COORDINATE-WISE VERSIONS OF THE GRID METHOD FOR THE ANALYSIS OF INTENSITIES OF NON-STATIONARY INFORMATION FLOWS BY MOVING SEPARATION OF MIXTURES OF GAMMA-DISTRIBUTION

Andrey Gorshenin
Institute of Informatics Problems, Russian Academy of Sciences
Vavilova str., 44-2, Moscow, Russia
Email: agorshenin@ipiran.ru

Victor Korolev
Moscow State University, Leninskie Gory, Moscow, Russia
Institute of Informatics Problems, Russian Academy of Sciences
Email: bruce27@yandex.ru

Victor Kuzmin
"Wi2Geo LLC", Russia
Email: shadesilent@gmail.com

Alexander Zeifman
Vologda State Pedagogical University, S.Orlova, 6, Vologda, Russia
Institute of Informatics Problems RAS, ISEDT RAS
Email: a_zeifman@mail.ru

KEYWORDS

Mixture of probability distributions; Gamma-distribution; Information flow; Grid method

ABSTRACT

A coordinate-wise modification of the grid method for separation of mixtures is proposed in the problem of the dynamical monitoring of the stochastic structure of information flows.

INTRODUCTION

Statistical analysis of the traffic in information flows in modern computational and telecommunication systems reveals their noticeable non-stationarity. Within the approach based of representation of the traffic as a stochastic process is very convenient to model this non-stationarity by representing the flow intensity as a non-stationary random process with mixed gamma one-dimensional distributions. As this is so, the evolution of the parameters of these mixtures in time reflects the non-stationarity of the original process and can be used as the estimate of the stochastic structure of the information flow.

GRID METHODS FOR SEPARATION OF MIXTURES OF PROBABILITY DISTRIBUTIONS

EM-type algorithms are traditionally used for the numerical solution of the problem of separation of mixtures of probability distributions when the number of components is comparatively large (Korolev, 2011). If the likelihood function is regular, then these method yields most likely estimates of the parameters of components and weights. But if the likelihood function is irregular and has many local extremes (possibly, infinite), then the EM-algorithm becomes extremely unstable and slow which is absolutely inadmissible for the on-line monitoring of the stochastic structure of information flows, see, e. g., (Korolev, 2011).

To overcome this drawback, recently in the papers (Korolev et al., 2008), (Korolev et al., 2010) the so-called grid algorithms were proposed for the separation of mixtures of probability distributions. The main idea that grid methods rely on is very close to the idea of harmonic analysis.

Consider an identifiable mixture of distribution functions of the form

$$F(x) = \sum_{i=1}^{k} p_i G(x; a_i, \sigma_i), \quad x \in \mathbb{R}, \tag{1}$$

where $k \geq 1$ is an integer. In the classical problem of separation of mixtures, the parameters to be estimated are the triples (p_i, a_i, σ_i), $i = 1, \ldots, k$, with $a_i \in \mathbb{R}$, $\sigma_i \in \mathbb{R}$, $p_i \geq 0$, $p_1 + \ldots + p_k = 1$.

Assume that the numbers $\underline{a}$, $\overline{a}$, $\underline{\sigma}$ and $\overline{\sigma}$ are given so that $\underline{a} \leq a_i \leq \overline{a}$ and $\underline{\sigma} \leq \sigma_i \leq \overline{\sigma}$ for all $i = 1, \ldots, k$. in other words, the finite ranges of the parameters a_i and σ_i are known.

The main idea of the grid methods is in the replacement of the *intervals* $[\underline{a}, \overline{a}]$ and $[\underline{\sigma}, \overline{\sigma}]$ of possible values of the parameters of the components of mixture (1) by *discrete* sets of *known* points. These points may be defined in the following way.

Let ε_a and ε_σ be positive numbers determining the prior requirements to the accuracy of the estimation of the parameters a_i and σ_i:

$$\max_i |a_i - \widehat{a}_i| \leq \varepsilon_a, \quad \max_i |\sigma_i - \widehat{\sigma}_i| \leq \varepsilon_\sigma, \tag{2}$$

Proceedings 27th European Conference on Modelling and Simulation ©ECMS Webjørn Rekdalsbakken, Robin T. Bye, Houxiang Zhang (Editors)
ISBN: 978-0-9564944-6-7 / ISBN: 978-0-9564944-7-4 (CD)

where $\widehat{a}_i$ and $\widehat{\sigma}_i$ are the desired estimates of the parameters. The numbers ε_a and ε_σ also can be interpreted as the thresholds of distinguishability of possible values of the parameters: the values a', a'' and σ', σ'' are, respectively, regarded as undistinguishable, if

$$|a' - a''| \le \varepsilon_a, \quad |\sigma' - \sigma''| \le \varepsilon_\sigma. \tag{3}$$

