Bounding Errors in Estimates of Genome Copy Number Variations Using SNP Array

Jorge Muñoz Minjares and Yuriy S. Shmaliy

***Abstract*—Measurements of chromosomal changes provided using the modern single nucleotide polymorphism (SNP) array technology are accompanied with extensive noise. This makes difficulties in the estimation of genome copy number variations (CNVs) essential for human life. We propose an efficient algorithm for computing the confidence upper and lower boundary limits in order to guarantee an existence of genomic changes with a required probability. The algorithm is designed to approximate the breakpoint jitter probability with the discrete skew Laplace distribution. We test some SNP-based measurements by the upper and lower confidence bound masks and show special cases when the estimated chromosomal change may not exist and the breakpoint locations cannot be estimated with sufficient accuracy.**

***Keywords*—Confidence bound mask, copy number variations, jitter probability, skew Laplace distribution.**

I. Introduction

STRUCTURAL chromosomal changes called copy-number variations (CNVs) in the deoxyribonucleic acid (DNA) are crucial to understand cancer biology. The cell with the DNA typically has a number of copies of one or more sections of the DNA that results in the structural chromosomal rearrangements - deletions, duplications, inversions and translocations of certain parts [1]. A brief survey of types of chromosome alterations involving copy number changes is given in [2]. A commonly accepted unit of measurement inmolecular biology is kilobase (kb) equal to 1000 base pairs of DNA [3]. The human genome with 23 chromosomes is estimated to be about 3.2 billion base pairs long and to contain $20000 - 25000$ distinct genes [4]. Each CNV may range from about one kb to several megabases (Mbs) in size [1].

One of the most modern and efficient technologies to identify the CNV utilizes the single nucleotide polymorphism (SNP) array [5]. Commercial probe-based SNP array platforms are able to genotype, with $> 99\%$ accuracy, about one million SNPs in one assay [6], [7]. The CNV measurements using this technology are represented by the Log R ratios (LRRs), which are the logtransformed ratios of experimental and normal reference SNP intensities centered at zero for each sample [8]. Cancer sample is often contaminated by normal stromal cells, which add normal DNA to measured material. Normal DNA addition smooth changes in Log R Ratio. Experimental noise makes Log R ratio profile much less variant than it is implied and the threshold approach can be used [8]. Note that intensive noise makes almost impossible visual identification of the breakpoints and copy numbers in small segments.

The following properties of the CNVs function can be recognized [2]: 1) It is piecewise constant (PWC) and sparse with a small number of alterations on a long base-pair length, 2) Its constant values are integer, although this property is not survived in the log R Ratio, and 3) The intensive measurement noise in the log R Ratio can be modeled as additive white Gaussian. The estimation problem associated with the CNVs is thus to predict the breakpoints locations and the segmental levels with a maximum possible accuracy and precision acceptable for medical applications. Because of intensive noise [9], the estimates have to be accompanied with the confidence upper bound (UB) and lower bound (LB) masks considered in [10], [11].

Y. S. Shmaliy is with the Department of Electronics Engineering, Universidad de Guanajuato, 36885, Salamanca, Mexico, e-mail: shmaliy@ugto.mx

J. M. Minjares is with the Department of Electronics Engineering, Universidad de Guanajuato, 36885, Salamanca, Mexico, e-mail: ju.munozminjares@ugto.mx

II. Confidence UB and LB Masks

Jitter in the breakpoints and segmental errors can be bounded by the probabilistic confidence UB and LB masks to guarantee an existence of the CNVs between the masks with the required probability.

A. Jitter Distribution

It follows from [10], [12] that jitter in the lth breakpoints n_l of a discrete sparse piecewise-constant signal such as the CNVs measured in white Gaussian noise can be approximated with the discrete skew Laplace probability density function (pdf) [13]:

$$p(k|d_l,q_l)=\frac{(1-d_l)(1-q_l)}{1-d_lq_l}\begin{cases} d_l^k, & k\geq 0\,,\\ q_l^{|k|}, & k\leq 0\,,\end{cases} \tag{1}$$

where $d_l = e^{-\frac{\kappa_l}{\nu_l}} \in (0,1)$ and $q_l = e^{-\frac{1}{\kappa_l\nu_l}} \in (0,1)$, $\kappa_l = \sqrt{\frac{\ln x_l}{\ln(x_l/\mu_l)}}$, $\nu_l = -\frac{\kappa_l}{\ln x_l}$,

$$x_l = \frac{\phi_l(1+\mu_l)}{2(1+\phi_l)}\left(1-\sqrt{1+\frac{4\mu_l(1-\phi_l^2)}{\phi_l^2(1+\mu_l)^2}}\right), \tag{2}$$

