Improving Jitter Distribution in the Breakpoints of Genome CNVs

JORGE MUNOZ MINJARES,YURIY S. SHMALIY
Universidad de Guanajuato
Department of Electronics Engineering
36885, Salamanca, Gto.
MEXICO

Abstract: The copy number variations (CNVs) are a form of structural genetic changes which are recognized to have an importance for diagnosing human disease. Therefore, accurate estimation of the CNVs using high resolution technologies has been under peer attention in both research and clinical applications during last decades. We propose a more accurate approximation for jitter distribution in the CNVs breakpoints based on the modif ed Bessel function of the second kind and zeroth order. We show that the modif ed distribution allows improving the estimates of the CNVs when the segmental signal-to-noise ratio is small and extremely small.

Key–Words: Copy number variations, jitter distribution, breakpoints.

1 Introduction

It is well-known that structural genetic variations [1, 2], called *genome copy number variations* (CNVs) [3, 4], are associated with disease such as cancer [5]. To measure the genome chromosomal structure the single nucleotide polymorphism (SNP) technology was developed in [9] and applied in [10] to high resolution measurements of the CNVs. But noise in the SNP measurements still remains at a high level [10] and eff cient estimators are required in order to extract information with suff cient accuracy. Unfortunately, no one estimator (optimal or robust) is able to detect the CNVs accurately in large noise [11]. The inability of providing multiple probing in short time [12] complicates the problem. Accordingly, some small changes may be diagnosed as unlikely existing if to test the estimates by the conf dence masks [13].

The SNP data are typically represented in SNP Index with the nth probe, $n_l \in [1, M]$, where M is the number of probes [14]. Figure 1a shows the CNVs picture, in which the n_lth discrete point corresponds to the ith edge or *breakpoint*. The CNVs are often normalized and plotted as $log_2 R/G = log_2 Ratio$, where R and G are the f uorescent Red and Green intensities, respectively [15]. The CNVs function demonstrates the following fundamental properties [16] which are of importance for the estimator design:

- It is piecewise constant (PWC) and sparse with a small number of alterations on a long base-pair length.
- Its constant values are integer, although this property is not survived in the log 2 Ratio.
- The measurement noise in the log R Ratio is highly intensive and can be modeled as additive white Gaussian.

In Fig. 1a, the lth segment a_l and $(l+1)$th segment a_{l+1} are represented with the noise standard deviations σ_l and σ_{l+1} and segmental difference $\Delta_l = a_{l+1} - a_l$ corresponding to the breakpoint at $k = 200$. The signal-to-noise ratios (SNRs) in the lth segment and $(l+1)$th segment can be specif ed as [17],

$$\gamma_l^- = \frac{\Delta_l^2}{\sigma_l^2}, \quad \gamma_l^+ = \frac{\Delta_l^2}{\sigma_{l+1}^2} \tag{1}$$

for supposedly constant segmental values. Intensive noise does not allow for an exact detection of the breakpoints and precise estimates of segmental levels. In fact, the white Gaussian segmental noise [16] which strongly affects the estimation accuracy [18]. Jitter in the breakpoints complicates the problem as has been shown in [19] and reproduced in Fig. 1b.

The estimation theory offers many approaches for signals with the aforementioned properties in order to provide denoising while preserving edges in such signals. One can employ the *wavelet-based* [20, 21, 22], *robust smoothers* [23, 24, 25, 26], *adaptive* and *time-variant smoothers* [27, 28, 29, 30] and *forward-backward* (FB) smoothers forward-backward [31, 32].

Although the skew discrete Laplace density shown as SkL in Fig. 1b [33] [13] *can approximate* the jitter distribution [19] in the breakpoints , the Laplace-based distribution has appeared to be accurate enough only for the SNR values exceeding unity

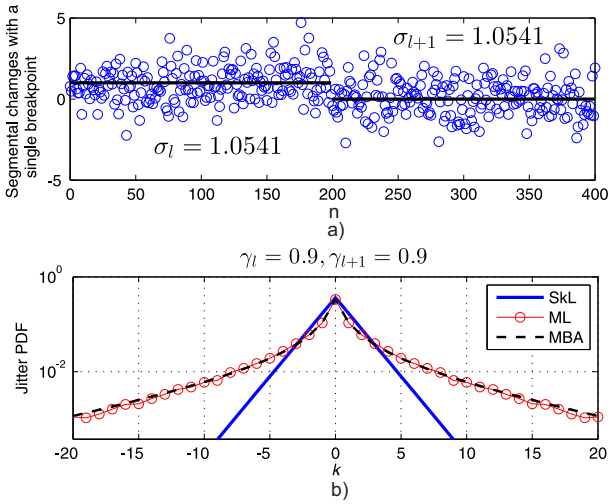

Figure 1: Simulated CNVs with a single breakpoint at $n = 200$ and segmental standard deviations σ_l and σ_{l+1} corresponding to SNRs $\gamma_l^- = \gamma_l^+ = 0.9$: (a) measurement and (b) jitter distribution. Here, ML (circled) is the jitter pdf obtained experimentally using a ML estimator via a histogram over 50×10^3 runs, SkL (solid) is the Laplace distribution, and MBA (dashed) is the proposed Bassel-based approximation.