Let $k_a = [(\overline{a} - \underline{a})/\varepsilon_a] + 1$, $k_\sigma = [(\overline{\sigma} - \underline{\sigma})/\varepsilon_\sigma] + 1$, where the symbol $[z]$ denotes the integer part of a number z. For $r = 1, 2, \ldots, k_a + 1$ set $\widetilde{a}_r = \underline{a} + (r-1)\varepsilon_a$. Similarly, for $l = 1, 2, \ldots, k_\sigma$ let $\widetilde{\sigma}_l = \underline{\sigma} + (l-1)\varepsilon_\sigma$. Then the points with coordinates $(\widetilde{a}_r, \widetilde{\sigma}_l)$ form the nodes of a finite grid covering the rectangle $\{(a, \sigma) : \underline{a} \le a \le \overline{a},\ \underline{\sigma} \le \sigma \le \overline{\sigma}\}$ representing the set of possible values of the parameters of components of mixture (1). The number of nodes of the grid is equal to $K = (k_a + 1)(k_\sigma + 1)$. For convenience, renumber somehow the nodes introducing a *single* index i for the coordinates $(\widetilde{a}_i, \widetilde{\sigma}_i)$ of the node with the number i after renumbering, $i = 1, \ldots, K$.

The basis of the proposed approach is the approximation of mixture (1) by a mixture with a deliberately larger number of *known* components:

$$F(x) = \sum_{i=1}^{k} p_i G(x; a_i, \sigma_i) \approx$$

$$\approx \sum_{i=1}^{K} \widetilde{p}_i G(x; \widetilde{a}_i, \widetilde{\sigma}_i) \equiv \widetilde{F}(x), \quad x \in \mathbb{R}. \tag{4}$$

Such an approximation is practically admissible since, by virtue of (2) and (3), for any pair (a_r, σ_r) of the parameters of a component of mixture (1) there surely exists a pair $(\widetilde{a}_i, \widetilde{\sigma}_i)$ of the parameters of a component of mixture $\widetilde{F}(x)$ which is practically undistinguishable of the original pair. The weights of the rest components of the mixture $\widetilde{F}(x)$, for the parameters of which there is no "close" pair (a_r, σ_r) of the parameters of a component of mixture (1), can be regarded as zero. Indeed, if in relation (4) instead of the *approximate* equality there were the *exact* equality, then due to the identifiability of mixture (1), by the definition of identifiability, the equalities

$$k = K, \quad p_i = \widetilde{p}_i,\ a_i = \widetilde{a}_i,\ \sigma_i = \widetilde{\sigma}_i,\ i = 1, \ldots, k. \tag{5}$$

would hold up to re-indexation. Note that in the mixture $\widetilde{F}(x)$ *only* the weights $\widetilde{p}_1, \ldots, \widetilde{p}_K$ are unknown parameters.

Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be an independent sample of observations each of which is a realization of a random variable with distribution function $F(x)$ defined by (1).

To find the estimates of the weights $\tilde{p}_i$, $i = 1, \ldots, K$, in the mixture $\tilde{F}(x)$ (see (4)) we will use the maximum likelihood method.

Let $(\tilde{a}_i, \tilde{\sigma}_i)$ be the nodes of the grid covering the set of the values of the parameters of components, $i = 1, \ldots, K$. For convenience by $g(x; a, \sigma)$ denote the density corresponding to the distribution function $G(x; a, \sigma)$,

$$g_{ij} = g(x_j; a_i, \sigma_i), \quad j = 1, \ldots, n;\ i = 1, \ldots, K. \tag{6}$$

To understand whether any numerical algorithm oriented at the maximization of the grid likelihood function will converge or not, we should answer the question whether the grid likelihood function

$$L(\boldsymbol{p}; \boldsymbol{x}) = \log \prod_{j=1}^{n} \sum_{i=1}^{K} \tilde{p}_i g_{ij} = \sum_{j=1}^{n} \log \left(\sum_{i=1}^{K} \tilde{p}_i g_{ij} \right) \tag{7}$$

is convex or concave. The answer to this question is given by the following theorem (Korolev et al., 2010) (also see (Korolev, 2011)).

THEOREM 1. *Any grid likelihood function is concave as a function of the weights.*

In (Korolev et al., 2010) it was demonstrated that the grid maximum likelihood method realized by the algorithm of conditional gradient appeared to be a very efficient tool for the approximate separation of mixtures of one-parameter distributions. However, the efficiency of a grid method strongly depends on the dimensionality of the set of parameters and those grid methods which work very rapidly and efficiently for one-parameter mixtures, become considerably slower when they are applied to multi-parameter mixtures since the volume of computations grows exponentially as the dimensionality of parameter increases. To somehow overcome this drawback (or moderate its effect), in the present communication we propose a modification of the grid method, namely, the coordinate-wise versions of the grid method for separation of mixtures of two-parameter distributions.

COORDINATE-WISE GRID METHOD FOR SEPARATION OF MIXTURES OF GAMMA-DISTRIBUTIONS. THE ALGORITHM

The first step of the algorithm consists in the application of the modified grid algorithm over the whole two-dimensional net. The modifications aim at the noise reduction and rejection of insignificant components with small weights.