$$\mu_l = \frac{P(A_l)[1-P(B_l)]}{P(B_l)[1-P(A_l)]}, \tag{3}$$

$$\phi_l = \frac{P(A_l)+P(B_l)-1}{[1-2P(A_l)][1-2P(B_l)]}, \tag{4}$$

$$P(A_l)=\begin{cases} 1+\frac{1}{2}[\mathrm{erf}(g_l^\beta)-\mathrm{erf}(g_l^\alpha)] & , \quad \gamma_l^- < \gamma_l^+\,,\\ \frac{1}{2}\mathrm{erfc}(g_1^\alpha) & , \quad \gamma_l^- = \gamma_l^+\,,\\ \frac{1}{2}[\mathrm{erf}(g_l^\beta)-\mathrm{erf}(g_l^\alpha)] & , \quad \gamma_l^- > \gamma_l^+\,,\end{cases} \tag{5}$$

$$P(B_l) = \begin{cases} \frac{1}{2}[\mathrm{erf}(h_l^{\alpha}) - \mathrm{erf}(h_l^{\beta})] & , \quad \gamma_l^- < \gamma_l^+ , \\ 1 - \frac{1}{2}\mathrm{erfc}(h_l^{\alpha}) & , \quad \gamma_l^- = \gamma_l^+ , \\ 1 + \frac{1}{2}[\mathrm{erf}(h_l^{\alpha}) - \mathrm{erf}(h_l^{\beta})] & , \quad \gamma_l^- > \gamma_l^+ , \end{cases} \quad (6)$$

where $g_l^{\beta} = \frac{\beta_l - \Delta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^-}{2}}$, $g_l^{\alpha} = \frac{\alpha_l - \Delta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^-}{2}}$, $h_l^{\beta} = \frac{\beta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^+}{2}}$, $h_l^{\alpha} = \frac{\alpha_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^+}{2}}$, $\mathrm{erf}(x)$ is the error function, $\mathrm{erfc}(x)$ is the complementary error function, and

$$\alpha_l, \beta_l \quad = \quad \frac{a_l\gamma_l^- - a_{l+1}\gamma_l^+}{\gamma_l^- - \gamma_l^+} \mp \frac{1}{\gamma_l^- - \gamma_l^+} \times \sqrt{(a_l - a_{l+1})^2\gamma_l^-\gamma_l^+ + 2\Delta_l^2(\gamma_l^- - \gamma_l^+)\ln\sqrt{\frac{\gamma_l^-}{\gamma_l^+}}} \quad (7)$$

if $\gamma_l^- \neq \gamma_l^+$. For $\gamma_l^- = \gamma_l^+$, set $\alpha_l = \Delta_l/2$ and $\beta_l = \pm\infty$. Here $\gamma_l^- = \frac{\Delta_l^2}{\sigma_l^2}$ and $\gamma_l^+ = \frac{\Delta_l^2}{\sigma_{l+1}^2}$ are the signal-to-noise ratios (SNRs) in the lth and $(l+1)$th segments related to a change $\Delta_l = a_{l+1} - a_l$ in the n_l breakpoint and σ_l^2 and σ_{l+1}^2 are the variances of the segmental white Gaussian noise.

1) Jitter Bounds: The left jitter bound (LJB) J_l^{L} and the right jitter bound (RJB) J_l^{R} can be determined with respect to the lth breakpoint $\hat{i}_l$ as follows. Consider the jitter distribution (1) for known γ_l^- and γ_l^+. Increase k in (1) from zero until $p_k < \xi, \%$. Accept the relevant value of k as the right jitter k_l^{R}. Next, reduce k from zero until $p_k < \xi, \%$ and accept the relevant value of k as the left jitter k_l^{L}. Form the LJB and RJB as

$$J_l^{\mathrm{L}} \quad \cong \quad \hat{n}_l - k_l^{\mathrm{R}} , \quad (8)$$

$$J_l^{\mathrm{R}} \quad \cong \quad \hat{n}_l + k_l^{\mathrm{L}} . \quad (9)$$

2) Segmental Bounds: The confidence UB and LB for segmental estimates can thus be specified in the ϑ-sigma sense as:

$$\hat{a}_j^{\mathrm{UB}} \quad \cong \quad \hat{a}_j + \epsilon = \hat{a}_j + \vartheta\sqrt{\frac{\sigma_j^2}{N_j}} = \hat{a}_j + \vartheta\hat{\sigma}_j , \quad (10)$$

$$\hat{a}_j^{\mathrm{LB}} \quad \cong \quad \hat{a}_j - \epsilon = \hat{a}_j - \vartheta\sqrt{\frac{\sigma_j^2}{N_j}} = \hat{a}_j - \vartheta\hat{\sigma}_j . \quad (11)$$

Here, ϑ indicates the bound wideness in terms of $\hat{\sigma_j}$. The probability ξ for the segmental estimate to exceed a threshold ϵ strongly depends on the segmental length N_j and can be determined using 1. Table I gives several values of ϑ, P, and ξ for likely existing genomic changes (50%) [11].

