$, the next $(k+1)$ -th update is given by

$$\alpha^{(k+1)} = G(\pi(\alpha|\mathbf{x}_n, \beta^{(k)})), \quad \beta^{(k+1)} = \arg \sup_{\beta} l(\alpha^{(k+1)}, \beta|\mathbf{x}_n). \quad (3)$$

Let the Fisher information matrix $\mathcal{I}(\theta) = \mathcal{I}(\alpha, \beta)$ be partitioned into sub-blocks $\mathcal{I}_{ij}(\theta)$ ($i, j = 1, 2$), and $\mathcal{I}(\alpha|\beta) = -E^{-2} \alpha \alpha' \log f(X|\alpha, \beta)$ be the conditional Fisher information given β . We have (Appendix)

Proposition 2—Assume $G(\pi(\mathbf{a}|\mathbf{x}_n, \boldsymbol{\beta})) - G(N(\hat{\mathbf{a}}_n, F^{-1}(\mathbf{a}|\boldsymbol{\beta})/n)) \rightarrow 0$ as $n \rightarrow \infty$, that $\sup_{\mathbf{a} \in A} \|\boldsymbol{\beta} [G(N(\hat{\mathbf{a}}_n, F^{-1}(\mathbf{a}|\boldsymbol{\beta})/n)] I_{12}(\boldsymbol{\theta}) I_{22}(\boldsymbol{\theta})^{-1}\| < 1$ for some $A \supset (\check{\mathbf{a}}_n, \hat{\boldsymbol{\beta}}_n)$, and that there is a unique stationary point $(\check{\mathbf{a}}_n, \hat{\boldsymbol{\beta}}_n)$ satisfying (1), then for large n , the algorithm given by (3) converges as $k \rightarrow \infty$:

$$(\alpha^{(k)}, \beta^{(k)}) \rightarrow (\check{\alpha}_n, \hat{\beta}_n).$$

Under the squared error or absolute error losses, $G(\cdot)$ is the posterior mean or median, the condition $G(\pi(\mathbf{a}|\mathbf{x}_n, \boldsymbol{\beta})) - G(N(\hat{\mathbf{a}}_n, F^{-1}(\mathbf{a}|\boldsymbol{\beta})/n)) \rightarrow 0$ in the L_1 sense is implied by Theorem 2.2 in Bickel and Yahav (1969).

Especially, under the squared error loss, (3) becomes

$$\alpha^{(k+1)} = \int \alpha \pi(\alpha|\mathbf{x}_n, \beta^k) d\alpha, \quad \beta^{(k+1)} = \arg \sup_{\beta} l(\alpha^{(k+1)}, \beta|\mathbf{x}_n).$$

Under the 0-1 loss, it becomes a double maximization procedure,

$$\check{\alpha}_n = \arg \sup_{\alpha} \frac{f(\mathbf{x}_n|\alpha, \hat{\beta}_n)}{m(\mathbf{x}_n|\hat{\beta}_n)} = \arg \sup_{\alpha} (l(\alpha, \hat{\beta}_n|\mathbf{x}_n) + \log \pi(\alpha)), \quad \hat{\beta}_n = \arg \sup_{\beta} l(\check{\alpha}_n, \beta|\mathbf{x}_n)$$

$$(\check{\alpha}_n, \hat{\beta}_n) = \arg \sup_{\alpha, \beta} (l(\alpha, \beta|\mathbf{x}_n) + \log \pi(\alpha)).$$

2.2 The Model for the CGH Data

Now we detail this model for the CGH data. Let $\mathbf{x}_n = (x_1, \dots, x_n)$ be the case-control log-ratio measurement of a genomic sequence (GS), with $x_i, i = 1, \dots, n$ be the observation at the i -th locus. The x_i 's are assumed to be arranged according to their chromosomal location. Here we focus on just one GS, in the case of multiple independent GS's the corresponding likelihood is a product of those for each GS, and the method is analogous. In practice, the true copy number changes cannot be observed directly, but rather, with a CGH data, in which the uorescence ratios between two samples, case and control, are measured across a genomic region. For loci with copy number deletion/amplification, the corresponding log-ratio measurement tend to be lower/higher. We are to answer the question of what is the probability that a given gene or region has increased or decreased copy number changes. In CGH data analysis, often a three-state mixture model is used: deletion state, normal state and amplification state, and we arbitrarily label them as state 1, 2 and 3. Although in some cases the number k of states can be arbitrary but fixed, the analysis is similar in nature. Some method (Rueda and Díaz-Uriarte, 2007) treat k as a parameter to be estimated using the reversible jump Markov model. Here we focus on the common situation for this problem of k being fixed at 3. Genes with copy number deletion tend to have smaller log-ratio measurements, those with normal status tend to have median measurements, and those with amplification tend to have larger measurements. The goal is to classify each the x_i s to one of the states. Without knowing these memberships, the density for the x_i s can be specified as the mixture

$$f(\mathbf{x}_n|\theta)=\prod_{i=1}^n\sum_{j=1}^3\gamma_j\varphi(x_i|\alpha_j,\sigma_j^2), \quad (4)$$

where $0 \leq \gamma_j \leq 1, \sum_{j=1}^3 \gamma_j = 1$ are the mixing proportions of the three states in the observed data, and $\varphi(\cdot|\alpha, \sigma^2)$ is the density of the normal distribution $N(\alpha, \sigma^2)$. Commonly for this problem, we have or can find good prior knowledge on $\mathbf{a} = (\alpha_1, \dots, \alpha_3)$ from existing such studies, as summarized by the prior density $\pi_j(\alpha_j) \sim N(\alpha_{j0}, \sigma_{j0}^2)$, with $(\alpha_{10}, \dots, \alpha_{30})$ known, and $\pi(\alpha) = \prod_{j=1}^3 \pi(\alpha_j)$. But we don't have not enough experience for the parameter $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_3)'$, $\boldsymbol{\beta}_j = (\gamma_j, \sigma_j^2)$ ($j=1, 2, 3$), as often $\boldsymbol{\beta}$ is difficult to infer from a prior initial study. One may also put a non-informative prior on $\boldsymbol{\beta}$ to formulate a full Bayesian model. But we will see below that by using the hybrid model, the parameter estimation can be formulated as a EM algorithm, which typically simplifies the computation considerably for mixture model as compared to a full Bayesian model. Thus we use a hybrid model with Bayesian estimate for $\mathbf{a}$ and the MLE for $\boldsymbol{\beta}$.