[18]. Otherwise, the Laplace distribution becomes too rough when the SNR is low, $\gamma_l^{\pm} < 1$, and when it is extremely low, $\gamma_l^{\pm} \ll 1$.

Accordingly, the conf dence masks created based on the Laplace distribution narrow possible bounds of the estimated chromosomal changes and cannot eff - ciently be used to improve the estimates. A more correct probabilistic model of jitter in the breakpoints is thus required.

2 Jitter Distribution in the Breakpoints

In this section we propose and analyse a new approximation to f t measurement data equally well for arbitrary segmental SNRs.

2.1 Approximation with discrete skew Laplace distribution

In order to derive the jitter distribution, the following supposition was made in [34, 11]. Suppose that all of the probes to the left of the breakpoint belong to the segment a_l and all of the probes to the right from the breakpoint belong to the segment a_{l+1}. Otherwise, the probability that one or more probes belong to another segment represents the jitter probability. It has been shown in [34, 11] that, under such a supposition, jitter in the breakpoints of the CNVs measured in white Gaussian noise can be approximated with the discrete skew Laplace probability density function (pdf) recently derived in [35],

$$p(k|d_l, q_l) = \frac{(1-d_l)(1-q_l)}{1-d_l q_l} \begin{cases} d_l^k, & k \geqslant 0\,, \\ q_l^{|k|}, & k \leqslant 0\,. \end{cases} \tag{2}$$

Several properties have been reveled with an extensive investigation of pdf (2) in applications to the CNVs-like signals measured in white Gaussian noise in [34] as the following:

- Density (2) is reasonably accurate if the SNRs exceed unity, $\gamma_l^-, \gamma_l^+ > 1$, and highly accuracy for $\gamma_l^-, \gamma_l^+ \gg 1$.
- It is also reasonably accurate if at least one of the segmental SNRs exceed unity, $\gamma_l^- > 1$ or $\gamma_l^+ > 1$, and highly accuracy if $\gamma_l^- \gg 1$ or $\gamma_l^- \gg 1$.
- The approximation error is large when $\gamma_l^-, \gamma_l^+ < 1$ and can be unacceptable if $\gamma_l^-, \gamma_l^+ \ll 1$.

An overall conclusion which can be made following [17, 18] is that only easily seen breakpoints can be f t with the Laplace distribution (2). The Laplace distribution can be useless in making any decision about the CNVs structures via the estimates if the chromosomal changes are not brightly pronounces. *A more correct jitter distribution is thus required.*

2.2 Approximation based on modified Bessel functions

Modern technologies such as the SNP still do not allow for multiply repeated probes of chromosomal changes in short time that makes it impossible to learn the jitter distribution experimentally. We therefore provide extensive long-term simulations of the CNVs probes in white Gaussian noise environment and learn their statistical properties. To this end, we generated several measurements of length M with one breakpoint $\hat{n}_l$ at $k = 200$ and two neighboring segments with known changes a_l and a_{l+1} as shown in Fig.1a. The standard deviations σ_l and σ_{l+1} of the segmental white Gaussian noise were set for the given SNR (1).

To detect the breakpoint, we use the ML estimator which is based on the Ordinary Least Squares (OSL). The MSE in the ML estimate is minimized for the stepwise CNVs signal, in which a_l, a_{l-1}, and the breakpoint location are used as variables. The breakpoint location is detected when the MSE in the ML estimate reaches a minimum.

In our simulation, detection of the breakpoint location was repeated 50×10^3 times for each generated

noise sequence with a constant SNR. The histogram was plotted as a number of the events in the k scale for each SNR. In order to avoid ripples, such a procedure has been repeated 9 times and the estimates were averaged. Normalized for a unit area, the histogram was accepted as the experimentally def ned *jitter pdf* as shown in Fig.1b with circles.

A complete picture of the experimentally def ned one-sided jitter pdf for equal SNRs in each segment is shown in Fig. 2. As can be concluded by analysing this f gure, the Laplace-based distribution (2) does not f t the true histogram over all k and another approach is required which we will consider next.

2.2.1 Modified Bessel functions

A preliminary analysis has shown that, among available special functions, the modif ed Bessel function of the second kind $K_0(x)$ and zeroth order is a most good candidate to f t the experimentally measured densities shown in Fig. 2. In our approximation, we use the following form of $K_0(x)$,

$$K_0[x(k)] = \int_0^\infty \cos[x(k)\sinh t]\,dt = \int_0^\infty \frac{\cos[x(k)t]}{\sqrt{t^2+1}}\,dt > 0\,, x(k) > 0 \quad (3)$$

in which variable $x(k)$ depends on index k which represents a discrete departure from the assumed breakpoint location (see Fig. 1b). Because $K_0[x(k)]$ is a positive-valued for $x(k) > 0$ smooth function decreasing with x to zero, we use it to approximate the measured probability densities shown in Fig. 2.