TABLE I. PROBABILISTIC MEASURES FOR GENOMIC CHANGES

	ϑ	$P(\%)$	$\xi(\%)$
Even chances	0.6745	50	50
1-Sigma	1	68.27	31.73
Probable	1.15035	75	25
Almost certain	1.81191	93	7
Typical confidence	1.96	95	5
2-Sigma	2	95.45	4.55
3-Sigma	3	99.73	0.27
Certain	∞	100	0

TABLE II. ALGORITHM FOR COMPUTING THE UB MASK $\mathcal{B}_n^{\mathrm{U}}$ AND LB MASK $\mathcal{B}_n^{\mathrm{L}}$ VIA SNP ARRAY CNVS MEASUREMENTS y_n AND THE BREAKPOINT LOCATIONS ESTIMATES $\hat{n}_l$. GIVEN: BOUND WIDENESS (ϑ-SIGMA).

Input: y_n, $\hat{n}_l$, ϑ

ξ =erfc($\frac{\vartheta}{\sqrt{2}}$), L =length($\hat{n}_l$), M =length(y_n)

$N_{L+1} = M - \hat{n}_L, \quad \hat{n}_0 = 0$

for $j = 1 : L+1$ **do**

$N_j = \hat{n}_j - \hat{n}_{j-1}, \quad \hat{a}_j = \frac{1}{N_j}\sum_{v=\hat{n}_{j-1}}^{\hat{n}_j - 1} y_v$

$\sigma_j = \sqrt{\frac{1}{N_j}\sum_{v=\hat{n}_{j-1}}^{\hat{n}_j - 1}(y_v - \hat{a}_j)^2}$

end for

$\left[k_l^{\mathrm{R}}, k_l^{\mathrm{L}}\right] = k_l^{\mathrm{R}}\_k_l^{\mathrm{L}}\_jitter(\hat{a}_j, \hat{\sigma}_j, L)$ ▷ right jitter ▷ left jitter

$\mathcal{I}_{L+1} = M - 1$, $\mathcal{E}_{L+1} = M - 1$

for $l = 1 : L$ **do**

$$\mathcal{I}_l = \begin{cases} \hat{n}_l - k_l^{\mathrm{R}} & \textbf{if} \quad \Delta_l > 0 \\ \hat{n}_l + k_l^{\mathrm{L}} & \textbf{if} \quad \Delta_l < 0 \end{cases}$$

$$\mathcal{E}_l = \begin{cases} \hat{n}_l + k_l^{\mathrm{L}} & \textbf{if} \quad \Delta_l > 0 \\ \hat{n}_l - k_l^{\mathrm{R}} & \textbf{if} \quad \Delta_l < 0 \end{cases}$$

end for

for $l = 1 : L$ **do**

$$\mathcal{I}_l = \begin{cases} \mathcal{I}_l & \textbf{if} \quad \mathrm{Im}\,\mathcal{I}_l = 0 \\ \mathcal{I}_{l-1} & \textbf{if} \quad \Delta_l \geq 0 \quad \wedge \quad \mathrm{Im}\,\mathcal{I}_l \neq 0 \\ \mathcal{I}_{l+1} & \textbf{if} \quad \Delta_l < 0 \quad \wedge \quad \mathrm{Im}\,\mathcal{I}_l \neq 0 \\ & \textbf{elseif} \ (\mathrm{Im}\,\mathcal{I}_{l+1}) \rightarrow \mathcal{I}_{l+2} \end{cases}$$

$$\mathcal{E}_l = \begin{cases} \mathcal{E}_l & \textbf{if} \quad \mathrm{Im}\,\mathcal{E}_l = 0 \\ \mathcal{E}_{l+1} & \textbf{if} \quad \Delta_l \geq 0 \quad \wedge \quad \mathrm{Im}\,\mathcal{E}_l \neq 0 \\ & \textbf{elseif} \ (\mathrm{Im}\,\mathcal{E}_l + 1) \rightarrow \mathcal{E}_{l+2} \\ \mathcal{E}_{l-1} & \textbf{if} \quad \Delta_l < 0 \quad \wedge \quad \mathrm{Im}\,\mathcal{E}_l \neq 0 \end{cases}$$

end for

$l = 1$, $k = 1$

for $n = 0 : M - 1$ **do**

$$l = \begin{cases} l & \textbf{if} \quad n < \mathcal{I}_l \\ l+1 & \textbf{if} \quad n \geq \mathcal{I}_l \quad \wedge \quad \mathcal{I}_{l+1} > C_l \\ l+2 & \textbf{if} \quad n \geq \mathcal{I}_l \quad \wedge \quad \mathcal{I}_{l+1} \leq C_l \end{cases}$$

$$k = \begin{cases} k & \textbf{if} \quad n < \mathcal{E}_l \\ k+1 & \textbf{if} \quad n \geq \mathcal{E}_l \quad \wedge \quad \mathcal{E}_{l+1} > C_l \\ k+2 & \textbf{if} \quad n \geq \mathcal{E}_l \quad \wedge \quad \mathcal{E}_{l+1} \leq C_l \end{cases}$$

$\mathcal{B}_n^{\mathrm{U}} = \hat{a}_l + 3\sqrt{\frac{\sigma_l^2}{N_l}}$ ▷ UB mask

$\mathcal{B}_n^{\mathrm{L}} = \hat{a}_k - 3\sqrt{\frac{\sigma_k^2}{N_k}}$ ▷ LB mask

end for

Output: $\mathcal{B}_n^{\mathrm{U}}$, $\mathcal{B}_n^{\mathrm{L}}$

TABLE III. ALGORITHM FOR COMPUTING THE KR JITTER k_l^{R} AND KL JITTER k_l^{L}. GIVEN: $\hat{a}_j$, $\hat{\sigma}_j$ AND NUMBER L OF BREAKPOINTS.