Given $\boldsymbol{\beta}$ and the data $\mathbf{x}_n$, the posterior density of $\mathbf{a}$ is

$$\pi(\alpha|\mathbf{x}_n, \boldsymbol{\beta}) \propto \pi(\alpha) \prod_{i=1}^n \sum_{j=1}^3 \gamma_j \varphi(x_i|\alpha_j, \sigma_j^2).$$

Note here $\pi(\mathbf{a}|\mathbf{x}_n, \boldsymbol{\beta})$ is different from the classical posterior $\pi(\mathbf{a}, \boldsymbol{\beta}|\mathbf{x}_n)$ of the full Bayesian model, the former does not contain any prior information about $\boldsymbol{\beta}$. Using the 0-1 loss on $\mathbf{a}$, its Bayesian solution is the posterior mode, the solution $(\check{\alpha}_n, \hat{\boldsymbol{\beta}}_n)$ in (1) is the maximizer of $\pi(\mathbf{a}|\mathbf{x}_n, \boldsymbol{\beta})$ over $(\mathbf{a}, \boldsymbol{\beta})$

$$(\check{\alpha}_n, \hat{\boldsymbol{\beta}}_n) = \arg \sup_{(\alpha, \boldsymbol{\beta})} \pi(\alpha) \prod_{i=1}^n \sum_{j=1}^3 \gamma_j \varphi(x_i|\alpha_j, \sigma_j^2).$$

Here $\pi(\mathbf{a}|\mathbf{x}_n, \boldsymbol{\beta})$ is the object function to be maximized.

The direct maximization in the above mixture model may not easy. Also, to classify each gene x_i , we need a status indicator for each gene. From computation aspect, more commonly and conveniently, the maximization problem in a mixture model is solved by an EM algorithm on the corresponding "complete data" model, such as in Meng and Rubin (1993). For this, let I_{ij} ($j=1, 2, 3$) be the unobserved status indicator for each x_i that is if x_i is in state j , then $I_{ij}=1$ and $I_{il}=0$ ($l \neq j$). Let $I_i = (I_{i1}, I_{i2}, I_{i3})$, $y_i = (x_i, I_i)$. Since the I_i s are unobserved, we treat them as missing data and $\mathbf{y}_n = (y_1, \dots, y_n)$ as the 'complete' data. We model the within chromosome dependence only through the I_i s. Then given the I_i s the x_i s are independent. Given $\mathbf{y}_n$ and $\boldsymbol{\beta}$, the posterior on $\mathbf{a}$ is

$$\pi(\alpha|\mathbf{y}_n, \boldsymbol{\beta}) \propto \pi(\alpha) \prod_{i=1}^n \prod_{j=1}^3 (\gamma_j \varphi(x_i|\alpha_j, \sigma_j^2))^{I_{ij}}$$

and omitting the proportionality constant, the corresponding logarithm is

$$l(\alpha, \beta | \mathbf{y}_n) = \sum_{i=1}^n \sum_{j=1}^3 I_{ij} (\log \gamma_j + \log \varphi(x_i | \alpha_j, \sigma_j^2)) + \log \pi(\alpha).$$

Here $l(\alpha, \beta | \mathbf{y}_n)$ is different from the traditional log-likelihood, as the prior $\pi(\alpha)$ is incorporated. Instead of maximizing the ‘incomplete data’ posterior $\pi(\alpha | \mathbf{x}_n, \beta)$, maximizing the ‘complete data’ log-likelihood is computationally much easier and this leads to an EM algorithm, but not the common EM algorithm as here the object function $\pi_0(\alpha, \beta | \mathbf{x}_n)$ does not have the full likelihood formulation. We will show it has the same property as the standard EM algorithm.

The estimator $(\tilde{\alpha}_{n,j}, \tilde{\sigma}_{n,j}^2)$ is the solution of two equations with two unknowns of order two, which can be easily obtained by the Newton-Raphson method. To get a closed form solution, we set $\sigma_{j0}^2 = \sigma_j^2$ ($j = 1, 2, 3$) (recall the α_{j0} ’s are known, but here the σ_j^2 are unknown and to be estimated). For given I_{ij} ’s, we get (Appendix), with $n_j = \sum_{i=1}^n I_{ij}$ for $j = 1, 2, 3$,

$$\tilde{\alpha}_{n,j} = \frac{\sum_{i=1}^n I_{ij} x_i + \alpha_{j0}}{n_j + 1}, \quad \tilde{\sigma}_{n,j}^2 = \frac{\sum_{i=1}^n I_{ij} (x_i - \tilde{\alpha}_{n,j})^2 + (\tilde{\alpha}_{n,j} - \alpha_{j0})^2}{n_j + 1}, \quad (j=1, 2, 3). \quad (5)$$

In comparison, if we used a full MLE for this problem, then we would have

$$\hat{\alpha}_{n,j} = \sum_{i=1}^n I_{ij} x_i / n_j, \quad \hat{\sigma}_{n,j}^2 = \sum_{i=1}^n I_{ij} (x_i - \hat{\alpha}_{n,j})^2 / n_j, \quad (j=1, 2, 3).$$

We see that, if the α_{j0} ’s represent useful information about α , $\tilde{\alpha}_n$ is more accurate than the MLE $\hat{\alpha}_n$, especially when the sample size is relatively small, or if the $\|\alpha_{j0}\|$ ’s have relatively large values.

On the other hand, if a full Bayes estimate is used, although we have the small sample advantage on $\tilde{\alpha}_n$, but since we don’t have accurate information on $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$, a prior on it may be miss leading, and the computation will be much more involved, even if we place a non-informative prior. It is known that if the prior is too “thin” around true parameter, the Bayes estimate can be inconsistent (LeCam and Yang, 2000).

Since the I_{ij} ’s are missing, we use the following EM algorithm for the computation. Specifically, start from given $(\alpha^{(0)}, \beta^{(0)})$, say,

$$\gamma_j^{(0)} = 1/3, \quad \alpha_j^{(0)} = \alpha_{j0}, \quad \sigma_j^{2(0)} = 1, \quad (j=1, 2, 3)$$

..., given the r -th iteration estimates $(\alpha^{(r)}, \beta^{(r)})$, the next $(r+1)$ -th updating take the following steps

- I. $I_{ij}^{(r+1)} = E(I_{ij} | \mathbf{x}_n, \alpha^{(r)}, \beta^{(r)}) = \gamma_j^{(r)} \varphi(x_i | \alpha_j^{(r)}, \sigma_j^{2(r)}) / \sum_{l=1}^3 \gamma_l^{(r)} \varphi(x_i | \alpha_l^{(r)}, \sigma_l^{2(r)})$.
- II. $\alpha_j^{(r+1)} = (\sum_{i=1}^n I_{ij}^{(r+1)} x_i + \alpha_{j0}) / (n_j^{(r+1)} + 1)$, and $n_j^{(r+1)} = \sum_{i=1}^n I_{ij}^{(r+1)}$, ($j = 1, 2, 3$).
- III. $\sigma_j^{2(r+1)} = (\sum_{i=1}^n I_{ij}^{(r+1)} (x_i - \alpha_j^{(r+1)})^2 + (\alpha_j^{(r+1)} - \alpha_{j0})^2) / (n_j^{(r+1)} + 1)$, ($j = 1, 2, 3$).
- IV. $\gamma_j^{(r+1)} = n_j^{(r+1)} / n$.