2.2.2 Approximation

In order to use (3) as an approximating function

$$B(k|\gamma) = K_0[x(k)] \quad (4)$$

conditioned on γ for the one-sided jitter probability densities shown in Fig. 2, we represent a variable x via k as $x(k,\gamma) = \ln(\Phi(k,\gamma))$ in a way such that small $k \geqslant 0$ correspond to large values of x and visa versa. Among several candidates, it has been found empirically that the following function $\Phi(k,\gamma)$ f ts the histograms with highest accuracy,

$$\Phi(k,\gamma) = (|k|+1)^{\beta+\alpha|k|}\left[\frac{1+\sqrt{\gamma}}{\gamma} - \epsilon\right], \quad (5)$$

if to set $\gamma = \gamma_l^-$ for $k < 0$, $\gamma = \frac{\gamma_l^- + \gamma_l^+}{2}$ for $k = 0$, and $\gamma = \gamma_l^+$ for $k > 0$, and represent the coeff cients $\alpha(\gamma)$, $\beta(\gamma)$, and $\epsilon(\gamma)$ as

$$\alpha(\gamma) = a_0\gamma + a_1\,, \quad (6)$$
$$\beta(\gamma) = \gamma(b_0\gamma^{b_1} + a_0) + b_2\,, \quad (7)$$
$$\epsilon(\gamma) = c_0\gamma^{c_1} + c_2\,. \quad (8)$$

where $a_0 = 0.02737$, $a_1 = -4.5\times10^{-3}$, $b_0 = 0.3674$, $b_1 = -0.3137$, $b_2 = 0.8066$, $c_0 = 0.8865$, $c_1 = -1.033$ and $c_2 = -1.233$ were found in the mean square error (MSE) sense. These values were found in several iterations until the MSE reached a minimum.

Table 1: MSEs produced by Laplace-based (2) and Bessel-based (4) approximations.

γ	MSE by (2)	MSE by (4)
0.1	$7.6e^{-5}$	$1.6e^{-6}$
0.2	$7.7e^{-5}$	$8.6e^{-7}$
0.3	$7.5e^{-5}$	$4.7e^{-7}$
0.5	$6.6e^{-5}$	$3.5e^{-7}$
0.7	$5.9e^{-5}$	$2.2e^{-7}$
0.9	$5.3e^{-5}$	$1.57e^{-7}$
1.37	$4.1e^{-5}$	$1.5e^{-7}$

In Table 1, we give the MSEs produced with the approximation function $B(k)$ in a comparison with the MSEs produced using the Laplace-based approximation (2). As can be seen in Fig. 1a, the approximation provided via $B(k)$ is much more accurate than (2) for any reasonably small γ (see Table 1).

3 Probabilistic Masks

3.1 Masks for Bessel-based approximation

It follows from Fig. 1 that, in view of large noise, estimates of the CNVs may have low conf dence, especially with small SNR $\gamma \leqslant 1$. Thus, each estimate requires conf dence boundaries within which it may exist with a given probability. The problem one faces here is coupled with the fact that, having a single chromosomal probing, we never know if the estimate is most probable or less probable regarding an actual unknown picture. This is well illustrated in Fig. 2 in [33] to show that the breakpoint can be detected in a wide range and far from an actual location if to repeat probing. In other words, some segmental levels and breakpoints can be detected by an estimator close to actual ones, whereas some others not. Even so, with no additional information, there is no other way but to f nd