Function $k_l^{\mathrm{R}}\_k_l^{\mathrm{L}}\_jitter$, **Input**: $\hat{a}_j$, $\hat{\sigma}_j$, L

for $l = 1 : L$ **do**

$\Delta_l = \hat{a}_{l+1} - \hat{a}_l, \quad \gamma_l^- = \frac{\Delta_l^2}{\sigma_l^2}, \quad \gamma_l^+ = \frac{\Delta_l^2}{\sigma_{l+1}^2}$

α_l by (7) with " $-$ " and $a_l = \hat{a}_l$

β_l by (7) with " $+$ " and $a_l = \hat{a}_l$

$g_l^\beta = \frac{\beta_l - \Delta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^-}{2}}, \quad g_l^\alpha = \frac{\alpha_l - \Delta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^-}{2}}$

$h_l^\beta = \frac{\beta_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^+}{2}}, \quad h_l^\alpha = \frac{\alpha_l}{|\Delta_l|}\sqrt{\frac{\gamma_l^+}{2}}$

P_l^A by (5), P_l^B by (6), ϕ_l by (4)

$\mu_l = \frac{P_l^A(1-P_l^B)}{P_l^B(1-P_l^A)}, \quad x_l$ by (2) , $\kappa_l = \sqrt{\frac{\ln x_l}{\ln(x_l/\mu_l)}}$,

$\nu_l = -\frac{\kappa_l}{\ln(x_l)}, \quad d_l = e^{-\frac{\kappa_l}{\nu_l}}, \quad q_l = e^{-\frac{1}{\kappa_l \nu_l}}$

$k_l^{\mathrm{R}} = \left\lfloor \frac{\nu_l}{\kappa_l} \ln \frac{(1-d_l)(1-q_l)}{\xi(1-d_l q_l)} \right\rfloor$ ▷ right jitter

$k_l^{\mathrm{L}} = \left\lfloor \nu_l \kappa_l \ln \frac{(1-d_l)(1-q_l)}{\xi(1-d_l q_l)} \right\rfloor$ ▷ left jitter

end for

Output: k_l^{R}, k_l^{L}

B. UB and LB Masks

By combining (8), (9), (10), and (11), the UB and LB masks can be formed to outline the region of existence for true CNVs. The algorithm for computing the UB mask $\mathcal{B}_n^{\mathrm{U}}$ and LB mask $\mathcal{B}_n^{\mathrm{L}}$ is developed in Table II.

Its input is measurements y_n, breakpoint estimates $\hat{n}_l$, allowed error probability ξ, number L of the breakpoints, and number of probes M. At the output, the algorithms produces the masks $\mathcal{B}_n^{\mathrm{U}}$ and $\mathcal{B}_n^{\mathrm{L}}$. The masks have the following basic properties [11]:

- The true CNVs exist between $\mathcal{B}_n^{\mathrm{U}}$ and $\mathcal{B}_n^{\mathrm{L}}$ with the given probability ξ: 99.73% in the 3-sigma sense.
- If $\mathcal{B}_n^{\mathrm{U}}$ or $\mathcal{B}_n^{\mathrm{L}}$ covering two or more breakpoints is uniform, then there is a probability of no changes in this region.
- If both $\mathcal{B}_n^{\mathrm{U}}$ and $\mathcal{B}_n^{\mathrm{L}}$ covering two or more breakpoints are uniform, then there is a high probability of no changes in this region.

III. EXPERIMENTAL VERIFICATION

In this section we apply the UB and LB probabilistic masks to determine possible false breakpoints. We employ the SNP array measurements published in [8] and available atavailable at http:// bioinfo−out.curie.fr/ projects/snp_gap/. The estimates $\hat{i}_l$ of the breakpoint locations and segmental levels $\hat{a}_l$ are obtained using the CBS algorithm [14], [15]. The UB and LB masks are used to identify possible false breakpoints or the ones with low probability of existence. Several typical examples are discussed below.

The database processed are the chromosomes 1st, 8th and 13th in files "BLC_B1_T45", "BLC_B1_T31" and "BLC_B1_T37", respectively. Figure 1 and Fig. 2 show measurements and the CNVs estimates of first and second samples. The measurements are normalized and plotted in the Log R ratios (LRRs) scale centered at zero. From

From the first measurement, one can obtain 23 segments and 22 breakpoints using the algorithm *cghcbs*. And for the second, we obtain obtain 59 segments and 60 breakpoints using the same algorithm. To show that not each breakpoint is detected with a high probability, below we consider several cases which indicate possible false breakpoints and their locations.