2.3 Dependence consideration

The genes in the human genome are more or less dependent each other. Generally such dependence varies in spatial location along the genome, from ethnical groups, population stratifications, etc. Although it is impossible to model such dependence accurately, a model with such considerations incorporated may still has some gains. It is known that in the same chromosome, genes closer in genetic distance tend to have stronger dependence with each other. To implement such spatial dependence, for each i , let R_i be the immediate neighbor of x_i which consists of the itself and the two neighboring measurements x_r and x_s in the given chromosome, except the measurements at the ends of the chromosome. Without loss of generality, we assume all the measurements are arranged according to their chromosomal location, thus x_1 is at the left end and x_n is at the right end of the chromosome. Then $R_i = \{x_{i-1}, x_i, x_{i+1}\}$ for ($i = 2, \dots, n-1$), $R_1 = \{x_1, x_2\}$ and $R_n = \{x_{n-1}, x_n\}$. Larger R_i can be considered similarly. The choice of the size and structure of the neighborhoods R_i 's is a trade-off between goodness-of-fit and applicability/simplicity. Here we used neighborhood size of 3 as experiences suggest that dependence among genes of immediate spatial members are more important generally, and increase the neighborhood structure may not gain much due to the complexity. We assume each of the I 's are spatially dependent only on R_i . Let $d(i, j)$ be the genetic distance, scaled to $[0, 1]$, (assumed known) between locus i and j . These distances will be used when spatial dependence is to be incorporated into the model. Here we distinguish such dependence from different chromosomes when there are more than one. Similarly as in Besag *et al.* (1991), we specify the spatial dependence relationship as $(I_{i1}, I_{i2}, I_{i3}) | R_i \sim \text{Multinomial}(1, \mathbf{q}_i)$, where $\mathbf{q}_i = (q_{i1}, q_{i2}, q_{i3})$ and

$$q_{ij} = P(I_{ij} = 1 | R_i) = \frac{\sum_{x_k \in R_i} \gamma_j e^{-d(i,k)} \varphi(x_k | \alpha_j, \sigma_j^2)}{\sum_{j=1}^3 \sum_{x_k \in R_i} \gamma_j e^{-d(i,k)} \varphi(x_k | \alpha_j, \sigma_j^2)}, \quad (i=1, \dots, n; j=1, 2, 3). \quad (6)$$

This can be viewed as posterior assignment of dependence parameters. Note in the above $d(i, j) = 0$ for all i . When the $d(i, j)$'s are not available, we just set $d(i, i+1) = d(i+1, i) = 1$ for all i and all the loci are viewed as equally spaced. Other method of dependence models for this problem can be found, for example in Fridlyand *et al.* (2004).

With the spatial dependence considered above, (I) in the above should be replaced by

$$I_{ij}^{(r+1)} = E(I_{ij} | \mathbf{x}_n, \alpha^{(r)}, \beta^{(r)}) = \frac{\sum_{x_k \in R_i} \gamma_j^{(r)} e^{-d(i,k)} \varphi(x_k | \alpha_j^{(r)}, \sigma_j^{2(r)})}{\sum_{j=1}^3 \sum_{x_k \in R_i} \gamma_j^{(r)} e^{-d(i,k)} \varphi(x_k | \alpha_j^{(r)}, \sigma_j^{2(r)})}.$$

Lastly, the resulting three clusters are classified as deletion, normal and amplification groups according to the magnitude of their means.

However the above is not the standard EM algorithm, here $l(\alpha, \beta | \mathbf{y}_n)$ is not a proper log-likelihood due to the extra term $\log \pi(\alpha)$. Define

$$S(\alpha | \mathbf{x}_n, \beta) = \pi(\alpha) \prod_{i=1}^n \sum_{j=1}^k \gamma_j \varphi(x_i | \alpha_j, \Omega_j).$$

The above algorithm still has the ascending property enjoyed by the standard EM algorithm, i.e.

Proposition 3—For all non-negative integer r , we have (Appendix)

$$\log S(\alpha^{(r+1)} | \mathbf{x}_n, \beta^{(r+1)}) \geq \log S(\alpha^{(r)} | \mathbf{x}_n, \beta^{(r)}),$$

so that, as for the standard EM algorithm, if there is a unique maximizer $(\check{\alpha}_n, \check{\beta}_n)$ in $\pi(\cdot | \mathbf{x}_n, \cdot)$, then as $r \rightarrow \infty$, we have $(\alpha^{(r)}, \beta^{(r)}) \rightarrow (\check{\alpha}_n, \check{\beta}_n)$.

Remark: Although we state the result only for the mixture model, Proposition 3 is true for general model with general form of missing data.

The algorithm stops at the last iteration R when the convergence criterion is met, and we set $(\hat{\alpha}_n, \hat{\beta}_n) = (\alpha^{(R)}, \beta^{(R)})$. Typically, we may choose the relative error criterion. i.e., specify a tolerance error bound ε (usually 10^{-2}), we stop the iteration if

$$\|(\theta^{(R)} - \theta^{(R-1)}) / \theta^{(R-1)}\| = \left(\sum_j (\theta_j^{(R)} - \theta_j^{(R-1)})^2 / \theta_j^{2(R-1)} \right)^{1/2} \leq \varepsilon$$

or just the Euclidian error, if there are some zero or nearly zero component(s) in θ , as

$$\|(\theta^{(R)} - \theta^{(R-1)})\| = \left(\sum_j (\theta_j^{(R)} - \theta_j^{(R-1)})^2 \right)^{1/2} \leq \varepsilon.$$

We classify each x_i to the j -th state if

$$\widehat{\gamma}_j \varphi(x_i | \check{\alpha}_j, \widehat{\sigma}_j^2) = \max_{1 \leq l \leq 3} \widehat{\gamma}_l \varphi(x_i | \check{\alpha}_l, \widehat{\sigma}_l^2). \quad (7)$$

This is the optimal classification rule in the sense of minimizing the expected loss (Anderson, 2003), it is also the so-called Bayesian assignment.

Let

$$D_j = \{y: y \in R, \widehat{\gamma}_j \varphi(y | \check{\alpha}_j, \widehat{\sigma}_j^2) = \max_{1 \leq l \leq 3} \widehat{\gamma}_l \varphi(y | \check{\alpha}_l, \widehat{\sigma}_l^2)\}.$$

The false discovery rate p_{jl} is the probability of mis-classifying a measurement to state j when its true state is l ,

$$p_{jl} = \int_{D_j} \varphi(y|\tilde{\alpha}_l, \tilde{\sigma}_l^2) dy, \quad (1 \leq j \neq l \leq 3).$$

This probability may not easy to compute directly in general. But it is easily computed by the following Monte Carlo simulation.

For given number M (typically $M = 1000$), for $m = 1, \dots, M$, do the following steps

- A. Sample $y_l^{(m)}$ from the distribution $\varphi(y|\tilde{\alpha}_l, \tilde{\sigma}_l^2)$, ($l = 1, 2, 3$).
- B. Define $v_{jl}^{(m)} = 1$ if

$$\widehat{\gamma}_j \varphi(y_l^{(m)}|\tilde{\alpha}_j, \tilde{\sigma}_j^2) = \max_s \widehat{\gamma}_s \varphi(y_l^{(m)}|\tilde{\alpha}_s, \tilde{\sigma}_s^2),$$

$$\text{else } v_{jl}^{(m)} = 0 \quad (j = 1, 2, 3).$$

Then π_{jl} is approximated by

$$\widehat{p}_{jl} = \frac{1}{M} \sum_{j=1}^M v_{jl}^{(m)}, \quad (j, l = 1, 2, 3; j \neq l).$$

Simulation studies in Yuan (2009) for the independence model case indicate that the hybrid method can improve the inference, especially in the small sample case, when the method applied properly. In the following we illustrate the method with a real CGH data.