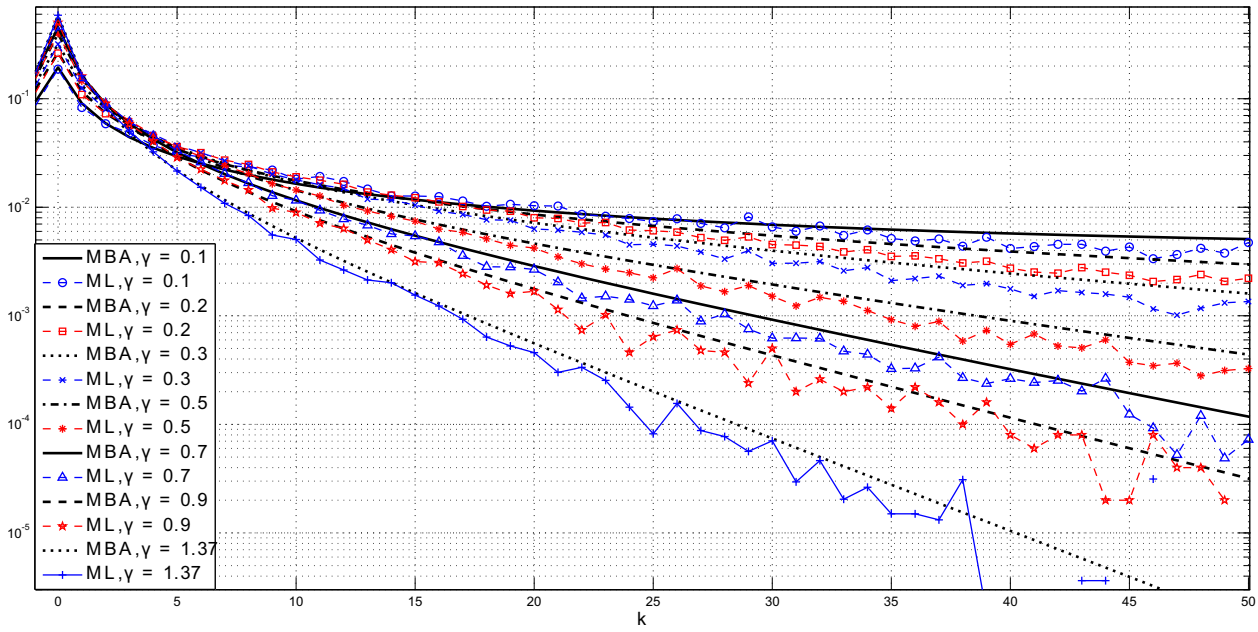

Figure 2: Experimentally def ned one-sided jitter probability densities (dotted) of the breakpoint location for equal segmental SNRs γ in the range of $M = 400$ points with a true breakpoint at $n = 200$. The experimental density functions were found using the ML estimator. The histogram was plotted over 50×10^3 runs repeated 9 times and averaged. Approximations (continuous) are provided using the proposed Bessel-based approximation depicted as MBA.

the conf dence boundaries and probabilistic masks for these estimates. Below, we will follow this approach referring to [13, 33].

Given an estimate $\hat{a}_l$ of the lth segmental level in white Gaussian noise, the probabilistic upper boundary (UB) and lower boundary (LB) can be specif ed for this estimate for the given conf dence probability $P(\vartheta)$ in the ϑ-sigma sense as [13]

$$\hat{a}_l^{\rm UB} \cong \hat{a}_l + \epsilon = \hat{a}_l + \vartheta\sqrt{\frac{\sigma_j^2}{N_l}} = \hat{a}_l + \vartheta\hat{\sigma}_l\,, \quad (9)$$

$$\hat{a}_l^{\rm LB} \cong \hat{a}_l - \epsilon = \hat{a}_l - \vartheta\sqrt{\frac{\sigma_j^2}{N_l}} = \hat{a}_l - \vartheta\hat{\sigma}_l\,. \quad (10)$$

where ϑ indicates the boundary wideness in terms of the segmental noise variance $\hat{\sigma}_l$ on an interval of N_l points, from $\hat{n}_{l-1}$ to $\hat{n}_l - 1$.

Assuming that the UB mask $\mathcal{B}_l^{\rm UB}$ and LB mask $\mathcal{B}_l^{\rm LB}$ for the Bessel-based approximation can be formed using the same equations as for the Laplace distribution. For the jitter probabilistic left boundary $J_l^{\mathcal{B}L}$ and right boundary $J_l^{\mathcal{B}R}$ only need be replaced the Bessel-based approximation instead of Laplace distribution using the equations def ned in [36]. In doing so, we f rst suppose that the Laplace pdf (2) is equal to the approximating function $B_l(k)$ at $k = 0$,

$$p(k = 0|d_l, q_l) = B_l(k = 0)\,, \quad (11)$$

that gives us $B_l(k = 0) = \frac{1}{\phi_l}$. Next, we def ne the probabilities $P^{\mathcal{B}}(A_l)$ at $k = -1$ and $P^{\mathcal{B}}(B_l)$ at $k = 1$ as

$$P^{\mathcal{B}}(A_l) = \frac{B_l(k=0)}{B_l(k=-1) + B_l(k=0)}\,, \quad (12)$$

$$P^{\mathcal{B}}(B_l) = \frac{B_l(k=0)}{B_l(k=1) + B_l(k=0)}\,. \quad (13)$$

We then substitute (12) and (13) into ξ_l,μ_l, ϕ_l, κ_l and ν_l def ned in [35]. That allows us to def ned in the ϑ-sigma sense if to specify the right-hand jitter $k_l^{\mathcal{B}\rm R}$ and left-hand jitter $k_l^{\mathcal{B}\rm L}$ by , respectively [36],

$$k_l^{\mathcal{B}\rm R} = \left\lfloor \frac{\nu_l^{\mathcal{B}}}{\kappa_l^{\mathcal{B}}} \ln\frac{1}{\xi B_l(k=0)} \right\rfloor\,, \quad (14)$$

$$k_l^{\mathcal{B}\rm L} = \left\lfloor \nu_l^{\mathcal{B}}\kappa_l^{\mathcal{B}} \ln\frac{1}{\xi B_l(k=0)} \right\rfloor\,. \quad (15)$$

where $\lfloor x \rfloor$ means a maximum integer lower than or equal to x. Note that functions (14) and (15) were obtained in [36] by equating (4) to $\xi(N_l) = \mathrm{erfc}(\vartheta/\sqrt{2})$ and solving for k_l.