A. Localized segmental change

We first select a part of measurements in Fig. 1 around the estimated segmental level $\hat{a}_{18}$ and two breakpoints, $\hat{i}_{17}$ and $\hat{i}_{18}$. The relevant confidence masks are shown in Fig. 3 to represent a normal case of CNVs detection with a given probability P taken from Table I. As can be seen, the breakpoints are localized in the 1-sigma sense ($P = 68.27\%$) with no errors, but the segmental level is detected with an error of about $\pm 20\%$. Admitting that the confidence probability of $P = 68.27\%$ may not be sufficient for medical decisions, we apply the masks in the 3-sigma sense ($P = 99.73\%$) and observe that the breakpoints can no longer be detected exactly and the segmental errors increase to about $\pm 50\%$. Although the CNV evidently exists in this case, there is a necessity of defining an exact *value of P which is sufficient for medical needs*.

B. False segmental change

Let us now consider a part of measurements in Fig. 1 around segment a_{10} with the estimates and confidence masks shown in Fig. 4. In this case, the confidence masks do not confirm and existence of segmental changes and there is a high probability of false breakpoints and CNV around a_{10}. Similar conclusions can be made while considering the masks in Fig. 5 which refers to the second chromosome measurement shown in Fig. 2.

C. Possible false segmental change

We finally select a part of measurements of cromosome 8 from the second file around segment a_8 which has a specific in the segmental change detection. The estimates around this points are tested by the confidence masks as shown in Fig. 6. It is seen that the 1-sigma sense ($P = 68.27\%$) does not reveal any changes in this region. The 2-sigma sense ($P = 95.45\%$) suggests that the right-hand breakpoint exists, but the left-hand does not exist. Surprisingly, the 3-sigma sense ($P = 99.73\%$) speaks in favor of the existence of the breakpoints and segmental changes. One thus may conclude that there is a probability that the segmental change around a_8 is false. One may arrive at the same conclusions by considering Fig. 7 and Fig. 8 related to the chromosome measurements shown in Fig. 2.

IV. CONCLUSION

In this paper, we have tested the SNP array measurements and estimates of the CNVs using the *cghcbs* algorithm by the confidence masks. The following inferences have been made. Massive measurement data obtained using the modern SNP array technology are still accompanied with intensive

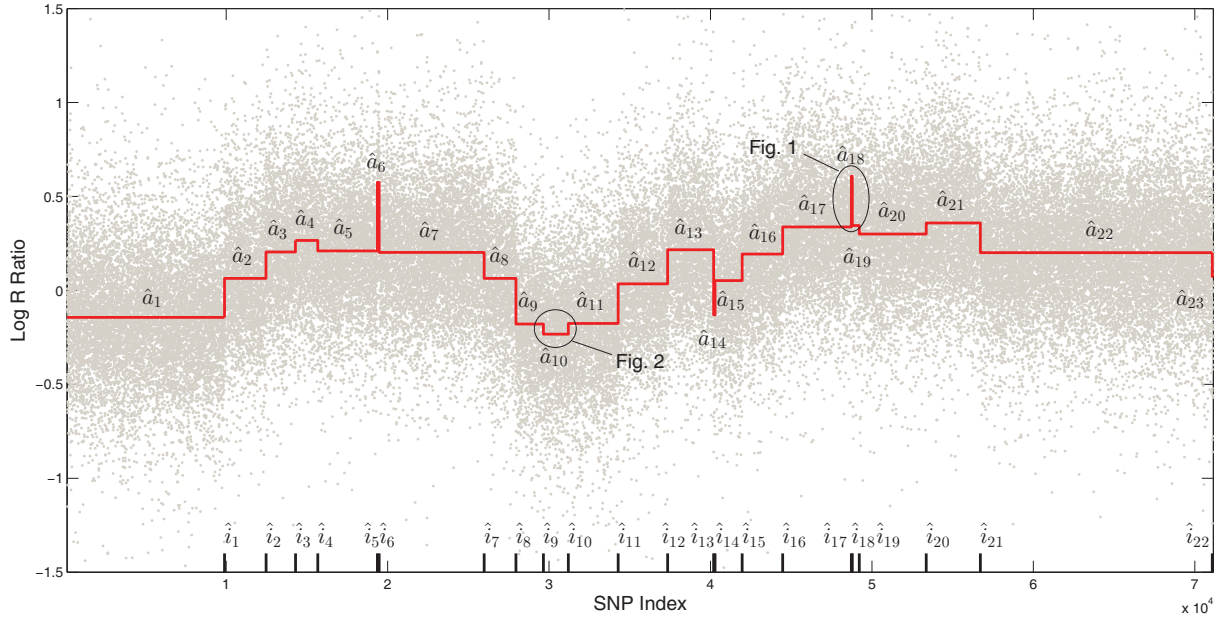

Fig. 1. Measurements and estimates of a part of the 1st chromosome taken from the files "BLC_B1_T45.txt". Segmental levels $\hat{a}_{10}$, $\hat{a}_{21}$ and $\hat{a}_{23}$ cannot be distinguished visually. If we apply the UB and LB masks, then there is a probability that the breakpoints $\hat{i}_9$, $\hat{i}_{10}$, $\hat{i}_{20}$, $\hat{i}_{21}$, $\hat{i}_{22}$ and $\hat{i}_{23}$ do not exist.

