Generalized Beta Convolution Model of the True Intensity for the Illumina Bead Arrays

Rohmatul Fajriyah [a, b]
[a] Institute of Statistics, TU Graz, 8010 Graz, Austria.
[b] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Jogjakarta, 55584, Indonesia.
Corresponding author; e-mail: fajriyah@student.tugraz.ac.at, rfajriyah@fmipa.uii.ac.id

Received: 16 June 2014
Accepted: 30 January 2015

Abstract

Microarray data, which come from many steps of production, have been known to contain noise. The pre-processing is implemented to reduce the noise, where the background is corrected. Prior to further analysis, many IlluminaBeadArrays users had applied the convolution model, a model which had been adapted from when it was first developed on the Affymetrix platform, to adjust the intensity value: corrected background intensity value.

Several models based on the different underlying distributions and or the parameters estimation methods have been proposed and applied. For instance: the exponential-gamma, the normal-gamma, and the exponential-normal convolutions with maximum likelihood estimation, non-parametric, Bayesian and moment methods of the parameters estimation, including two recent exponential-lognormal and gamma-lognormal convolutions.

In this paper, we propose models and derive the corrected background intensity based on the generalized betas and the generalized beta-normal convolutions as a generalization of the existing models.
Keywords: background correction, additive error, generalized beta distribution family, IlluminaBeadArrays and convolution model.

1. Introduction

It has become common knowledge that data from microarray experiments will contain some non-biological noise. Therefore, the data needs to be adjusted. In this case, implementing the pre-processing will adjust [1-3] or correct the background intensity value.

There are several steps in pre-processing where one of the steps is the background correction. In the background correction, the noise can be modeled as additive or multiplicative (See, Huber et al. [1,2], Bolstad et al. [4], Irizarry et al. [5-7], Li and Wong [8], Silver et al. [9] and Wu et al. [10].

In the robust multi-array average (RMA), Irizarry et al. [5-7] have modeled the noise as an additive, to adjust the intensity value. Although the RMA was developed for the Affymetrix platform initially, it was also been used for the data from the Illumina platform.

Currently, there are some models to correct the intensity value of the Illumina platform available, for instance: the model-based background correction method (MBCB) from Ding, et al. [11] and Xie, et al. [12], the exponential-gamma from Chen et al. [13], the gamma-normal from Plancade et al. [14] and the exponential(gamma)-lognormal from Fajriyah [15].

Posekany’s et al. study [16], show us that by using the Affymetrix and the Invitrogen platforms the noises in microarray data is not Gaussian but far more heavy-tailed. On the other hand, Chen et al. [13] show that the noise distribution in the Illumina platform is usually skewed in different degrees.

Therefore, while the intensity values are widely accepted as a skewed distribution, the noise distribution could possibly symmetrical or skewed. Note that in this paper, noise and intensity mean the negative control probes intensity and the observed probes intensity values respectively.

McDonald and Xu [17] have introduced a distribution tree of generalized beta distribution, which is used to model the income distribution. It is similar in nature to the microarray data where the random variable is a non-negative value. This distribution tree helps us to understand the relationship among the available distributions. Moreover, quite recently, Leemis and McQueston [18] have explained the relationships among the univariate distributions in statistics. See the distribution tree from McDonald and Xu [17].
This paper aims to present the estimator of the true intensity value, the corrected background intensity, where the noise is symmetric and skewed distribution. If the noise is skewed distribution, the underlying distributions of the proposed convolution model are the generalized beta distributions, a generalized model of the existing ones. If the noise is a symmetrically distributed, the proposed model is generalized beta-normal convolution, which is a generalized model of the model of Plancade et al. [14].

In general, the background correction is applied toward each array, where in each array there are probes (perfect match and mismatch probes), probesets and genes (terminology for the Affymetrix platform) or bead and bead-type level probes (terminology for the Affymetrix platform).

The current publicly available benchmarking data set for the Illumina platform is the raw data from the bead studio, which is the average of the bead-type level probes, not corrected background and of unnormalized intensity. Therefore, the background correction in this paper is applied to the gene (bead-type level probes) intensity in each array.

Suppose we have $J$ arrays and for each array there are $I$ regular genes and $W$ negative control genes. Throughout the paper, the convolution model is applied for each array $j$ and represented as follows:

Figure 1. Distribution tree, McDonald and Xu [17].
\[ P_i = S_i + B_i \]  \hspace{1cm} (1)

where \( P_i \), \( S_i \) and \( B_i \) are the regular (observed) true/corrected background and noise intensity values respectively of the \( i \)th gene, \( i = 1, 2, \ldots, I \). For a negative control gene \( w \) at array \( j \), \( w = 1, 2, \ldots, W \), the observed intensity, denoted by \( P_{ow} \), is assumed to be \( P_{ow} = B_{ow} \) is the noise intensity. \( P_i \) and \( P_{ow} \) are assumed to be independent.

This paper is organized as follows: Section 2 reviews previous work related to the background correction for the IlluminaBeadArrays, Section 3 explains the results of our investigation and Section 4 provides discussion and remarks.

2. Previous work

2.1. Basic concepts

Definition 2.1. Suppose \( X \) is a random variable of generalized beta distribution. McDonald and Xu [17] define the probability function of the generalized beta distribution as follows:

\[
X \sim GB(x; a, c, d, u, v) = \frac{[u]^{a+c-1} \left(1 - (1-c)(\frac{u}{v})^a\right)^{d-1}}{\beta(a+c)(1+c)} \left(\frac{u}{v}\right)^{a-d} \left(1 + \frac{u}{v}\right)^{a-d} \]  \hspace{1cm} (2)

\( B(u, v) \) is the beta function, \( 0 \leq c \leq 1, a, d, u \) and \( v \) positive.

Definition 2.2. Let \( X \) and \( Y \) be two continuous random variables with density functions \( f_1(x) \) and \( f_2(y) \) respectively. Assume that both \( f_1(x) \) and \( f_2(y) \) are defined for all real numbers. Then the convolution \( f_1(x) \) and \( f_1 \ast f_2 \) of \( f_1 \) and \( f_2 \) is the function given by

\[
(f_1 \ast f_2)(x) = \int_{-\infty}^{\infty} f_1(z - y) f_2(y) \, dy = \int_{-\infty}^{\infty} f_1(x) f_2(z - x) \, dx \]  \hspace{1cm} (3)

Theorem 2.1. Let \( X \) and \( Y \) be two independent random variables with density function \( f_2(x) \) and \( f_2(y) \) defined for all \( x \) and \( y \). The sum \( Z = X + Y \) is a random variable with a density function of \( f_2(z) \), where \( f_2(z) \) is the convolution of \( f_X \) and \( f_Y \).