We f nally def ne the jitter left boundary $J_l^{\mathcal{B}L}$ and right boundary $J_l^{\mathcal{B}R}$ as, respectively ,

$$J_l^{\mathcal{B}\rm L} \cong \hat{n}_l - k_l^{\mathcal{B}\rm R}\,, \quad (16)$$

$$J_l^{\mathcal{B}\rm R} \cong \hat{n}_l + k_l^{\mathcal{B}\rm L}\,, \quad (17)$$

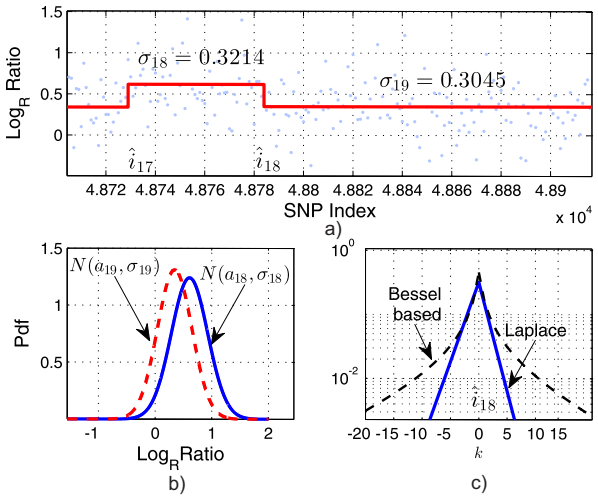

Figure 3: Probes of the 1st chromosome taken from "BLC_B1_T45.txt" around the breakpoint i_{18}: (a) Measurements and estimates with segmental SNRs $\gamma_{18}^{-} = 0.708$ and $\gamma_{19}^{+} = 0.789$, (b) segmental Gaussian densities for σ_6 and σ_7, and (c) Laplace density and Bessel-based approximation for jitter in the breakpoint.

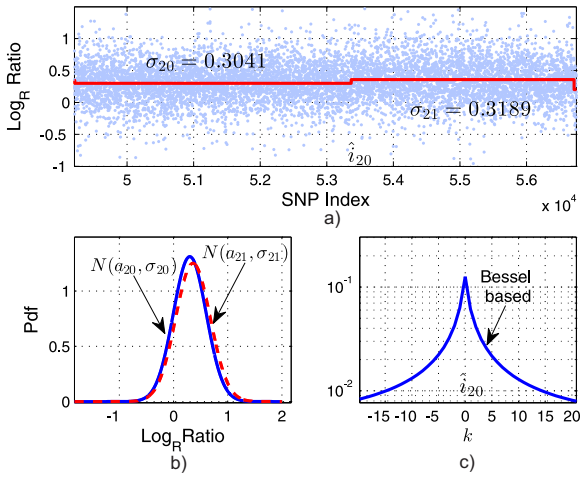

Figure 4: Probes of the 1st chromosome taken from "BLC_B1_T45.txt" around the breakpoint i_{20}: (a) Measurements and estimates with segmental SNRs $\gamma_{20}^{-} = 0.038$ and $\gamma_{21}^{+} = 0.034$, (b) segmental Gaussian densities for σ_{20} and σ_{21}, and (c) Bassel-based approximation of the breakpoint pdf.

and use in the algorithm [13] previously designed for the conf dence masks based on the Laplace distribution.

By combining (9) and (10) with (16) and (17), the probabilistic masks can be formed as shown in [13] to bound the CNV estimates in the ϑ-sigma sense for the given conf dence probability $P(\vartheta)$. An important property of these masks is that they can be used not only to bound the estimates and show their possible locations on a probabilistic f eld [13, 33], but also to remove supposedly wrong breakpoints. Such situations occur each time when the masks reveal double UB and LB uniformities in a gap of three neighbouring detected breakpoints. If so, then the unlikely existing intermediate breakpoint ought to be removed.

4 Testing Estimates Obtained Using SNP Array

In order to demonstrate eff ciency of the Bessel-based probabilistic masks formed in Section 3 and get practically useful results, in this section we employ the probes available using the modern SNP array technology and the CNV estimates obtained by different methods. We test the estimates by the proposed Bessel-based masks and old Laplace-based masks [13] in order to show a difference for the given conf - dence probability.

4.1 Confidence of the Detected Breakpoints

To testing practical estimates for the conf dence in the breakpoint detection, we employ the SNP array chromosomal measurements [10] which are available from http://bioinfoout.curie.fr/diagup projects/snp_ gap/. First, we test the estimates $\hat{i}_l$ of the breakpoint locations and segmental levels $\hat{a}_l$ which are obtained in [37, 38] using the circular binary segmentation (CBS) algorithm. The database processed correspond to the 1st chromosome in "BLC_B1_T45". Figure 3a shows a zoomed sample which represents probing using the 300K Illumina SNP_array.

The measurements are normalized and plotted in the Log R Ratios (LRRs) scale centered at zero.

We f rst select in Fig. 3a a part of probes around the breakpoint $\hat{i}_{18}$ with two estimated segmental levels $\hat{a}_1 8$ and $\hat{a}_1 9$. Segmental Gaussian densities are shown in Fig. 3b and we notice the segmental difference is $\Delta = \hat{a}_{19} - \hat{a}_{18} = -0.2705$. Figure 3c sketches the Laplace density and proposed Bessel-based approximation computed for the breakpoint $\hat{i}_{18}$. As can be seen, the departure from the breakpoint is accompanied with an increasing difference between the Laplace and Bessel-based densities, especially when $k \gg 1$.