3. Analysis of real CGH Array Data

We use the Genetics Analysis Workshop 15 (GAW15) data set with 14 pedigrees of CEPH Utah families, each with three generations and about a dozen normal individuals. Expression level of genes in lymphoblastoid cells of the above subjects were obtained using the Affymetrix Human Focus Arrays that contain probes for 8,500 transcripts. Gene copy number variations in normal people within human genome has been the subject of study (Freeman et al. 2006; Pugh et al. 2008). We analyze the genes, over 7 time interval, of the second individual in this data set. As an illustration, we use the data in the first time interval as a training set to get the prior information about the means, but the information on the variances and the proportion parameters are not used, as it is known that these parameters are difficult to estimate, and the initial estimates may not be reliable on these parameters based on a small sample size. Therefore, the hybrid method is adequate for this problem, and use this partial prior information on the mean, we analyze the 21562 genes on the all the remaining time intervals, and only report the results on the 7-th time interval of the second individual. Results on all the six time intervals and all the observed individuals are similar and omitted due to space limit. The model parameter estimates are compared for the traditional maximum likelihood estimator (MLE), Bayes estimate (BYE) and the proposed hybrid estimator (HBD) for the parameter $\theta = (\gamma_1, \gamma_2, \gamma_3, \alpha_1, \alpha_2, \alpha_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)$. In the hybrid estimation, the components $(\alpha_1, \alpha_2, \alpha_3)$ are estimated Bayesian, with a prior $\pi(\alpha) \sim \mathcal{N}(\alpha_0,$

$\text{diag}(\sigma^2)$), where $\alpha_0 = (-0.85, 0.22, 1.35)$ is obtained from the data on the first time interval, and $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$. For the full Bayesian estimate, we used $\pi(\theta) = \pi(\gamma) \pi(\alpha) \pi(\sigma^2)$, with $\pi(\alpha)$ the same as for the HBD, $\pi(\gamma)$ be uniform on $(0, 1)^3$, and $\pi(\sigma^2) = N((0.6, 1.5, 2.5), (0.5, 1, 2))$. Here we do not intentionally choose the prior mean too far away from the other estimates to avoid mis-leading results. The commonly used squared error loss is used for the full BYE, so that the estimates are the posterior means. For the hybrid estimate, we use the 0-1 loss on α , so that the Bayesian part estimate on α is the posterior mode, and thus the hybrid estimator $\theta_n = (\tilde{\alpha}_n, \hat{\beta}_n)$ is evaluated in closed form given in (5). By the rule of thumb, the minimal effective sample size to estimate one parameter is about 30. As there are nine parameters in θ , the effective sample size should be at least around 300. To see the effects of the sample size on the estimators, we computed the estimates with the three methods using the first 300, 1000, 10000 and all the genes respectively. We run the EM algorithm 500 iterations for all the estimates, with the convergence criterion been well achieved. There are six time intervals, we only display the results on the last time interval, those on the other intervals are similar and omitted. We did the analysis using both the independence gene model and the spatial dependence model via (6). Genetic distance information is not available for this data, so in (6) we just set the distances as a constant.

Since the large number of observations, instead of analyzing the classification result for each of the x_i 's, we discuss the results of the parameter estimations under the three different models, as better parameter estimates correspond to better classification results. The results are reported in Table 1 for the independence model, and Table 2 for dependence model. In the parentheses are the estimated standard errors of the parameters.

We can see that for both independence and dependence models, when the sample size is small (300) for this problem, compared to the estimates from the full data as a reference, the accuracy for the MLE, BYE and HBD are not good, but the overall performance of the HBD is better than those of the MLE and BYE, in that estimates from HBD and those of MLE are closer each other, reflecting the fact that good prior information can adjust the results from MLE. While for the BYE, as we do not have reliable prior information on the mixing proportions and variance parameters, its estimates show some larger deviation from those of the MLE, which we regard as more or less mis-leading. When the sample increase to the moderate size of 1000, all the estimators see some improvements, but still the HBD outperforms the MLE and BYE in overall in terms of accuracy. This suggest that for data with small or moderate size, the hybrid method can have some overall advantage than the MLE or Bayesian alone, when we have sound prior information about part of the parameters. When the sample size reaches the full 21562, all estimates show little difference, as expected, a confirmation of the statement in Remark 2 in Yuan (2009), that the Bayesian, MLE and hybrid estimators are asymptotically first order equivalent, normal and efficient. The results from the dependence model seems more reasonable, in that the performances of the three estimators are more homogeneous and stable. This may reflect the fact that the genes are dependent by nature and thus implement such feature improves the fitting, also as in the dependence model the effects are averaged in each gene neighborhood, the outcome is more smooth. Thus our interpretations of results are based on the HBD method with gene-gene spatial dependence model. In contrast to the case-control analysis of this problem, in which the data are collected as log ratios of signal measurements of cases to controls, thus deletion status is associated with negative observations, normal status with near zero observations and amplification with positive ones; here our data are from normal individuals, the means of all the three statuses are positive, with deletion status associated with the smallest means of about 0.215, and those for normal and amplification statuses are about 0.712 and 2.897. The corresponding proportions are 0.258, 0.737 and 0.006, thus about 73.7% genes have no alteration during the past history in the group of normal people

under investigation. About 25.8% of the genes have undergone repeat number deletion, but the average amount of deletion is much smaller (with a mean change of -0.497) in scale as compared to that for the amplification (with a mean change of 2.185), but the latter occurs with very small proportion, about 0.6%.

The classification results based on all the genes, using the hybrid method (almost the same by the other two methods for such a large sample size) of the second individual on the last time interval are shown in Figure 1 below without gene-gene spatial dependence, and in Figure 2 for the spatial dependence model. Horizontal axis represents the sequential numbering of genes from 1 to 21562, as separated into nine panels in each figure, and the vertical axis indicates the classified statuses of the genes with 1, 2 and 3 representing deletion, normal and amplification statuses of genes respectively. For this data, the graphical appearances of the independence model and that with gene spatial dependence have no significant differences, as in figures 1 and 2. Each dot in the figures roughly represents two events. Some minor difference can be observed in the region of gene loci 5000–7500, or panel 3 in each figure. Based on our estimates from the dependence model with the HBD method as displayed in Figure 2, more amplifications occurred in the 1–2500 gene loci region, with about 40 amplification events, where deletion is relatively less heavy. Deletions are heavy on loci 17500–21562, but very few amplifications observed in this region. Amplification is sparse in the 2200–2300 region, with about only four amplifications occurred, and with about eight of them occurred in the region 2500–5000. As pointed out before, although the number of amplifications is small compared to that of deletions, the average scale, i.e., the magnitude of copy number changes in each event of the former is much bigger than that of the latter, which is not shown in the figures. Our consensus is that in the analysis of gene copy number variations, along with the proportions of amplifications and deletions, the associated magnitude should also be reported, which is often missing in many such studies. The combined analysis should be more informative than just reporting the number of gains/losses. According to our results on this given normal individual, the average magnitude of gain is about 4.4 times that of losses. If we consider the number of gene alteration of some status times the corresponding average magnitude of mean change (in absolute value) as an index to describe the effect of copy number changes, for the individual under investigation, this index in the deletion status is about $5088 \times 0.497 \approx 2529$, while that for amplification is about $40 \times 2.185 \approx 80$. Thus it seems that in normal individual, at least for this one under investigation, most gene alterations are in the form of copy number losses, this is in contrast to some other reports in cancer patients (Nymark et al. 2006), in which the proportion of copy number gains is significantly higher than that of losses. As this individual is non-affected, these copy number changes reflect the result of cumulative changes over the generations during the human history.