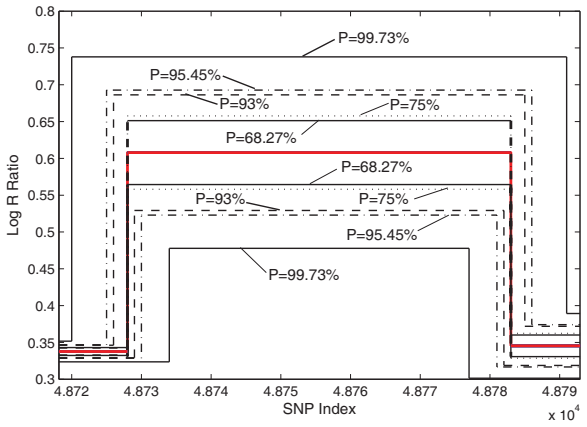

Fig. 3. The UB and LB masks around the segmental level a_{18} for confidence probabilities taken from Table I: the CNV evidently exists, but the breakpoints and segmental level cannot be estimated with high probability.

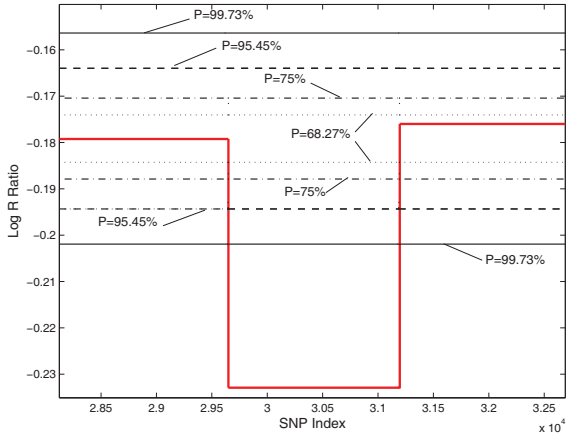

Fig. 4. The UB and LB masks around the segmental level a_{18} for confidence probabilities taken from Table I: the confidence masks do not confirm an existence of segmental changes.

noise that does not guarantee reliable detection of the CNVs. The confidence masks used to test the estimates suggest that some segmental changes may not exists or breakpoints cannot be detected with sufficient accuracy. Moreover, difference confidence limits may lead to contradictive results caused by uncertainties.

We thus formulate the estimation problems which still remain unsolved. The confidence probability has to be defined and optimized from the medical prospective in order to avoid uncertainties and false decisions. On the other hand, the jitter probability that was previously justified approximately via the skew discrete Laplace distribution is not sufficient accurate for low and extra low signal-to-noise ratios. Thus, the jitter distribution in such signals needs further improvements. These