Another example is shown in Fig. 4 for the probes taken around the breakpoint $\hat{i}_{20}$ with two estimated segmental levels $\hat{a}_{20}$ and $\hat{a}_{21}$. A specif c of this chromosomal section is that the segmental differences are very small, $\Delta = \hat{a}_{20} - \hat{a}_{21} = 0.0592$, and such that the Laplace distribution cannot be applied in view of

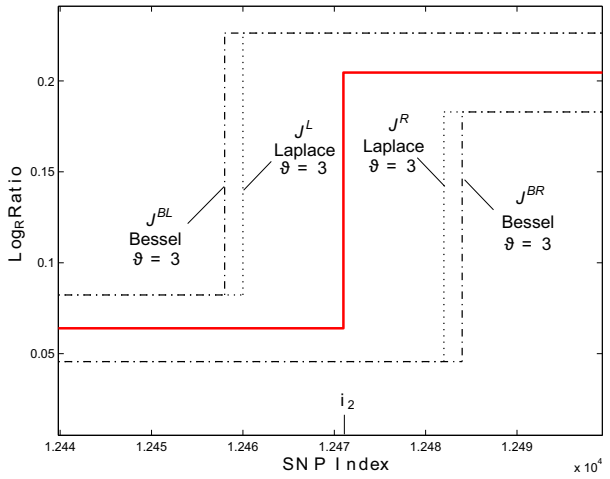

Figure 5: Jitter left boundaries $J_l^{\mathcal{B}L}$, J_l^L and right boundaries $J_l^{\mathcal{B}R}$, J_l^R for the breakpoint i_2.

imaginary values. In turn, the Bessel-based approximation serves well as shown in Fig 4c.

4.2 Testing Estimates by Confidence Masks

Our purpose now is to test the complete CNV estimates by the probabilistic masks. Specif cally, we employ the probes of the 1st chromosome available from "BLC_B1_T45.txt" obtained using the SNP array technology.

Inherently, the more accurate Bessel-based approximation extends the jitter probabilistic boundaries with respect to the Laplace-based ones, especially for low SNRs. We illustrate it in Fig. 5, where the estimates of the 1st chromosome were tested by $\mathcal{B}_l^{UB}$, $\mathcal{B}_l^{LB}$, $\mathcal{L}_l^{UB}$, and $\mathcal{L}_l^{LB}$ for $\vartheta = 3$ (conf dence probability $P = 99.73\%$).

In Fig. 6, we show masks $\mathcal{B}_l^{UB}$ and $\mathcal{B}_l^{LB}$ placed in the vicinity of segment $\hat{a}_{18}$ for several conf dence probabilities: $\vartheta = 0.6745$ ($P = 50\%$), $\vartheta = 1$ ($P = 68.27\%$), $\vartheta = 2$ ($P = 95.45\%$), and $\vartheta = 3$ ($P = 99.73\%$). What the masks suggest here is that the CNV evidently exists with high probability, but the segmental levels and the breakpoint locations cannot be estimated with high accuracy, owing to low SNRs.

It also worth emphasizing on a special case when the masks $\mathcal{L}_l^{\mathrm{UB}}$ and $\mathcal{L}_l^{\mathrm{LB}}$ are not able to conf rm or deny an existence of segmental changes with high probability, owing to an inability of computing the Laplace-based masks for extremely low SNRs. Figure 7 and Fig. 8 illustrate such situations. Just on the contrary, the masks $\mathcal{B}_l^{UB}$ and $\mathcal{B}_l^{LB}$ can be computed for any reasonable SNR.

A conclusion that can be made based on the results illustrated in Fig. 5 - Fig. 8 is that the Bessel-based probabilistic masks can be used to improve es-

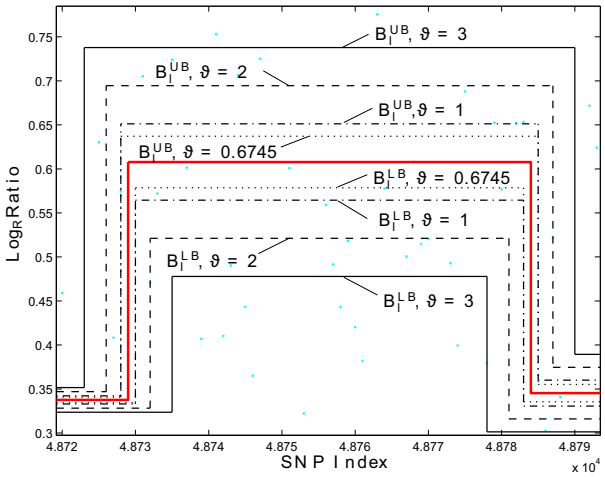

Figure 6: The $\mathcal{B}_l^{UB}$ and $\mathcal{B}_l^{LB}$ masks placed around the segmental level a_{18} for several conf dence probabilities [13]. Here, the CNV exists with high probability, but the segmental levels and the breakpoint locations cannot be estimated with high accuracy.