2.2. Background correction by RMA

In the RMA model ([4] and [5-7]), it is assumed that the intensity values are affected by the noises of the chip. The RMA model is as in the Equation (1), where \( P_i = \)
PM<sub>i</sub> is the observed probe level intensity of perfect match probes of the \(i^{th}\) gene, \(S_i\) is the true intensity of the \(i^{th}\) gene, with \(S_i \sim f_i(s; \theta) = \text{Exp}(\theta), \theta > 0\), and \(B_i\) is the background noise of the \(i^{th}\) gene with \(f_2 = N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\).

Assuming independence, the joint density of the two-dimensional random variables \((S_i, B_i)\) is:

\[
f_{S_i, B_i}(s_i, b_i; \mu_j, \sigma_j^2, \theta_j) = \theta_j e^{-\theta_j b_i} f_2(b_i; \mu_j, \sigma_j^2), s_i, b_i > 0
\]

Furthermore, the transformation formula for two-dimensional densities gives the joint density of \(S_i\) and \(R_i\) is

\[
f_{S_i, R_i}(s_i, p_i; \mu_j, \sigma_j^2, \theta_j) = \theta_j e^{-\theta_j (p_i - \mu_j)} f_2(s_i; p_i - \mu_j - \sigma_j^2 \theta_j, \sigma_j^2), 0 < s_i < p_i
\]

From equation (4) we get the marginal density of \(P_i\) and the conditional density of \(S_i\) given \(P_i\) in equations (5) and (6) below, respectively:

\[
f_{P_i}(p_i) = \theta_j e^{\frac{\sigma_j^2}{2}(p_i - \mu_j)} \left( \Phi \left( \frac{\mu_j - p_i}{\sigma_j} \right) + \Phi \left( \frac{p_i - \mu_j}{\sigma_j} - 1 \right) \right)
\]

\[
f_{S_i|P_i}(s_i|p_i) = \frac{f_{S_i,R_i}(s_i,p_i; \mu_j,\sigma_j^2)}{\Phi \left( \frac{\mu_j - p_i}{\sigma_j} \right) + \Phi \left( \frac{p_i - \mu_j}{\sigma_j} - 1 \right)}
\]

where \(\mu_{S,P,j} = p_i - \mu_j - \sigma_j^2 \theta_j\)

The corrected background intensity is computed by the conditional expectation

\[
E(S_i|P_i = p_i) = \frac{1}{\Phi \left( \frac{\mu_j - p_i}{\sigma_j} \right) + \Phi \left( \frac{p_i - \mu_j}{\sigma_j} - 1 \right)} \int_0^{p_i} f_2(s_i; \mu_{S,P,j}, \sigma_j^2) ds_i.
\]

The substitution \(s_i = \mu_{S,P,j} + \sigma_j \epsilon_i\), yields the corrected background intensity in the Equation (7) equal to

\[
\mu_{S,P,j} + \sigma_j \Phi \left( \frac{\mu_j - \mu_{S,P,j}}{\sigma_j} \right) - \Phi \left( \frac{p_i - \mu_{S,P,j}}{\sigma_j} \right) \frac{\Phi \left( \frac{p_i - \mu_j}{\sigma_j} - 1 \right)}{\Phi \left( \frac{\mu_j - p_i}{\sigma_j} \right) + \Phi \left( \frac{p_i - \mu_j}{\sigma_j} - 1 \right)}
\]
2.3. Exponential-normal MBCB

Xie et al. [12] use the same underlying distributions as the RMA for the background correction. The differences between the MBCB and the RMA ([4] and [5-7]) are

1. Xie et al. [12] take the infinite value for the upper bound of the integral to compute the marginal density function and the conditional expectation of the true intensity value. On the other hand, the RMA put \( p \) as the upper bound of the integral. The corrected background intensity of this model is

\[
\mu_{c,x,j} + \sigma_j \phi \left( \frac{x - \mu_j}{\sigma_j} \right)
\]

(9)

2. Under convolution model of (1), where the true intensity value is assumed exponentially distributed and the noise is normally distributed, we then need to estimate the parameters and \( \theta, \mu_j, \) and \( \sigma_j \). Xie et al. [12] offer three parameters estimation methods: the non-parametric, maximum likelihood and Bayesian. On the other hand, RMA apply the ad-hoc method.

Ding et al. [11] use the exponential-normal convolution model to correct the background of the Illumina platform by using a Markov chain Monte Carlo simulation.

2.4. Gamma-normal convolution

Plancade et al. [14] introduced the gamma-normal convolution to model the background correction of the IlluminaBeadArrays. The model is based on the RMA background correction of AffymetrixGeneChips. Plancade et al. [14] assume that the true intensity value is gamma distributed and the noise is normally distributed.

Under the model background correction in (1), \( f_y \) is the convolution product of \( f_yi \) and \( f_{yi} \). The true intensity \( S_i \) is computed by the conditional expectation of \( S_i \) given \( P_i = p_i \):

\[
E(S_i | P_i = p_i) = \tilde{S}_i(p_i) = \int \frac{f_x \cdot f_{xm} \cdot f_{nm}(p_i-x) \, dx}{\int f_x \cdot f_{xm} \cdot f_{nm}(p_i-x) \, dx}
\]

(10)

where \( f_{xm} = \left( x; \alpha_j, \theta_j \right) = \frac{\theta_j^{\alpha_j} x^{\alpha_j-1} e^{-\theta_j x}}{\Gamma(\alpha_j)}, \alpha_j, \theta_j, x_i > 0, \) is the gamma density.