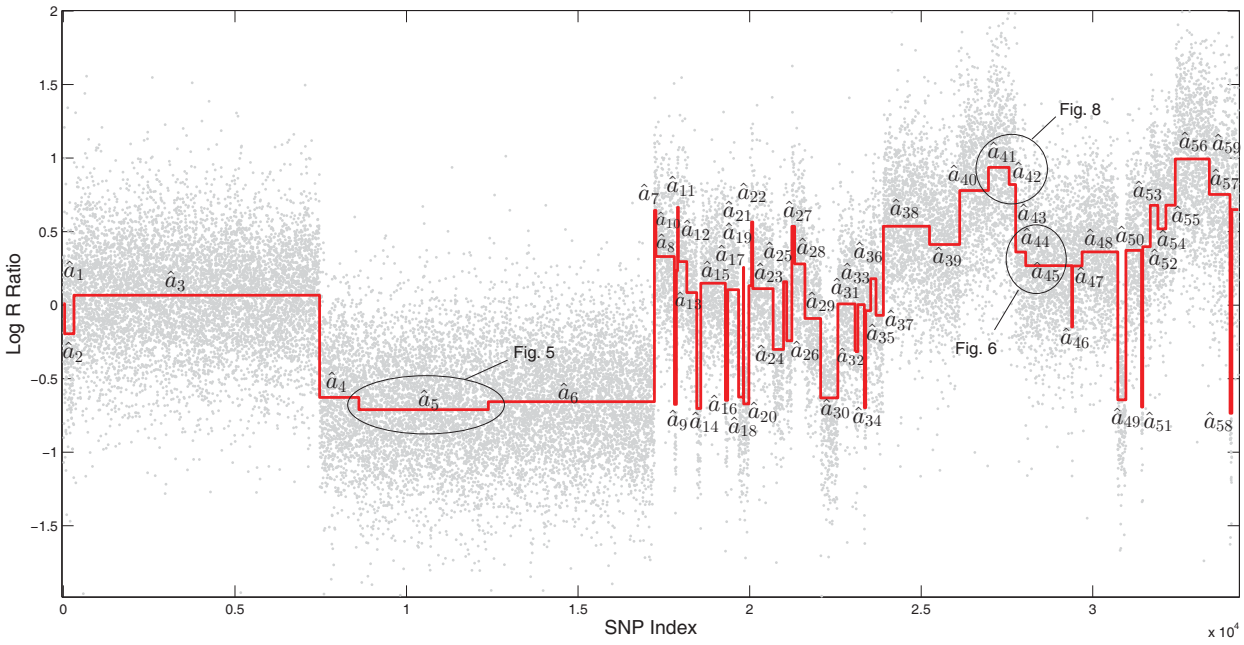

Fig. 2. Measurements and estimates of a part of the 1st chromosome taken from the files "BLC_T37.txt". Segmental levels $\hat{a}_{10}$, $\hat{a}_{21}$ and $\hat{a}_{23}$ cannot be distinguished visually. If we apply the UB and LB masks, then there is a probability that the breakpoints $\hat{i}_9$, $\hat{i}_{10}$, $\hat{i}_{20}$, $\hat{i}_{21}$, $\hat{i}_{22}$ and $\hat{i}_{23}$ do not exist.

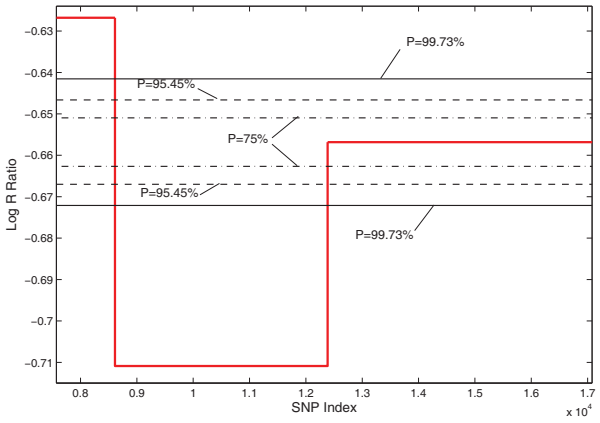

Fig. 5. The UB and LB masks around the segmental level $\hat{a}_5$ from "BLC_T37, for confidence probabilities taken from Table I: the confidence masks do not confirm an existence of segmental changes.

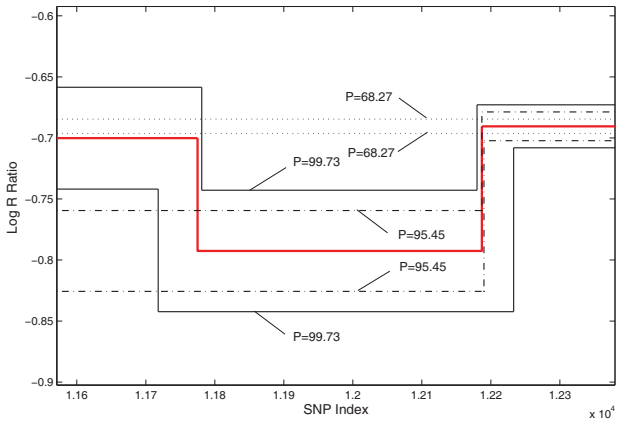

Fig. 6. UB mask and LB to BLC_BI_T31 Chromosome 8 probably does not exist.

problems are currently under the investigation.

REFERENCES

[1] P. Stankiewicz and J. R. Lupski, *Structural variation in the human genome and its role in disease*, Annual Review of Medicine. Vol. 61, pp. 437-455, Feb. 2010.

[2] R. Pique-Regi, A. Ortega, A. Tewfik, and S. Asgharzadeh, *Detection changes in the DNA copy number*, IEEE Signal Processing Mgn. Vol. 29, pp. 98-107, Jan. 2012.

[3] A. Cockburn, M.J. Newkirk, and R.A. Firtel, *Organization of the RNA*

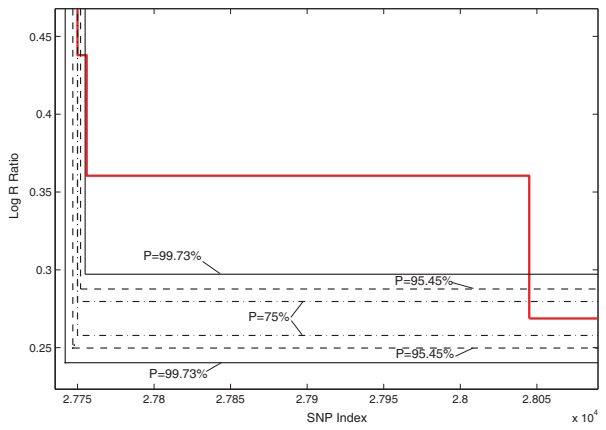

Fig. 7. The UB and LB masks around the segmental level $\hat{a}_{43}$ and $\hat{a}_{44}$ from "BLC_T37, for confidence probabilities taken from Table I: the confidence masks do not confirm an existence of segmental changes.