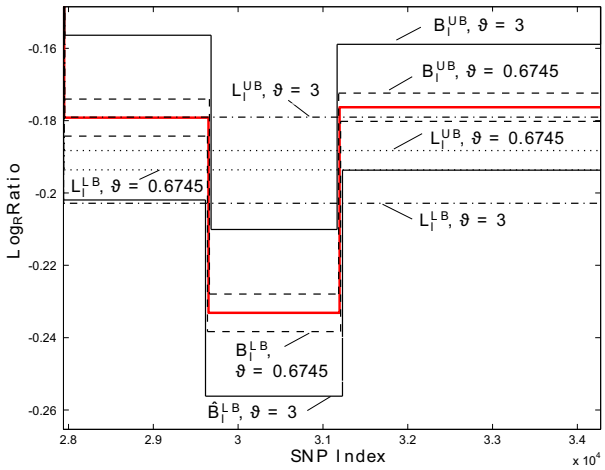

Figure 7: The conf dence masks placed around a_{10} for $\vartheta = 0.6745$ ($P = 50\%$) and $\vartheta = 3$ ($P = 99.73\%$). Masks $\hat{\mathcal{L}}_l^{\mathrm{UB}}$ and $\hat{\mathcal{L}}_l^{\mathrm{LB}}$ do not conf rm an existence of segmental changes while $\hat{\mathcal{B}}_l^{UB}$ and $\hat{\mathcal{B}}_l^{LB}$ indicate a small change.

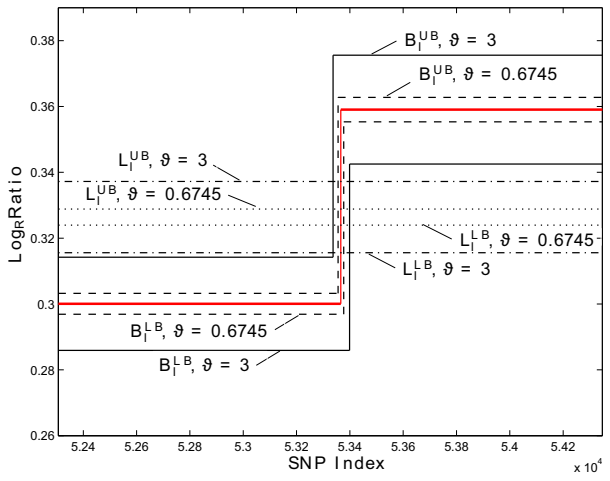

Figure 8: The conf dence masks $\mathcal{L}_l^{\mathrm{UB}}$, $\mathcal{L}_l^{\mathrm{LB}}$, $\mathcal{B}_l^{UB}$ and $\mathcal{B}_l^{LB}$ placed around the breakpoint i_{20} for $\vartheta = 0.6745$ and $\vartheta = 3$.

timates of the chromosomal changes for the required probability that we do in the next section.

5 Conclusion

The proposed Bessel-based approximation of the jitter distribution in the breakpoints is more accurate than the Laplace distribution justif ed earlier. For low and extra low SNR values often observed in probes of small chromosomal changes this is particularly true. The conf dence probabilistic masks formed in this paper for the Bessel-based approximation give a more accurate locations of chromosomal changes on a probabilistic f eld. These masks argue that the CNV estimates may be improved when the SNR reaches values.

References:

[1] A. Reymond , C. N. Henrichsen, L. Harewood, G. Merla, *Side effects of genome structural changes*. Current Opinion in Genetics & Development Volume 17, Issue 5, Oct. 2007, pp. 381–386.

[2] C. Alkan, B. P. Coe, and E.E Eichler, *Genome structural variation discovery and genotyping*. Nat Rev Genet, vol. 12, no. 5, May 2011, pp. 363–376.

[3] R. Redon, S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D., Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, S. W. and M. E. Hurles, *Global variation in copy number in the human genome*. Nature, vol. 444, no. 7118, Nov. 2006, pp. 444–454.

[4] P. J. Hastings, J. R. Lupski, S. M. Rosenberg, and G. Ira, *Mechanisms of change in gene copy number*. Nat Rev Genet, vol. 10, no. 8, Aug. 2009, pp. 551–564.

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

[6] F. Forozan, R. Karhu, J. Kononen, A. Kallioniemi, and O. P. Kallioniemi, *Genome screening by comparative genomic hybridization*. Trends in Genetics, vol. 13, 1997, pp. 405–409.

[7] M. R. Speicher, and N. P. Carter, *The new cytogenetics: blurring the boundaries with molecular biology.* Nat Rev Genet., vol. 6, Oct. 2005, pp. 782–792.

[8] P.C. Ng and E.F. Kirkness, *Whole genome sequencing*. Methods Mol Biol, vol. 628, Feb. 2010, pp. 215–226.

[9] D. G. Wang, J.B. Fan, C.J. Siao, A. Bermo, P. Young, *et. al.*: *Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome*. I Science, vol. 280, May. 1998, pp. 1077–1082.

[10] 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*. Sequence and Genome Analysis I Humans, Animals and Plants. Edited by Ltd iP. iConcept Press Ltd. ISBN: 978–1–477554–913. Aug. 2013.

[11] J.U. Muñoz, J. Cabal and Y.S. Shmaliy, *Jitter probability in the breakpoints of discrete sparse piecewise-constant signals*. Proc. 21st European Signal Process. Conf. (EUSIPCO), Marrakech, Marocco, Sep. 2013.