When \( S_i \) is gamma distributed and \( B_i \) is normally distributed, then the equation (10) does not have analytic expression as it dies in Equations (8) and (9). Therefore, Plancade et al. [14] implemented the Fast Fourier Transform to estimate the parameters and to correct the background. For the background correction with Fast Fourier
Transform, Equation (10) is rewritten as

\[
\hat{S}_i(p|\Theta) = \frac{a_i \theta_i \int f_{\text{gamm}}(s_i|\alpha_i, \beta_i) f_{\text{gamm}}(p_i-s_i|\alpha, \beta) ds_i}{\int f_{\text{gamm}}(s_i|\alpha_i, \beta_i) f_{\text{gamm}}(p_i-s_i|\alpha, \beta) ds_i}
\]  (11)

where \(\Theta = (\mu, \sigma, \alpha, \beta)\) and \(s_f \text{gamm}(s_i) = a_i \beta_i f_{\text{gamm}}(s_i)\) is valid for every \(s_i > 0\).

2.5. Exponential-gamma convolution

Chen et al. [13] proposed in favor of the distribution of the true intensity and its noise, under the convolution model of Equation (1), the exponential and gamma distribution respectively. Therefore, \(S_i \sim f_1(s_i; \theta_j) = \text{Exp}(\theta_j)\), and \(B_i \sim f_2(b_i; \alpha, \beta_j) = \text{GAM}(\alpha_j, \beta_j)\), where \(s_i, b_i, \theta_j, \alpha_j, \beta_j > 0\).

The corrected background intensity for the proposed model ([13]) is

\[
p_i = \frac{e^{\theta_i s_i} e^{-\frac{(s_i)}{\theta_j}}}{\int_0^\infty e^{\theta_i s_i} e^{-\frac{(s_i)}{\theta_j}} ds_i} (12)
\]

2.6. Exponential-lognormal convolution, Fajriyah [15]

Under model (1), when the true intensity \(S_i\) is assumed to be exponentially distributed \(S_i \sim f_1(s_i; \theta_j) = \theta_j e^{-\theta_j s_i}; \theta_j, s_i > 0\), and the background noise \(B_i\) is assumed to be log normally distributed, \(B_i \sim f_2(b_i; \mu, \sigma_b) = e^{-\frac{(\ln b_i - \mu)^2}{2\sigma^2_b}}; \mu \in \mathbb{R}, \sigma^2_b, b_i > 0\), the corrected background intensity value is

\[
p_i = \frac{e^{\theta_i s_i^2/2 \sigma^2_b}}{C_{1,i}} (13)
\]

where

\[
C_{2,i} = \sum_{k=0}^\infty \frac{\theta_i^k e^{k \frac{(\ln b_i - \mu)^2}{\sigma^2_b}}}{k!} \phi \left( \frac{\ln b_i - (\mu + k\sigma^2_b)}{\sigma_b} \right), \text{ and } C_{1,i} = \sum_{k=0}^\infty \frac{\theta_i^k e^{k \frac{(\ln b_i - \mu)^2}{\sigma^2_b}}}{k!} \phi \left( \frac{\ln b_i - (\mu + k\sigma^2_b)}{\sigma_b} \right)
\]

2.7. Gamma-lognormal convolution, Fajriyah [15]

Under model (1), when the true intensity \(S_i\) is assumed to be gamma distributed \(S_i \sim f_1(s_i; \alpha_j, \beta_j) = \frac{e^{-\frac{s_i}{\beta_j}} s_i^{\alpha_j-1}}{\beta_j \Gamma(\alpha_j)}; \alpha_j, \beta_j, s_i > 0\) and the background noise \(B_i\) is assumed to be...
lognormally distributed, \( B_i \sim f_2(b_i; \mu_j, \sigma_j^2) = \frac{(\ln b_i - \mu_j)^2}{2\sigma_j^2} \); \( \mu_j \in \mathbb{R}, \sigma_j^2, b_i > 0 \), the corrected background intensity value is

\[
\frac{p_i C_{4,j}}{C_{3,j}}
\]

where

\[
C_{4,j} = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(-1)^k e^{(k+n)(\mu_j + (k+n)\sigma_j^2)}}{p_i^k \beta_j^n n!} \Phi \left( \frac{\ln p_i - (\mu_j + (k+n)\sigma_j^2)}{\sigma_j} \right)
\]

and

\[
C_{3,j} = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{(-1)^k (\alpha_j - 1) e^{(k+n)(\mu_j + (k+n)\sigma_j^2)}}{p_i^k \beta_j^n n!} \Phi \left( \frac{\ln p_i - (\mu_j + (k+n)\sigma_j^2)}{\sigma_j} \right)
\]


3. Results

In the subsequent sections, we will explain the generalized beta convolution model and its corrected background intensity value.

3.1. Generalized beta distribution convolution

3.1.1. The joint density function

Under the convolution model of Equation (1), where \( P_i \) is the observed intensity of regular probes of the \( i^{th} \) gene, \( S_i \) is the true intensity of the \( i^{th} \) gene, with

\[
S_i \sim f_1(s; a_{i,j}, c_{1,j}, d_{1,j}, u_{i,k,j}, v_{i,j}) = \frac{\left[ a_{i,j} \right]^{a_{i,j} u_{i,j} - 1} \left( 1 - c_{i,j} \right) \left( \frac{u_{i,j}}{d_{i,j}} \right)^{a_{i,j} u_{i,j} v_{i,j} - 1}}{d_{1,j}^{a_{i,j} u_{i,j}} B \left( u_{i,j}, v_{i,j} \right) \left( 1 + c_{i,j} \left( \frac{u_{i,j}}{d_{i,j}} \right) \right)^{a_{i,j} u_{i,j} v_{i,j}}} \\
0 \leq c_{i,j} \leq 1, a_{i,j}, d_{1,j}, u_{i,k,j} and v_{i,j} positive s_i > 0.
\]
And $B_i$ is the background noise with

$$B_i \sim f_2(b_i; a_{z,i}, d_{z,i}, u_{z,i}, v_{z,i}) = \frac{|a_{z,i}| b_i^{a_{z,i} u_{z,i} - 1} (1 - (1 - c_{z,i}) \left( \frac{b_i}{d_{z,i}} \right)^{a_{z,i} u_{z,i}})^{v_{z,i} - 1}}{d_{z,i}^{a_{z,i} u_{z,i} + B} (u_{z,i}, v_{z,i}) \left( 1 + c_{z,i} \left( \frac{b_i}{d_{z,i}} \right)^{a_{z,i} u_{z,i}} \right)^{v_{z,i} + v_{z,i}}}$$

$$0 \leq c_{z,i} \leq 1, a_{z,i}, d_{z,i}, u_{z,i} \text{ and } v_{z,i} \text{ positive } b_i > 0.$$