4. Brief Discussion

We considered a Bayesian-frequentist hybrid method to analyze the CGH data, in which the mean parameters are estimated Bayesian and the other parameters frequentist. The method is used to classify a real CGH data, we find that for the individual analyzed, a moderate proportion of gene copy number deletion is observed, while relatively sparse loci with copy number amplification, and around 94.9% of the genes are in normal status, for this normal individual.

The method can be easily extended to multivariate CGH observations, and the spatial dependence specification can be implemented for neighborhood of arbitrary size.

As one reviewer point out, our situation is similar to a hierarchical setting, in which a prior dataset D_0 has information on part of the parameters α but not β , and a main dataset D

which has information on both α and β . Thus, parameters can be inferred from the joint likelihood $P(D, D0|\alpha, \beta)$. We think this approach has both pros and cons. The advantage is that it uses both data set so that more information in the data is used. But as D and $D0$ has different format, this increases the modeling complexity, and the corresponding results may not directly comparable to other methods based on $P(D|\alpha, \beta)$. However this modeling is of practical interest and will be the among topics in our future research.

Acknowledgments

This work is supported in part by the National Center for Research Resources at NIH grant 2G12RR003048, and by the Center for Research on Genomics and Global Health (CRGGH) at NHGRI/NIH.

References

- Anderson, TW. An Introduction to Multivariate Statistical Analysis. 3. John Wiley & Sons, Inc; Hoboken, New Jersey: 2003.
- Barash Y, Friedman N. Context-specific Bayesian clustering for gene expression data. *Journal of Computational Biology*. 2002; 9:169–191. [PubMed: 12015876]
- Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991; 43:1–59.
- Bickel PJ, Yahav JA. Some contributions to the asymptotic theory of Bayes solutions, *Z. Wahrscheinlichkeitstheorie view. Geb.* 1969; 11:257–276.
- Blackwell, D.; Girschick, MA. *Theory of Games and Statistical Decisions*. Wiley; New York: 1954.
- Blackwell D, Bickel P. A Note on Bayes Estimates. *Annals of Mathematical Statistics*. 1967; 38:1907–1911.
- Broët P, Richardson S. Detection of gene copy number changes in CGH microarrays using a spatially correlated mixture model. *Bioinformatics*. 2006; 22:911–918. [PubMed: 16455750]
- Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli G, Rossi E, Ludovini V, Gregorc V, Toschi L, Franklin W, Crino L, Gazdar A, Buun P, Hirsch F. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *Journal of Clinical Oncology*. 2005; 23(22): 5007–5018. [PubMed: 16051952]
- Cheng C, Kimmel R, Neiman P, Zhao L. Array rank order regression analysis for the detection of gene copy-number changes in human cancer. *Genomics*. 2003; 82:122–129. [PubMed: 12837263]
- Daruwala RS, Rudar A, Ostrer H, Lucito R, Wigler M, Mishra B. A versatile statistical analysis algorithm to detect genome copy number variation. *Proc Natl Acad Sci USA*. 2004; 101:1629216297.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser B*. 1977; 39:1–38.
- Eilers PH, De Menezes RX. Quantile smoothing of array CGH data. *Bioinformatics*. 2004 Nov 30. Epub.
- Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C. Copy number variation: New insights in genome diversity. *Genome Research*. 2006; 10.1101/gr.3677206
- Fridlyand J, Snijders A, Pinkel D, Albertson DG, Jain A. Hidden Markov models approach to the analysis of array CGH data. *Special Genomic Issue on Multivariate Methods in Genomic Data Analysis*. 2004; 90(1):132–153.
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs R, Freedman B, Quinones M, Bamshad M, Murthy K, Rovin B, Bradley W, Clark R, Anderson S, O'Connell R, Agan B, Ahuja SS, Bologna R, Sen L, Dolan M, Ahuja SK. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005; 307(5714):1422. [PubMed: 15746416]
- Guha, S.; Li, Y.; Neuberg, DD. Harvard Univ Biostatistics Working Paper Series-Working Paper 24. 2008. Bayesian hidden Markov modeling of array CGH data.

- Huang T, Wu B, Lizardi P, Zhao H. Detection of DNA copy number alterations using penalized least squares regression. *Bioinformatics*. 2005; 21(20):3811–3817. [PubMed: 16131523]
- Hupé P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*. 2004; 20(18):341–322.
- Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Nowak N, Albertson DG, Pinkel D, Collins C, Hanahan D, Gray JW. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nature Genetics*. 2001; 29:459–464. [PubMed: 11694878]
- Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics*. 2004; 20:3636–3637. [PubMed: 15201182]
- Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*. 1992; 258(5083):818–821. [PubMed: 1359641]
- LeCam, L.; Yang, G. *Asymptotics in Statistics: some basic concepts*. Springer-Verlag; New York: 2000.
- Lingjaerde OC, Baumbush LO, Liestol K, Glad IK, Borresen-Dale AL. CGH-explorer: a program for analysis of array-CGH data. *Bioinformatics*. 2005; 21(6):821–822. [PubMed: 15531610]
- Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*. 1993; 80:267–278.
- Myers CL, Dunham ML, Kung SY, Troyanskaya OG. Accurate detection of aneuploidies in array CGH and gene expression microarray data. *Bioinformatics*. 2004; 20(18):3533–3543. [PubMed: 15284100]
- Nymark P, Wikman H, Rousaari S, Hollmen J, Vanhala E, Karjalainen A, Anttila S, Knuutila S. Identification of specific gene copy number changes in asbestos-related lung cancer. *Cancer Research*. 2006; 66:5737–5743. [PubMed: 16740712]
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004; 5(4):557–572. [PubMed: 15475419]
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics*. 1998; 20(2): 207–211. [PubMed: 9771718]
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown P. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA*. 2002; 99:12963–12968. [PubMed: 12297621]
- Pugh TJ, Delaney AD, Farnoud N, Flibotte S, Griffith M, Li HI, Qian H, Farinha P, Gascoyne RD, Marra MA. Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acids Research*. 2008; 36(13):e80. [PubMed: 18559357]
- Rueda OM, Díaz-Uriarte R. Flexible and accurate detection of genomic copy-number changes. *PLoS Computational Biology*. 2007; 3(6):e122.10.1371/journal.pcbi.0030122 [PubMed: 17590078]
- Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genetics*. 2001; 29(3):263–264. [PubMed: 11687795]
- Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*. 1997; 20(4):399–407. [PubMed: 9408757]
- Wang J, Meza-Zepeda LA, Kresse SH, Myklebost O. M-CGH: analyzing microarray-based CGH experiments. *BMC Bioinformatics*. 2004; 74:1–4.
- Wang P, Kim Y, Pollack J, Balasubramanian N, Tibshirani R. A method for calling gains and losses in array CGH data. *Biostatistics*. 2005; 6(1):45–58. [PubMed: 15618527]
- Yuan A. Bayesian-frequentist hybrid inference. *Annals of Statistics*. 2009; 37:2458–2501.
- Yuan A, Chen G, Zhou Z, Bonney G, Rotimi C. Gene copy number analysis for family data using semiparametric copula model. *Bioinformatics and Biology Insights*. 2008; 2:349361.