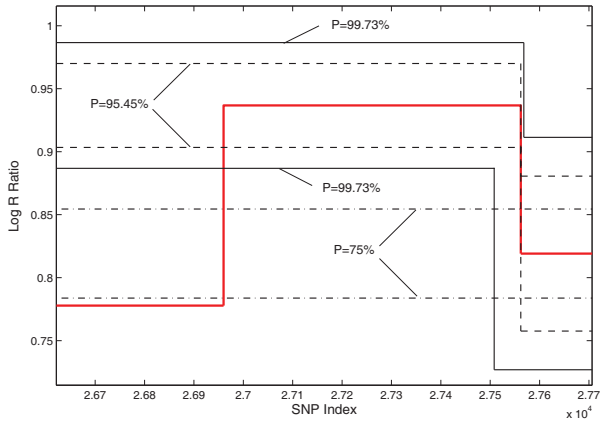

Fig. 8. UB mask and LB to BLC_BI_T31 Chromosome 13 probably does not exist from "BLC_T37, .

genes of dictyostelium: Mapping of the nontrascribed spacer regions, Cell. Vol. 9, pp. 605-613, Dec. 1976.

[4] International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome*, Nature. Vol.431, pp. 931-945, Oct. 2004.

[5] L. Engle, C.Simpson, and J. Landers, *Using high-throughput SNP technologies to study cancer*, Oncogene. Vol. 25, pp.1594–1601, Mar., 2006.

[6] *Affymetrix Genome-Wide Human SNP Array 6.0 data sheet*. Santa Clara (California): Affymetrix, 2007.

[7] *Genome-Wide DNA Analysis BeadChips Data Sheet*. San Diego, California: Illumina, 2009.

[8] T. Popova, V. Boeva, E. Manie, Y. Rozenholc, E. Barillot, and M. H. Stern, *Analysis of Somatic Alterations in Cancer Genome: From SNP Arrays to Next Generation Sequencing*, In Sequence and Genome Analysis I Humans, Animals and Plants. Edited by Ltd iP. iConcept Press Ltd. ISBN: 978–1–477554–913. Aug. 2013.

[9] J. MuñozMinjares, J. CabalAragon, and Y.S. Shmaliy, *Effect of noise on estimate bounds for genome DNA structural changes*, WSEAS Trans. on Biology and Biomedicine. Vol. 11, pp. 52–61, Apr. 2014.

[10] J. MuñozMinjares, J. CabalAragn, and Y. S. Shmaliy, *Confidence limits for genome DNA copy number variations in HR-CGH array measurements*, Biomedical Signal Processing and Control. Vol. 13, pp. 166–173, Mar. 2014.

[11] J. MuñozMinjares, J. CabalAragn, and Y. S. Shmaliy, *Confidence masks for genome DNA copy number variations in applications to HR–CGH array measurements*. Biomedical Signal Processing and Control. Vol. 13, pp. 337–344, Sep. 2014.

[12] J. MuñozMinjares, J. CabalAragón, and Y. S. Shmaliy, *Jitter probability in the breakpoints of discrete sparse piecewise-constant signals*, in Proc. 21st European Signal Process. Conf. (EUSIPCO-2013). pp. 1-5, Marrakech, Marocco, Aug. 2013.

[13] T. J. Kozubowski and S. Inusah, *A skew Laplace distribution on integers*, Annals of the Inst. of Statist. Math. Vol. 58, pp. 555-571, Sep. 2006.

[14] A. B. Olshen, E. S. Venkatraman, R. Lucito, and M. Wigler, *Circular binary segmentation for the analysis of arraybased DNA copy number data*, Biostatistics. Vol. 5, no. 4, pp. 557–572, Oct. 2004.

[15] E. S. Venkatraman and A. B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data*, Bioinformatics. Vol. 23, pp. 657–663, Ene. 2007.

[16] R. Lucito, J. Healy, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J. A. West, S. Rostan, K. C. Nguyen, S. Powers, K. Q. Ye, A. Olshen, E. Venkatraman, L. Norton, and M. Wigler, *Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation,* Genome Research. Vol. 10, pp. 2291–2305, Sep. 2003.

[17] *Representational oligonucleotide microarray analysis (ROMA)*. http://Roma.cshl.org.

[18] J. Muñoz-Minjares, O. Ibarra-Manzano, and Y. S. Shmaliy, *Maximum likelihood estimation of DNA copy number variations in HR-CGH arrays data*, in Proc. 12th WSEAS Int. Conf. on Signal Process., Comput. Geometry and Artif. Vision (ISCGAV'12). Istanbul, Turkey, 2012, pp. 45–50.

[19] O. Vite-Chavez, R. Olivera-Reyna, O. Ibarra-Manzano, Y. S. Shmaliy, and L. Morales-Mendoza, *Time-variant forward-backward FIR denoising of piecewise-smooth signals*, Int. J. Electron. Commun. (AEU). Vol. 67, pp. 406–413, May 2013.

[20] T. Popova, E. Manie, D. Stoppa-Lyonnet, G. Rigaill, E. Barillot, and M. H. Stern, *Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays*, Genome Biol. Vol. 10, no. 11, R128. Nov. 2009.

[21] T. Popova, E. Manie, and M. H. Stern, *Genomic signature of homologous recombination deficiency in breast and ovarian cancers*, Bio-Protocol. Vol. 3, Iss 13, May. 2013.