The joint density function of $S_i$ and $P_i$ is

$$f_{S_i,P_i}(s_i, p_i) = \frac{|a_{s,i}| s_i^{a_{s,i} u_{s,i} - 1} (1 - (1 - c_{s,i}) \left( \frac{s_i}{d_{s,i}} \right)^{a_{s,i} u_{s,i}})^{v_{s,i} - 1}}{d_{s,i}^{a_{s,i} u_{s,i} + B} (u_{s,i}, v_{s,i}) \left( 1 + c_{s,i} \left( \frac{s_i}{d_{s,i}} \right)^{a_{s,i} u_{s,i}} \right)^{v_{s,i} + v_{s,i}}}$$

$$\times \frac{|a_{z,i}| (p_i - s_i)^{a_{z,i} u_{z,i} - 1} (1 - (1 - c_{z,i}) \left( \frac{p_i - s_i}{d_{z,i}} \right)^{a_{z,i} u_{z,i}})^{v_{z,i} - 1}}{d_{z,i}^{a_{z,i} u_{z,i} + B} (u_{z,i}, v_{z,i}) \left( 1 + c_{z,i} \left( \frac{p_i - s_i}{d_{z,i}} \right)^{a_{z,i} u_{z,i}} \right)^{v_{z,i} + v_{z,i}}}$$

### 3.1.2 The marginal density function

The marginal density function of $P_i$ is

$$f_{P_i}(p_i) = \int_{0}^{p_i} f_{S_i,P_i}(s_i, p_i) \, ds_i$$

$$= K \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(-1)^{l+m+n+r} (1 - c_{s,i}) (1 - c_{z,i}) m^{m} e_{s,i}^{n} e_{z,i}^{r}}{d_{s,i}^{a_{s,i} u_{s,i} + B} d_{z,i}^{a_{z,i} u_{z,i} + B}}$$

$$\times \left( \frac{v_{s,i} + 1}{l} \right)^{v_{s,i} + 1} \left( \frac{v_{z,i} + 1}{m} \right)^{v_{z,i} + 1} \left( \frac{n + 1}{n} \right)^{(u_{s,i} + v_{s,i} + n - 1)(u_{z,i} + v_{z,i} + r - 1)}$$

$$\times \int_{0}^{p_i} s_i^{a_{s,i} u_{s,i} + n} (p_i - s_i)^{a_{s,i} u_{s,i} + m} \, ds_i$$

(15)

Let $\frac{s_i}{p_i} = z_i$, then the equation (15) becomes

$$K_i p_i^{a_{s,i} u_{s,i} + a_{z,i} u_{z,i} - 1} c_{z,i}$$

(16)
where

\[ K_1 = \frac{\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m+n} \sum_{l=0}^{m+n} (-1)^{i+m+n+r} (1 - c_{1,j}) (1 - c_{2,j})^m c_{1,j}^r c_{2,j}^r \times \left( \frac{u_{1,j} + v_{1,j} + n - 1}{m} \right) \left( \frac{u_{2,j} + v_{2,j} + r - 1}{n} \right) \times p_i a_{1,j}^{(i+n)+a_{2,j}(m+r)} B \left( a_{1,j}(u_{1,j} + l + n) - 1, a_{2,j}(u_{2,j} + m + r) - 1 \right) }{d_{1,j} a_{2,j} B \left( u_{1,j}, v_{1,j} \right) B \left( u_{2,j}, v_{2,j} \right)} \],

and

\[ C_{5,j} = \sum_{i=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} (-1)^{i+m+n+r} (1 - c_{1,j}) (1 - c_{2,j})^m c_{1,j}^r c_{2,j}^r \times \left( \frac{u_{1,j} - 1}{l} \right) \left( \frac{v_{2,j} - 1}{m} \right) \left( \frac{u_{1,j} + v_{1,j} + n - 1}{n} \right) \left( \frac{u_{2,j} + v_{2,j} + r - 1}{r} \right) \times p_i a_{1,j}^{(i+n)+a_{2,j}(m+r)} B \left( a_{1,j}(u_{1,j} + l + n) - 1, a_{2,j}(u_{2,j} + m + r) - 1 \right) \].

3.1.3. The conditional density function

The conditional density function of \( S_i \) where it is known that \( P_j = p_i \) is

\[ f_{s_i | p_i}(s_i | p_i) = \frac{f_{s_0 | p_0}(s_0 | p_0)}{f_{p_0}(p_0)} = s_i a_{1,i}^{u_{1,i} - 1} \left( 1 - (1 - c_{1,j}) \left( \frac{s_i}{u_{1,i}} \right) a_{1,j}^{u_{1,j} - 1} (p_i - s_i) a_{2,j}^{u_{2,j} - 1} \left( 1 - (1 - c_{2,j}) \left( \frac{p_i - s_i}{u_{2,j}} \right) a_{2,j}^{u_{2,j} - 1} \right) \right) \]

3.1.4. The corrected background intensity value

The corrected background intensity under this generalized beta convolution is

\[ \frac{C_{5,j}}{p_i} \]

where

\[ C_{6,j} = \sum_{i=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} (-1)^{i+m+n+r} (1 - c_{1,j}) (1 - c_{2,j})^m c_{1,j}^r c_{2,j}^r \times \left( \frac{u_{1,j} + v_{1,j} + n - 1}{n} \right) \left( \frac{u_{2,j} + v_{2,j} + r - 1}{r} \right) p_i a_{1,j}^{(i+n)+a_{2,j}(m+r)} B \left( a_{1,j}(u_{1,j} + l + n) - 1, a_{2,j}(u_{2,j} + m + r) - 1 \right) \]
3.1.5. The likelihood function