Appendix

Proof of Proposition 2

Let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})$, (3) can be formulated as $\boldsymbol{\theta}^{(k+1)} = \pi(\boldsymbol{\theta}^{(k)})$ ($k = 0, 1, \dots$) for some function $\pi(\boldsymbol{\theta}^{(k+1)}) = (T_1(\boldsymbol{\beta}^{(k)}), T_2(\boldsymbol{\alpha}^{(k)}))$ determined by (3). Apparently, the convergence of one of the sequences $\{\boldsymbol{\alpha}^{(k)}\}$ and $\{\boldsymbol{\beta}^{(k)}\}$ implies the other. So we only need to prove the convergence of $\{\boldsymbol{\alpha}^{(k)}\}$. Denote $\dot{T} = (\dot{T}_1, \dot{T}_2)$ the derivative matrix of T , we have for some $\tilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}^{(k+m)}$ and $\boldsymbol{\theta}^{(k)}$,

$$\begin{aligned} \|\alpha^{(k+m)} - \alpha^{(k)}\| &= \|\dot{T}_1(\tilde{\boldsymbol{\beta}})(\beta^{(k-1+m)} - \beta^{(k-1)})\| = \|\dot{T}_1(\tilde{\boldsymbol{\beta}})\dot{T}_2(\tilde{\boldsymbol{\alpha}})(\alpha^{(k-1+m)} - \alpha^{(k-1)})\| \\ &\leq \sup_{\boldsymbol{\theta} \in A} \|\dot{T}_1(\boldsymbol{\beta})\dot{T}_2(\boldsymbol{\alpha})\| \cdot \|\alpha^{(k-1+m)} - \alpha^{(k-1)}\| \leq \dots \leq C^k \|\alpha^{(m)} - \alpha^{(0)}\| \\ &\leq C^k \sum_{j=0}^{m-1} \|\alpha^{(j+1)} - \alpha^{(j)}\| \leq C^k \sum_{j=0}^{m-1} C^j \|\alpha^{(1)} - \alpha^{(0)}\|, \quad (k=1, 2, \dots; m=1, 2, \dots) \end{aligned}$$

where $C = \sup_{\boldsymbol{\theta} \in A} \|\dot{T}_1(\boldsymbol{\beta})\dot{T}_2(\boldsymbol{\alpha})\|$. So if we show $C < 1$, then

$$C^k \sum_{j=0}^{m-1} C^j \|\alpha^{(1)} - \alpha^{(0)}\| \leq \frac{C^k}{1-C} \|\alpha^{(1)} - \alpha^{(0)}\|,$$

thus $\{\alpha^{(k)}\}$ is a Cauchy sequence and the conclusion will be true.

In fact, for fixed $\boldsymbol{\beta}$, by Theorem 2 in Yuan (2009) and the given condition, for large n ,

$$\dot{T}_1(\boldsymbol{\beta}) \approx \frac{\partial G(N(\hat{\alpha}_n, I^{-1}(\alpha|\boldsymbol{\beta})/n))}{\partial \alpha}.$$

Also, $\hat{\boldsymbol{\beta}}$ is determined by the equation $0 = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{x}_n)/\boldsymbol{\beta}$, so we have

$$0 = \frac{\partial^2}{\partial \beta \partial \alpha'} \ell(\alpha, \beta|\mathbf{x}_n) + \frac{\partial^2}{\partial \beta \partial \beta'} \ell(\alpha, \beta|\mathbf{x}_n) \frac{\partial \beta}{\partial \alpha'},$$

or, by the SLLN,

$$\dot{T}_2(\boldsymbol{\alpha}) = -\frac{\partial \beta}{\partial \alpha} = -\left(\frac{\partial^2}{\partial \alpha \partial \beta'} \ell(\alpha, \beta|\mathbf{x}_n) \right) \left(\frac{\partial^2}{\partial \beta \partial \beta'} \ell(\alpha, \beta|\mathbf{x}_n) \right)^{-1} \approx -I_{12}(\alpha, \beta) I_{22}^{-1}(\alpha, \beta),$$

thus by the given condition, $C < 1$ and the conclusion is true.

Proof of (5), (I), (I') and (IV)

The proof is for the general case where the observation is a sequence of vectors $\mathbf{x}_n = (x_1, \dots, x_n)$, the corresponding density if $\varphi(x_i|\alpha_j, \boldsymbol{\Omega}_j^{-1})$. Let $\boldsymbol{\Lambda}_j = \boldsymbol{\Omega}_j^{-1}$, and $d = \dim(\boldsymbol{\alpha})$. For (5), we have

$$\begin{aligned}
l(\alpha, \beta | \mathbf{y}_n) &= \sum_{i=1}^n \sum_{j=1}^3 I_{ij} (\log \gamma_j + \log \varphi(x_i | \alpha_j, \boldsymbol{\Omega}_j^{-1})) + \log \pi(\alpha) \\
&= \sum_{i=1}^n \sum_{j=1}^k I_{ij} \left(\log \gamma_j - \frac{1}{2} (x_i - \alpha_j)' \boldsymbol{\Lambda}_j (x_i - \alpha_j) + \frac{1}{2} \sum_{j=1}^k \log |\boldsymbol{\Lambda}_j| - \frac{d}{2} \log(2\pi) \right) \\
&\quad + \sum_{j=1}^k \left(-\frac{1}{2} (\alpha_j - \alpha_{j0})' \boldsymbol{\Lambda}_j (\alpha_j - \alpha_{j0}) + \frac{1}{2} \log |\boldsymbol{\Lambda}_j| - \frac{d}{2} \log(2\pi) \right).
\end{aligned}$$

To get $\check{\boldsymbol{\alpha}}_{nj}$, take partial derivative of $l(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}_n)$ with respect to $\boldsymbol{\alpha}_{nj}$, note $[(x_i - \boldsymbol{\alpha}_j)' \boldsymbol{\Lambda}_j (x_i - \boldsymbol{\alpha}_j)]' / \boldsymbol{\alpha}_j = -2\boldsymbol{\Lambda}_j (x_i - \boldsymbol{\alpha}_j)$ and $[(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j0})' \boldsymbol{\Lambda}_j (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j0})]' / \boldsymbol{\alpha}_j = 2\boldsymbol{\Lambda}_j (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j0})$. We have

$$\mathbf{0} = \frac{\partial l(\alpha, \beta | \mathbf{y}_n)}{\partial \alpha_j} = \sum_{i=1}^n I_{ij} \boldsymbol{\Lambda}_j (x_i - \alpha_j) - \boldsymbol{\Lambda}_j (\alpha_j - \alpha_{j0})$$

or, since $\boldsymbol{\Lambda}_j$ is non-singular, equivalently

$$0 = \sum_{i=1}^n I_{ij} x_i + \alpha_{j0} - (n_j + 1) \alpha_j,$$

This gives the expression for $\check{\boldsymbol{\alpha}}_{nj}$.