[12] A. R. Tobler, S. Short, M. R. Andersen, T. M. Paner, J.C. Briggs, S. M. Lambert, ... H. M. Wenz, *The SNPlex Genotyping System: A Flexible and Scalable Platform for SNP Genotyping*. Journal of Biomolecular Techniques: JBT, 16(4), Dec. 2005, pp. 398–406.

[13] J.U. Muñoz, J. Cabal and Y. S. Shmaliy, *Confidence masks for genome DNA copy number variations in applications to HR–CGH array measurements. Biomed. Signal Process. Contr.*, vol. 13, Sep. 2014, pp. 337–344.

[14] A. Abe, S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, H. Matsumura, K. Yoshida, C. Mitsuoka, M. Tamiru, H. Innan, L. Cano, S. Kamoun and R. Terauchi, *Genome sequencing reveals agronomically important loci in rice using MutMap*. Nat Biotech, vol. 30, no. 2, Feb. 2012, pp. 174–178.

[15] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed, *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation*. Nucleic Acids Res., vol. 30, no. 4, p. e15, 2002.

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

[17] Y. S. Shmaliy, *On the multivariate conditional probability density of a signal perturbed by Gaussian noise*. *IEEE Trans. on Inform. Theory*, vol. 53, Dec. 2007, pp. 4792–4797.

[18] J.U. Muñoz, Y.S Shmaliy and A. J. Cabal, *Noise Studies in Measurements and Estimates of Stepwise Changes in Genome DNA Chromosomal Structures*. Advances in Applied and Pure Mathematics, ISBN: 978-1-61804-240-8, 2014.

[19] F. Picard, S. Robin, M. Lavielle, C. Vaisse and J.-J. Daudin, *A statistical approach for array CGH data analysis. BMC Bioinformatics*, vol. 6, no. 1, 2005, pp. 2737.

[20] D. L. Donoho, *De-noising by soft thresholding.* IEEE Trans. Inform. Theory, vol. 41, Mar. 1995, pp. 613–627.

[21] X. P. Zhang and M. D. Desai, *Adaptive denoising based on SURE risk*. IEEE Signal Process. Let., vol. 5, Oct. 1998, pp. 265–267.

[22] S. G. Chang, B. Yu and M. Vetterli, *Adaptive wavelet thresholding for image denoising and compression*. IEEE Trans. Image Process., vol. 9, Sep. 2000, pp. 1532–1546.

[23] J. W. Tukey, *Exploratory Data Analysis*. Menlo Park, CA: Addison-Wesley, 1971.

[24] D. R. K. Brownrigg, *The weighted median filter*. Commun. ACM, vol. 27, Aug. 1984, pp. 807–818.

[25] S. Kalluri and G. R. Arce, *Adaptive weighted myriad filter algorithms for robust signal processing in α-stable environments*. IEEE Trans. Signal Process., vol. 46, Feb. 1998, pp. 322–334.

[26] G. R. Arce, *Nonlinear signal processing: a statistical approach*. New York: Willey, 2005.

[27] O. V. Lepski, E. Mammen and V. G. Spokoiny, *Optimal spatial adaptation to inhomogenous smoothness: an approach based on kernel estimates with variable bandwidth selectors*. The Annals of Statistics, vol. 25, Jun. 1997, pp. 929–947.

[28] J.U. Munoz, O. Ibarra and Y.S. Shmaliy, *Maximum likelihood estimation ofDNAcopy number variations in HR-CGH arrays data*. Proc. 12th WSEAS Int. Conf. on Signal Process., Comput. Geometry and Artif. Vision (ISCGAV' 12), Proc. 12th WSEAS Int. Conf. on Systems Theory and Sc. Comput. (ISTASC' 12), Istanbul, Turkey, 2012, pp. 45–50.

[29] A. Goldenshluger and A. Nemirovski, *Adaptive denoising of signals satisfying differential inequalities*. IEEE Trans. Inf. Theory, vol. 43, Mar. 1997, pp. 872–889.

[30] Y. S. Shmaliy and L. Morales, *FIR smoothing of discrete-time polynomial models in state space*. IEEE Trans. Signal Process., vol. 58, May. 2010, pp. 2544–2555.

[31] B. D. Rao and K. V. S. Hari, E*ffect of spatial smoothing on the performance of MUSIC and the minimum-norm method*. IEEE Proc., vol. 137(F), no. 6, Dec. 1990, pp. 449–458.

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

[33] J.U. Muñoz, Y.S. Shmaliy and J. Cabal-Aragon, *Confidence limits for genome DNA copy number variations in HR-CGH array measurements*. *Biomed. Signal Process. Contr.*, vol. 10, Mar. 2014, pp. 166–173.

[34] J. Muñoz-Minjares, J. Cabal-Aragon and Y.S. Shmaliy, *Effect of noise on estimate bounds for genome DNA structural changes.* WSEAS Trans. on Biology and Biomedicine, vol. 11, , Apr. 2014, pp. 52–61.

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

[36] J. Muñoz-Minjares, J. Cabal-Aragon and Y. S. Shmaliy, *Probabilistic bounds for estimates of genome DNA copy number variations using HR-CGH microarray*. Proc. 21st European Signal Process. Conf. (EUSIPCO), Marrakech, Marocco, Sep. 2013.

[37] 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, Oct. 2004, pp. 557–572.

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