The likelihood function ($L$) to estimate $a_{1,j}, c_{1,j}, d_{1,j}, u_{1,j}, v_{1,j}, a_{2,j}, c_{2,j}, d_{2,j}, u_{2,j}$ and $v_{2,j}$ is

$$L = \prod_{i=1}^{t} \left| a_{1,i} \right| \left| a_{2,i} \right| \prod_{i=1}^{t} p_{i}^{a_{1,i}u_{1,i}+a_{2,i}u_{2,i}-1} c_{1,i}^{d_{1,i}} d_{2,i}^{a_{2,i}u_{2,i}-1} c_{2,i}^{d_{2,i}}$$

$$= \prod_{i=1}^{t} \left| a_{1,i} \right| \left| a_{2,i} \right| \prod_{i=1}^{t} p_{i}^{a_{1,i}u_{1,i}+a_{2,i}u_{2,i}-1} c_{1,i}^{d_{1,i}} d_{2,i}^{a_{2,i}u_{2,i}-1} c_{2,i}^{d_{2,i}}$$

The log-likelihood function $l$ is

$$l = \sum_{i=1}^{t} \ln \left[ a_{1,i} \right] + \ln \left[ a_{2,i} \right] + (a_{1,i} u_{1,i} + a_{2,i} u_{2,i} - 1) \ln p_{i}$$

$$+ \ln \left[ c_{1,i} \right] - a_{1,i} u_{1,i} \ln \left[ d_{1,i} \right] - a_{2,i} u_{2,i} \ln \left[ d_{2,i} \right]$$

$$- \ln B \left( u_{2,i}, v_{2,i} \right) + \sum_{w=1}^{W} \ln \left[ a_{2,i} \right] + (a_{2,i} u_{2,i} - 1) \ln (b_{ow})$$

$$+ (v_{2,i} - 1) \ln \left( 1 - (1 - c_{2,i}) \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \right)$$

$$- a_{2,i} u_{2,i} \ln \left( d_{2,i} \right) - \ln \left( B \left( u_{2,i}, v_{2,i} \right) \right) - (u_{2,i} + v_{2,i}) \ln \left( 1 + c_{2,i} \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \right)$$

The likelihood equations are as follows

$$\frac{\partial l}{\partial a_{2,i}} = \sum_{w=1}^{W} \left[ \frac{1}{a_{2,i}} + u_{2,i} \ln (b_{ow}) + (v_{2,i} - 1) \left( \frac{1}{1 - (1 - c_{2,i}) \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}}} \right) - u_{2,i} \ln \left( d_{2,i} \right)$$

$$- \left( u_{2,i} + v_{2,i} \right) \left( \frac{c_{2,i} \ln \left( \frac{b_{ow}}{d_{2,i}} \right) \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}}}{1 + c_{2,i} \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}}} \right) \right] = 0$$

$$\frac{\partial l}{\partial c_{2,i}} = \sum_{w=1}^{W} \left[ v_{2,i} - 1 \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \right] - \left( u_{2,i} + v_{2,i} \right) \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \left( \frac{b_{ow}}{d_{2,i}} \right)^{a_{2,i}} \right] = 0$$
\[
\frac{\partial l}{\partial d_{2,j}} = \sum_{w=1}^{W} \left[ (v_{2,j} - 1) \frac{(1 - c_{2,j}) b_{ow} a_{2,j} (-a_{2,j}) d_{a,j}^{a_{2,j}+1} (1 - c_{2,j}) (b_{ow} a_{2,j})}{1 + (1 - c_{2,j}) (b_{ow} a_{2,j})} \right] + \left( u_{2,j} + v_{2,j} \right) \frac{c_{2,j} b_{ow} a_{2,j} (-a_{2,j}) d_{a,j}^{a_{2,j}+1}}{1 + c_{2,j} (b_{ow} a_{2,j})} = 0
\]

\[
\frac{\partial l}{\partial u_{2,j}} = \sum_{w=1}^{W} \left[ a_{2,j} \ln(b_{ow}) + a_{2,j} \ln(d_{2,j}) - \frac{\partial h(u_{2,j}, v_{2,j})}{\partial u_{2,j}} - \ln \left( \frac{(1 + c_{2,j}) (b_{ow} a_{2,j})}{B(u_{2,j}, v_{2,j})} \right) \right] = 0
\]

\[
\frac{\partial l}{\partial v_{2,j}} = \sum_{w=1}^{W} \left[ \ln \left( 1 - (1 - c_{2,j}) \frac{b_{ow}}{d_{2,j}} \right) \right] - \frac{\partial h(u_{2,j}, v_{2,j})}{\partial v_{2,j}} - \ln \left( \frac{(1 + c_{2,j}) (b_{ow} a_{2,j})}{B(u_{2,j}, v_{2,j})} \right) = 0
\]

\[
\frac{\partial l}{\partial a_{1,j}} = \sum_{i=1}^{I} \left[ \frac{1}{a_{1,j}} + u_{1,j} \ln(p_{i}) + \frac{\partial c_{1,j}}{\partial a_{1,j}} + \frac{\partial c_{1,j}}{\partial C_{7,j}} - \mu_{1,j} \ln(d_{1,j}) \right] = 0
\]

\[
\frac{\partial l}{\partial c_{1,j}} = \sum_{i=1}^{I} \left[ \frac{\partial c_{1,j}}{\partial C_{7,j}} \right] = 0
\]

\[
\frac{\partial l}{\partial d_{1,j}} = \sum_{i=1}^{I} \left[ \frac{\partial c_{1,j}}{\partial d_{1,j}} \right] = 0
\]

\[
\frac{\partial l}{\partial u_{1,j}} = \sum_{i=1}^{I} \left[ a_{1,j} \ln(p_{i}) + \frac{\partial c_{1,j}}{\partial u_{1,j}} \right] = 0
\]

\[
\frac{\partial l}{\partial v_{1,j}} = \sum_{i=1}^{I} \left[ \frac{\partial c_{1,j}}{\partial v_{1,j}} \right] = 0
\]