To get $\hat{\boldsymbol{\Omega}}_j$, take partial derivative of $l(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}_n)$ with respect to $\boldsymbol{\Lambda}_j$, note $(x_i - \boldsymbol{\alpha}_j)' \boldsymbol{\Lambda}_j (x_i - \boldsymbol{\alpha}_j) / \boldsymbol{\Lambda}_j = (x_i - \boldsymbol{\alpha}_j)(x_i - \boldsymbol{\alpha}_j)'$ and $\partial \log |\boldsymbol{\Lambda}_j| / \partial \boldsymbol{\Lambda}_j = \boldsymbol{\Lambda}_j^* / |\boldsymbol{\Lambda}_j|$, where $\boldsymbol{\Lambda}_j^*$ is the conjugate of $\boldsymbol{\Lambda}_j$ in the sense $\boldsymbol{\Lambda}_j^{-1} = \boldsymbol{\Lambda}_j^* / |\boldsymbol{\Lambda}_j| = \boldsymbol{\Omega}_j$. We have

$$\mathbf{0} = 2 \frac{\partial l(\alpha, \beta | \mathbf{y}_n)}{\partial \boldsymbol{\Lambda}_j} = - \sum_{i=1}^n I_{ij} (x_i - \alpha_j) (x_i - \alpha_j)' - (\alpha_j - \alpha_{j0}) (\alpha_j - \alpha_{j0})' + (n_j + 1) \boldsymbol{\Omega}_j,$$

plugging in $\check{\boldsymbol{\alpha}}_{nj}$ for $\boldsymbol{\alpha}_j$, this gives (5).

In the above, if $\pi_j(\cdot) = \varphi(\cdot | \boldsymbol{\alpha}_{j0}, \boldsymbol{\Omega}_{j0})$, with $\boldsymbol{\Omega}_{j0}$ known, then $(\check{\boldsymbol{\alpha}}_j, \hat{\boldsymbol{\Omega}}_j)$ is the solution of the equations

$$\begin{cases} \sum_{i=1}^n I_{ij} (x_i - \alpha_j) - \boldsymbol{\Omega}_j \boldsymbol{\Omega}_{j0}^{-1} (\alpha_j - \alpha_{j0}) = \mathbf{0} \\ \sum_{i=1}^n I_{ij} (x_i - \alpha_j) (x_i - \alpha_j)' - n_j \boldsymbol{\Omega}_j = \mathbf{0} \end{cases},$$

which can be obtained by the Newton-Raphson method.

To derive (I), note $P(I_{ij} = 1 | \beta^{(r)}) = \gamma_j^{(r)}$ ($i = 1, \dots, n; j = 1, 2, 3$), so

$$E(I_{ij}|\mathbf{x}_n, \alpha^{(r)}, \beta^{(r)}) = P(I_{ij}=1|\mathbf{x}_n, \alpha^{(r)}, \beta^{(r)}) = P(I_{ij}=1|x_i, \alpha^{(r)}, \beta^{(r)}) \\ = \frac{\varphi(x_i|\alpha_j^{(r)}, \boldsymbol{\Omega}_j^{(r)})P(I_{ij}=1|\beta^{(r)})}{\sum_{l=1}^3 \varphi(x_i|\alpha_l^{(r)}, \boldsymbol{\Omega}_l^{(r)})P(I_{il}=1|\beta^{(r)})} = \frac{\gamma_j^{(r)} \varphi(x_i|\alpha_j^{(r)}, \boldsymbol{\Omega}_j^{(r)})}{\sum_{l=1}^3 \gamma_l^{(r)} \varphi(x_i|\alpha_l^{(r)}, \boldsymbol{\Omega}_l^{(r)})}.$$

The derivation of (I') is similar under the conditional probability for I_{ij} given in (6), and noting in this case $E(I_{ij}|\mathbf{x}_n, \alpha^{(r)}, \beta^{(r)}) = P(I_{ij}=1|R_i, \alpha^{(r)}, \beta^{(r)})$.

For (IV), note the maximization over the γ_j 's are under the constraint $\sum_{j=1}^3 \gamma_j = 1$. So by the Lagrange method, we are to maximize $l(\alpha, \beta|\mathbf{y}_n) - r \sum_{j=1}^3 \gamma_j$. For $j = 1, 2, 3$, set

$$0 = \frac{\partial}{\partial \gamma_j} \left(l(\alpha, \beta|\mathbf{y}_n) - r \sum_{j=1}^3 \gamma_j \right) = \sum_{i=1}^n I_{ij} / \gamma_j - r$$

we get $\sum_{i=1}^n I_{ij} = r \gamma_j$. Summing over j on both sides we get $n = \sum_{i=1}^n \sum_{j=1}^3 I_{ij} = r$, thus

$$\widehat{\lambda}_j = \sum_{i=1}^n I_{ij} / r = n_j / n.$$

Proof of Proposition 3

Define

$$b(\mathbf{y}_n|\theta) = \pi(\alpha) \prod_{i=1}^n \prod_{j=1}^k (\gamma_j \varphi(x_i|\alpha_j, \boldsymbol{\Omega}_j))^{I_{ij}}, \quad a(\mathbf{x}_n|\theta) = \pi(\alpha) \prod_{i=1}^n \sum_{j=1}^k \gamma_j \varphi(x_i|\alpha_j, \boldsymbol{\Omega}_j),$$

and define $Q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = E[\log b(\mathbf{y}_n|\boldsymbol{\theta}')|\mathbf{x}_n, \boldsymbol{\theta}]$, and $H(\boldsymbol{\theta}'|\boldsymbol{\theta}) = E[\log g(\mathbf{y}_n|\boldsymbol{\theta}')|\mathbf{x}_n, \boldsymbol{\theta}]$, where $g(\mathbf{y}_n|\boldsymbol{\theta}) = b(\mathbf{y}_n|\boldsymbol{\theta})/a(\mathbf{x}_n|\boldsymbol{\theta})$, the expectations are over the missing data I_{ij} 's. Then $\log S(\boldsymbol{\alpha}, |\mathbf{x}_n, \boldsymbol{\theta}') = \log a(\mathbf{x}_n|\boldsymbol{\theta}') = Q(\boldsymbol{\theta}'|\boldsymbol{\theta}) - H(\boldsymbol{\theta}'|\boldsymbol{\theta})$. It is seen that $g(\mathbf{y}_n|\boldsymbol{\theta})$ is just the conditional density of $\mathbf{y}_n$ given $\mathbf{x}_n$, thus $Q(\cdot|\cdot)$ and $H(\cdot|\cdot)$ here play the same roles as they do in the standard EM algorithm (Dempster et al., 1977). Since $\boldsymbol{\theta}^{(r+1)} = \arg \sup_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$, we have

$$Q(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) \geq Q(\boldsymbol{\theta}^{(r)}|\boldsymbol{\theta}^{(r)}).$$

Also by Lemma 1 in Dempster et al (1977), or just using the property of relative entropy,

$$H(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) \leq H(\boldsymbol{\theta}^{(r)}|\boldsymbol{\theta}^{(r)}),$$

so we have

$$\begin{aligned}\log\pi(\alpha^{(r+1)}|\mathbf{x}_n, \beta^{(r+1)}) &= \log a(\mathbf{x}_n|\theta^{(r+1)}) = Q(\theta^{(r+1)}|\theta^{(r)}) - H(\theta^{(r+1)}|\theta^{(r)}) \\ &\geq Q(\theta^{(r)}|\theta^{(r)}) - H(\theta^{(r)}|\theta^{(r)}) = \log a(\mathbf{x}_n|\theta^{(r)}) = \log S(\alpha^{(r)}|\mathbf{x}_n, \beta^{(r)}),\end{aligned}$$

i.e. Proposition 3 is true.