By virtue of the inaccuracy of the representation of zero in a computer, it is reasonable to regard a component as insignificant not if $p_i = 0$, but if $p_i < \epsilon$, where $\epsilon > 0$ is a small number. Since the components with zero weights cannot gain a positive probability as the grid algorithm proceeds, the increase of ϵ noticeably speeds up the algorithm. Moreover, the increase of ϵ makes it possible to reject those components whose weights go to zero as

the algorithm proceeds. As a result of empirical experiments it was decided to take $\epsilon = \frac{1}{8(K+1)^2}$.

The second modification consists in rejection of the insignificant components, if the total number of existing components exceeds some preassigned critical number Q (in practice we used $Q = 10$). After every S steps the component with the minimum weight is sifted out if its current weight is less than at the previous steps. After this the weights of the remaining components are normalized.

As a stopping rule we considered the condition

$$\sum_{i=1} (K+1)^2 |p_i^m - p_i^{m-1}| < \delta, \qquad (8)$$

where p_i^m is the weight of the ith component (the weight of the ith node of the grid) on the mth step of the algorithm, $\delta > 0$ is the pre-assigned accuracy. In practice we took $\delta \le 10^{-6}$.

The second step of the algorithm is determined as follows. After the first step is accomplished, we obtain the set of components with non-zero weights $(p_1, r_1, s_1), \ldots, (p_{k_1^r}, r_{k_1^r}, \theta_{k_1^r})$ with some $k_1 \in \mathbb{N}$. Choose all the corresponding different estimates $r_1, \ldots, r_{\widetilde{k}_1^r}$ of the parameter r (clearly, $\widetilde{k}_1^r \le k_1^r$). Construct a new grid based on these fixed values of the parameter r, spreading the grid over the possible values of θ in the way it was done on the first step. So, the obtained grid will have $\widetilde{k}_1^r(K+1) \le (K+1)^2$ nodes.

Again start the grid iterative algorithm on this reduced grid. As the result, we obtain the set of components with non-zero weights $(p_1, r_1, s_1), \ldots, (p_{k_1^\theta}, r_{k_1^\theta}, \theta_{k_1^\theta})$ with some $k_1^\theta \in \mathbb{N}$. Choose all the corresponding different estimates $\theta_1, \ldots, \theta_{\widetilde{k}_1^\theta}$ of the parameter θ (clearly, $\widetilde{k}_1^\theta \le k_1^\theta$). Construct a new grid based on these fixed values of the parameter θ, spreading the grid over the possible values of r in the way it was done on the first step. So, the obtained grid will have $\widetilde{k}_1^\theta(K+1) \le (K+1)^2$ nodes.

And so on. Each next step consists of two applications of the usual grid method on the adaptively reduced grids: on the first step we obtain new estimates for the parameter θ with the fixed values of r obtained on the preceding stage and on the second step we obtain new estimates for the parameter r with the fixed values of θ obtained on the preceding stage.

DISCUSSION

The modified coordinate-wise grid method possesses some advantages as compared to the "classical" grid method.

First, at each mth step the coordinate-wise method works with approximately $(\widetilde{k}_m^r + \widetilde{k}_m^\theta)K$ nodes whereas the classical grid method works with approximately K^2 nodes. Therefore, if $\widetilde{k}_m^r + \widetilde{k}_m^\theta$ is noticeably less than K, then the modified algorithm has noticeably less computational complexity and, hence, is more rapid and more efficient.

Second, since the grid is re-adjusted at each step, the modified algorithm is more sensitive. This circumstance is especially important for processing data obtained from non-stationary processes. For example, the modified grid method processes quite reliably those windows which appear to be problematic for the classical grid algorithm. Unlike the standard grid algorithm with the uniform grid, the modified algorithm shows all the peculiarities of mixtures of gamma-distributions with noticeably stratified (or clustered) values of the parameters of components.

Finally, the use of filters which reject insignificant components makes it possible to obtain more illustrative and sharp pictures of the evolution of the parameters of the mixture in the moving mode.

This research was supported by the Russian Foundation for Basic Research, projects no. 11-07-00112a, 12-07-00115a, 12-07-31267mol_a.

REFERENCES

Korolev, V. Yu. 2011. "Probabilistic and Statistical Methods of Decomposition of Volatility of Chaotic Processes." Moscow: Moscow University Publishing House (in Russian), 31-76.

Korolev, V. Yu., Nepomnyshchiy E. V.; Rybalchenko A. G. and Vinogradova A. V. 2008. "Grid methods for separation of mixtures of probability distributions and their application to the decomposition of volatility of financial indexes." *Informatics and Its Applications.* 2, No. 2, 3-18 (in Russian).

Korolev, V. Yu., Nazarov A. L. 2010. "Separating mixtures of probability distributions with the grid method of moments and the grid maximal likelihood method." *Autom. Remote Control.* 71, 455-472.

AUTHOR BIOGRAPHIES

ANDREY GORSHENIN is Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences.

VICTOR KOROLEV is Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences.

VICTOR KUZMIN is researcher, "Wi2Geo LLC", Russia.

ALEXANDER ZEIFMAN is Doctor of Science in physics and mathematics; professor, Dean of the Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; leading scientist, Institute of Territories Socio-Economic Development, Russian Academy of Sciences.