where

\[
\frac{\partial B(u_{2,j}, v_{2,j})}{\partial u_{2,j}} = \Gamma(v_{2,j}) \frac{\Gamma(u_{2,j})}{\Gamma(u_{2,j})} \left[ (1 - c_{2,j}) (b_{ow} a_{2,j} (-a_{2,j}) d_{a,j}^{a_{2,j}+1}) - \frac{\partial h(u_{2,j}, v_{2,j})}{\partial u_{2,j}} \right] = B(u_{2,j}, v_{2,j}) \left( \sum_{k=1}^{k} \frac{1}{k} \right)
\]
\[
\frac{\partial B(u_{2,i},v_{2,j})}{\partial v_{2,j}} = \Gamma(u_{2,i}) \left( -\gamma + \sum_{k=1}^{v_{2,j}-1} \frac{1}{k} \right) - \Gamma(v_{2,j}) \left( -\gamma + \sum_{k=1}^{u_{2,i}+v_{2,j}-1} \frac{1}{k} \right)
\]
\[
= B(u_{2,i},v_{2,j}) \left( \sum_{k=1}^{v_{2,j}-1} \frac{1}{k} - \sum_{k=1}^{u_{2,i}+v_{2,j}-1} \frac{1}{k} \right)
\]

suppose \( C_{7,i} \) is written as

\[
\sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{7mn}
\]

then

\[
\frac{\partial C_{7,i}}{\partial a_{1,j}} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ C_{7mn} \left( (l + m) \ln \left( \frac{p_{i} - \mu_{1,i}}{d_{i,j}} \right) \right) + (u_{1,j} + m + n) \left( \sum_{k=1}^{a_{1,i}(u_{1,j}+m+n)-1} \frac{1}{k} - \sum_{k=1}^{a_{1,i}(u_{1,j}+m+n)-n-1} \frac{1}{k} \right) \right]
\]

\[
\frac{\partial C_{7,i}}{\partial c_{1,j}} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ C_{7mn} \left( \frac{(1 - c_{1,j})m - l c_{1,j}}{c_{1,j}(1 - c_{1,j})} \right) \right]
\]

\[
\frac{\partial C_{7,i}}{\partial d_{1,j}} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ C_{7mn} \left( -\frac{a_{1,j}(l + m)}{d_{i,j}} \right) \right]
\]

\[
\frac{\partial C_{7,i}}{\partial u_{1,j}} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ C_{7mn} \left( \sum_{k=1}^{u_{1,i}+v_{2,j}+m-1} \frac{1}{k} - \sum_{k=1}^{u_{1,i}+v_{2,j}-1} \frac{1}{k} \right) + a_{1,j} \left( \sum_{k=1}^{a_{1,i}(u_{1,j}+m+n)-1} \frac{1}{k} - \sum_{k=1}^{a_{1,i}(u_{1,j}+m+n)-n-1} \frac{1}{k} \right) \right]
\]

\[
\frac{\partial C_{7,i}}{\partial v_{1,j}} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ C_{7mn} \left( \sum_{k=1}^{v_{2,j}-1} \frac{1}{k} - \sum_{k=1}^{v_{2,j}-l-1} \frac{1}{k} \right) + a_{1,j} \left( \sum_{k=1}^{u_{1,i}+v_{2,j}+m-1} \frac{1}{k} - \sum_{k=1}^{u_{1,i}+v_{2,j}+m+n-1} \frac{1}{k} \right) \right]
\]
3.2. Generalized beta - Normal convolution

Although Figure 1. covers normal distribution, we cannot derive the estimator of the true intensity value when the noise is normal, from the Equation (1). The normal distribution in Figure 1. is the normal distribution with one parameter. Therefore, in this section, we derive the formula to compute the corrected background intensity when the noise is symmetrically distributed, a normal distribution.

3.2.1. The joint density function

Under the convolution model in Equation (1), where \( P_i \) is the observed intensity of the regular \( i^{th} \) gene, \( S_i \) is the true intensity of the \( i^{th} \) gene, with

\[
S_i \sim f_1(s_i; a_j, c_j, d_j, u_j, v_j) = \frac{|a_j|^{a_j/v_j-1} \left(1 - (1 - c_j) \left(\frac{u_j}{a_j}\right)\right)^{v_j-1}}{d_j^{a_j/v_j} B(u_j, v_j) \left(1 + c_j \left(\frac{u_j}{a_j}\right)\right)^{u_j+v_j}},
\]

\[0 \leq c_j \leq 1, a_j, d_j, u_j \text{ and } v_j \text{ positive } s_i > 0.\]

And \( B_i \) is the background noise with

\[B_i \sim f_2(b_i; \mu, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2}(b_i-\mu)^2}}{\sigma\sqrt{2\pi}}; \mu \in \mathbb{R}, \sigma^2 > 0, b_i > 0.\]

The joint density function of \( S_i \) and \( B_i \) is

\[
f_{s_i, b_i}(s_i, b_i) = \frac{|a_j|^{a_j/v_j-1} \left(1 - (1 - c_j) \left(\frac{u_j}{a_j}\right)\right)^{v_j-1} e^{-\frac{1}{2\sigma^2}(b_i-\mu)^2}}{d_j^{a_j/v_j} B(u_j, v_j) \left(1 + c_j \left(\frac{u_j}{a_j}\right)\right)^{u_j+v_j} \sigma\sqrt{2\pi}}.
\]

The joint density function of \( S_i \) and \( P_i \) is

\[
f_{s_i, p_i}(s_i, p_i) = \frac{|a_j|^{a_j/v_j-1} \left(1 - (1 - c_j) \left(\frac{u_j}{a_j}\right)\right)^{v_j-1} e^{-\frac{1}{2\sigma^2}(p_i-\mu)^2}}{d_j^{a_j/v_j} B(u_j, v_j) \left(1 + c_j \left(\frac{u_j}{a_j}\right)\right)^{u_j+v_j} \sigma\sqrt{2\pi}}.
\]
3.2.2. The marginal density function

The marginal density function of $P_i$ is

$$f_{P_i}(p_i) = \frac{|a_i|}{d_i^{1+i}|B(u_i,v_i)\sigma_i\sqrt{2\pi}}$$

$$\times \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ (-1)^{l+m}(1-c_i)^l c_i^m \left(\frac{v_j}{1} - l\right) \left(\frac{u_j + v_j + m - 1}{m}\right) \int_0^{a_i(u_i + l + m) - 1} e^{-\frac{1}{2}(p_i - s_i)^2} ds_i \right]$$

(19)