Figure 1.
Classification results from independence model: 1. GCN deletion; 2. Normal; 3. Amplification.

Figure 2.
Classification results from dependence model: 1. GCN deletion; 2. Normal; 3. Amplification.

Table 1

Estimation by MLE, BYE and HBD: Independence Model

n	Method	γ_1	γ_2	γ_3	α_1	α_2	α_3	σ_1^2	σ_2^2	σ_3^2
300	MLE	0.128 (0.147)	0.854 (0.062)	0.018 (0.372)	0.051 (0.211)	0.855 (0.141)	3.140 (0.003)	0.006 (15.479)	0.122 (0.305)	5.225 (0.059)
	BYE	0.015 (0.066)	0.970 (0.058)	0.015 (0.409)	0.519 (0.002)	0.757 (0.127)	3.520 (0.003)	0.192 (0.136)	0.188 (0.182)	4.689 (0.070)
	HBD	0.256 (0.071)	0.726 (0.060)	0.018 (0.364)	0.663 (0.064)	1.002 (0.100)	3.005 (0.003)	0.060 (0.634)	0.193 (0.187)	4.531 (0.085)
1000	MLE	0.083 (0.090)	0.911 (0.033)	0.006 (0.320)	0.085 (0.078)	0.839 (0.075)	2.965 (0.001)	0.006 (6.756)	0.140 (0.154)	4.881 (0.065)
	BYE	0.116 (0.075)	0.877 (0.033)	0.006 (0.318)	0.125 (0.051)	0.863 (0.074)	3.087 (0.001)	0.019 (2.094)	0.129 (0.161)	4.087 (0.085)
	HBD	0.133 (0.071)	0.860 (0.033)	0.007 (0.284)	0.146 (0.050)	0.873 (0.075)	2.698 (0.001)	0.024 (1.541)	0.122 (0.171)	3.797 (0.102)
10000	MLE	0.122 (0.021)	0.873 (0.010)	0.005 (0.120)	0.167 (0.020)	0.770 (0.024)	3.192 (0.000)	0.012 (0.858)	0.125 (0.056)	5.246 (0.025)
	BYE	0.128 (0.021)	0.867 (0.010)	0.005 (0.120)	0.173 (0.020)	0.773 (0.024)	3.211 (0.000)	0.013 (0.755)	0.124 (0.057)	5.138 (0.026)
	HBD	0.125 (0.021)	0.870 (0.010)	0.005 (0.120)	0.170 (0.020)	0.772 (0.024)	3.209 (0.000)	0.013 (0.755)	0.124 (0.057)	5.132 (0.026)
21562	MLE	0.233 (0.011)	0.761 (0.007)	0.006 (0.071)	0.200 (0.018)	0.701 (0.015)	2.876 (0.000)	0.016 (0.338)	0.115 (0.045)	4.966 (0.019)
	BYE	0.235 (0.011)	0.759 (0.007)	0.006 (0.071)	0.201 (0.017)	0.702 (0.015)	2.892 (0.000)	0.016 (0.329)	0.115 (0.045)	4.941 (0.019)
	HBD	0.236 (0.011)	0.758 (0.007)	0.006 (0.071)	0.202 (0.017)	0.702 (0.015)	2.869 (0.000)	0.016 (0.327)	0.115 (0.045)	4.924 (0.019)

Table 2

Estimation by MLE, BYE and HBD: Dependence Model

n	Method	γ_1	γ_2	γ_3	α_1	α_2	α_3	σ_1^2	σ_2^2	σ_3^2
300	MLE	0.139 (0.141)	0.843 (0.062)	0.018 (0.374)	0.058 (0.147)	0.864 (0.143)	3.159 (0.003)	0.011 (7.008)	0.117 (0.314)	5.140 (0.061)
	BYE	0.150 (0.131)	0.833 (0.062)	0.017 (0.382)	0.087 (0.087)	0.874 (0.132)	3.176 (0.003)	0.023 (2.585)	0.122 (0.288)	4.440 (0.083)
	HBD	0.139 (0.141)	0.843 (0.062)	0.018 (0.378)	0.064 (0.143)	0.868 (0.138)	3.158 (0.004)	0.012 (6.658)	0.120 (0.297)	4.337 (0.087)
1000	MLE	0.111 (0.078)	0.883 (0.033)	0.007 (0.303)	0.112 (0.056)	0.859 (0.075)	2.857 (0.001)	0.016 (2.679)	0.129 (0.163)	4.576 (0.072)
	BYE	0.126 (0.072)	0.867 (0.033)	0.007 (0.304)	0.138 (0.047)	0.869 (0.074)	2.886 (0.001)	0.024 (1.529)	0.127 (0.163)	4.013 (0.093)
	HBD	0.110 (0.078)	0.883 (0.033)	0.007 (0.305)	0.114 (0.056)	0.860 (0.074)	2.890 (0.001)	0.015 (2.688)	0.130 (0.160)	3.971 (0.095)
10000	MLE	0.143 (0.020)	0.853 (0.010)	0.005 (0.119)	0.188 (0.018)	0.781 (0.024)	3.167 (0.000)	0.017 (0.558)	0.122 (0.058)	5.210 (0.026)
	BYE	0.146 (0.020)	0.850 (0.010)	0.005 (0.119)	0.191 (0.018)	0.782 (0.023)	3.162 (0.000)	0.018 (0.524)	0.122 (0.058)	5.108 (0.027)
	HBD	0.142 (0.020)	0.853 (0.010)	0.005 (0.119)	0.188 (0.018)	0.781 (0.024)	3.162 (0.000)	0.017 (0.560)	0.122 (0.058)	5.097 (0.027)
21562	MLE	0.258 (0.011)	0.737 (0.007)	0.006 (0.072)	0.215 (0.017)	0.712 (0.015)	2.895 (0.000)	0.020 (0.259)	0.114 (0.046)	5.013 (0.018)
	BYE	0.259 (0.011)	0.735 (0.007)	0.006 (0.072)	0.216 (0.017)	0.713 (0.015)	2.900 (0.000)	0.020 (0.255)	0.114 (0.046)	4.984 (0.019)
	HBD	0.258 (0.011)	0.737 (0.007)	0.006 (0.072)	0.215 (0.017)	0.712 (0.015)	2.897 (0.000)	0.020 (0.259)	0.114 (0.046)	4.974 (0.019)