Let $\frac{(z_i - (p_i - \mu))}{\alpha_j} = z_j$ therefore, the equation (21) becomes

$$= \frac{|a_i|}{d_i^{1+i}|B(u_i,v_i)\sigma_i\sqrt{2\pi}}$$

$$\times \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ (-1)^{l+m}(1-c_i)^l c_i^m \left(\frac{v_j}{1} - l\right) \left(\frac{u_j + v_j + m - 1}{m}\right) \int_0^{a_i(u_i + l + m) - 1} e^{-\frac{1}{2}(p_i - s_i)^2} ds_i \right]$$

(20)

Let $\frac{z_j^2}{2} = x_j$, therefore, the equation (22) becomes

$$K_2 C_{T,j}$$

(21)

where

$$K_2 = \frac{|a_i|}{2\sqrt{\pi}} p_i^{a_i - 1} d_i^{1+i}|B(u_i,v_i)$$

$$C_{T,j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ (-1)^{l+m}(1-c_i)^l c_i^m \left(\frac{v_j}{1} - l\right) \left(\frac{u_j + v_j + m - 1}{m}\right) \int_0^{a_i(u_i + l + m) - 1} e^{-\frac{1}{2}(p_i - s_i)^2} ds_i \right]$$

and $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function.
3.2.3. The conditional density function

The conditional density function of $S_i$ where it is known that $P_i = p_i$ is

$$f_{S_i|P_i}(s_i|p_i) = \frac{\sqrt{2} p_i^{1-a_iu_i} s_i^{a_iu_i-1}(1 - (1 - c_i) \left(\frac{a_i}{\sigma_i}\right)^a)^{v_i-1} e^{-\frac{1}{2\sigma_i^2} (p_i - \mu_i)^2}}{C_{7,i} \sigma_i \left(1 + c_i \left(\frac{s_i}{\sigma_i}\right)^{a_iu_i+v_i}\right)}$$

3.2.4. The corrected background intensity value

The corrected background intensity under this generalized beta-normal convolution is

$$\frac{C_{8,j}}{p_i - C_{7,j}}$$

where

$$C_{8,j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ \frac{(-1)^{l+m}(1 - c_j)^l c_j^m v_j^{m-n}} {d_j^{n(l+m)} (p_i - \mu_j)^n} \frac{(u_j + v_j + m - 1) \left(\frac{a_j(u_j + l + m)}{n}\right)} {\Gamma(u_j + v_j + m)} \right] \times \left( p_i - \mu_j \right)^{a_j(l+m)} \sigma_j^{2n+2} \gamma \left( \frac{n+1}{2}, \frac{\left(\frac{\mu_j}{\sigma_j}\right)^2}{2}\right) - \gamma \left( \frac{n+1}{2}, \frac{\left(\frac{p_i - \mu_j}{\sigma_j}\right)^2}{2}\right)$$

and $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function.

3.2.5. The likelihood function

The likelihood function ($L$) to estimate $a_j, c_j, d_j, u_j, v_j, \mu_j$ and $\sigma_j^2$ is

$$= \prod_{i=1}^{l} \left[ \ln |a_j| + (a_j u_j - 1) \ln p_i + \ln(C_{7,i}) - \ln(2) - \frac{\ln(\pi)}{2} - a_j u_j \ln(d_j) - \ln \left( B(u_j, v_j) \right) \right]$$
The likelihood equations are as follows

\[
\frac{\partial l}{\partial \mu_j} = \sum_{w=1}^{W} \left[ \frac{(b_{ow} - \mu_j)^2}{\sigma_j^2} \right] = 0
\]

\[
\frac{\partial l}{\partial \sigma_j} = \sum_{w=1}^{W} \left[ \frac{(b_{ow} - \mu_j)^2}{\sigma_j^2} - \frac{1}{\sigma_j} \right] = 0
\]

\[
\frac{\partial l}{\partial a_j} = \sum_{i=1}^{I} \left[ \frac{1}{a_j} + u_j \ln(p_i) + \frac{\partial C_{i,j}}{C_{i,j}} \right] = 0
\]

\[
\frac{\partial l}{\partial c_j} = \sum_{i=1}^{I} \left[ \frac{\partial C_{i,j}}{c_j} \right] = 0
\]

\[
\frac{\partial l}{\partial d_j} = \sum_{i=1}^{I} \left[ \frac{\partial C_{i,j}}{d_j} - \frac{\partial \mu_j}{d_j} \right] = 0
\]

\[
\frac{\partial l}{\partial u_j} = \sum_{i=1}^{I} \left[ \frac{\partial C_{i,j}}{u_j} \right] = 0
\]

\[
\frac{\partial l}{\partial v_j} = \sum_{i=1}^{I} \left[ \frac{\partial C_{i,j}}{v_j} \right] = 0
\]

where

\[
\frac{\partial B(u_j, v_j)}{\partial u_j} = \Gamma(v_j) \frac{\Gamma(u_j) \left( -\gamma + \sum_{k=1}^{u_j-1} \frac{1}{k} \right) - \Gamma(u_j) \left( -\gamma + \sum_{k=1}^{u_j+v_j-1} \frac{1}{k} \right)}{\Gamma(u_j, v_j)} = B(u_j, v_j) \left( \sum_{k=1}^{u_j-1} \frac{1}{k} - \sum_{k=1}^{u_j+v_j-1} \frac{1}{k} \right)
\]

\[
\frac{\partial B(u_j, v_j)}{\partial v_j} = \Gamma(u_j) \frac{\Gamma(v_j) \left( -\gamma + \sum_{k=1}^{v_j-1} \frac{1}{k} \right) - \Gamma(v_j) \left( -\gamma + \sum_{k=1}^{u_j+v_j-1} \frac{1}{k} \right)}{\Gamma(u_j, v_j)}
\]
\[ = B(u_j v_j) \left( \sum_{k=1}^{u_j-1} \frac{1}{k} - \sum_{k=1}^{u_j v_j-1} \frac{1}{k} \right) \]

Suppose \( C_{i,j} \) is written as

\[
\sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n}
\]

then

\[
\frac{\partial C_{i,j}}{\partial a_j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n} \left( (l + m) \ln \left( \frac{p_l - \mu_i}{d_j} \right) + (u_j + m + n) \left( \sum_{k=1}^{a_j(u_j+m+n)-1} \frac{1}{k} - \sum_{k=1}^{a_j(u_j+m+n)-n-1} \frac{1}{k} \right) \right)
\]

\[
\frac{\partial C_{i,j}}{\partial c_j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n} \left( \frac{1-c_j(m-lc_j)}{c_j(1-c_j)} \right)
\]

\[
\frac{\partial C_{i,j}}{\partial d_j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n} \left( \frac{a_j(l+m)}{d_j} \right)
\]

\[
\frac{\partial C_{i,j}}{\partial u_j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n} \left( \left( \sum_{k=1}^{u_j v_j+m-1} \frac{1}{k} - \sum_{k=1}^{u_j v_j-1} \frac{1}{k} \right) + a_j \left( \sum_{k=1}^{u_j v_j+m-1} \frac{1}{k} - \sum_{k=1}^{u_j v_j-1} \frac{1}{k} \right) \right)
\]

\[
\frac{\partial C_{i,j}}{\partial v_j} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{l,m,n} \left( \left( \sum_{k=1}^{v_j-1} \frac{1}{k} - \sum_{k=1}^{v_j-1} \frac{1}{k} \right) + a_j \left( \sum_{k=1}^{u_j v_j+m-1} \frac{1}{k} - \sum_{k=1}^{u_j v_j-1} \frac{1}{k} \right) \right)
\]

4. Discussion and Remarks

We have studied the additive models of background correction for BeadArrays and proposed the generalized model where the true intensity and the noise are assumed skewed distribution and where the true intensity is skewed but the noise is symmetric distribution. In this paper, we have shown the corrected background intensity value of the
This proposed model is a generalization of the available convolution models as in papers [4, 5–7] and [12–15]. The generalization comes from the property of the tree-generalized beta distributions [17] and is explained in [17, 19]. The parameters of the generalized beta distribution are \( a, d, c, u, \) and \( v \). The gamma, exponential, and lognormal distributions are the special cases of the generalized beta distribution.

The gamma distribution is the generalized beta distribution when \( c = 1, v \to \infty, d = \beta v^\frac{1}{2} \) and \( a = 1 \); the exponential distribution is the generalized beta distribution when \( c = 1, v \to \infty, d = \beta v^\frac{1}{2}, a = 1 \) and \( p = 1 \); and the lognormal distribution is the generalized beta distribution when \( c = 1, v \to \infty, d = \beta v^\frac{1}{2}, \beta = \sqrt{\sigma^2 u^2}, u = \frac{\mu + 1}{\sigma^2 u^2} \) and \( \to 0 \).

There are some aspects to be considered while implementing these models:

1. Parameters estimation

   In parameters estimation, there are some methods have been suggested by some researchers. McDonald and Xu [17] used and suggested: the method of maximum likelihood (also was used by Fajriyah [20–22]), the method of moments and the maximum product spacing estimation.

   When \( c = 1 \), the generalized beta distribution is a generalized beta of second kind. Graf and Nedyalkova [23] and Graf et al. [24] have observed that the pseudo maximum likelihood [25–27], the nonlinear least squares on the quantile function [28] and the nonlinear fit for indicator can be implemented to estimate the parameters of the generalized beta of the second kind. The available \textit{VGAM} package in R helps to estimate the parameters of this distribution.

   The existing convolution models use various methods:

   a) The ad-hoc method which is implemented by the RMA method, more details can be found in [5–7], [12] and [19]

   b) Markov chain Monte Carlo simulations, more details can be found in [11]

   c) Maximum likelihood, nonparametrics and method of moments, more details can be found in [12, 13, 15]

   d) Plug-in method, more details can be found in [15]

   e) Fast Fourier transform, more details can be found in [14]

In general, we first need to provide the initial parameters to optimize the log-likelihood function in Equations (18) and (23). The initial parameters of the noise
are easily provided since the benchmarking data set of the negative control probe is available publicly. The initial parameters of the true intensity can be estimated from the observed intensity data subtracted by the mean (or median) of the negative control intensity.

Secondly, once the initial parameters are available, then they will be used to optimize the likelihood function by implementing the optimization method. There are some packages in R which can be used to compute the parameters of the model, for example the optim or optimx package. These parameters are then used to compute the corrected background intensity based on the formula of the chosen model. Remember that the background correction is implemented for each array.

2. the corrected background intensity computation

The corrected background intensity computation includes computations of the infinite summations: $C_{5j}$, $C_{6j}$, $C_{7j}$ and $C_{8j}$. In the author’s experience (in [15]) these infinite summations are close to being constant after certain terms. As a consequence, the ratios of $C_{6j}$ and $C_{8j}$ are able to be computed. Therefore the $C_{5j}$ and $C_{7j}$ difficulty in computing the summations used to compute the corrected background intensity can be eliminated. A sophisticated program written in R, C, Python and its parallelization, could help to speed up the computation.

3. the benchmarking data set

During the implementation of this generalized estimator, the Illumina users need to be aware of the availability of the Illumina Spike-in data set. Once the model is fitted into this data set, the model can then be used to adjust the intensity value.

Apart from the benchmarking criteria for the AffymetrixGeneChips, in the author’s knowledge, the benchmarking criteria for the IlluminaBeadArrays have not been formalized yet. Some researchers, i.e. [12–14, 30] have developed the criteria to assess which background correction methods perform better than the others for the IlluminaBeadArrays.

These criteria together with the criteria in the Affycomp package [31, 32] can be used as the benchmarking criteria for the IlluminaBeadArrays. These criteria have been implemented by Fajriyah [15]. The method has been used by Shi et al. [33] also can be used to assess the best performance of the background correction methods.
4. the benchmarking data set

It is possible that the negative control probes data set is unavailable. In this case, we can adapt the proposed model to the convolution model for background correction without the negative control probes intensities, as in the RMA model.

Considering these statements, clearly the application of this generalized model towards other platforms is possible.

Acknowledgements

This paper is part of the author’s PhD dissertation written under the direction of Professor István Berkes. We would like to thank Paulo Canas Rodrigues, PhD for his comments. Financial support from the Austrian Science Fund (FWF), Project P24302-N18 is gratefully acknowledged. We would also like to thank the anonymous reviewers for their valuable remarks in leading to an improvement of this paper.

References


