Regularization methods for unfolding of discrete distributions

Vladimir B. Anykeyev [1]

Institute for High Energy Physics, Protvino, Russia

The problem to estimate "true" distribution on the basis of "observed" one in model and detector independent manner is known to be ill-posed inverse problem in mathematics. This problem has the specific for discrete distibution – number of parameters I can exceed the number of measurements N. The solution of the problem on the basis of pseudoinverse operator and regularization procedures were proposed.

Key words: Unfolding; Discrete distribution; Regularization

One of the goals of modern experiments in high energy physics is the investigation of discrete distributions. Among these distributions are multiplicities of charged particles or of hadron jets produced in elementary particles interactions. As the result of finite detector resolution and ineffectiveness of track recognition or jet reconstruction algorithms the migration of events from one class to another occurs. Formally this process is decsribed by the following mathematical model

$$N_n^{obs} = \bar{N}_n^{obs} + \varepsilon_n = \sum_{i=1}^{I} A_{ni} N_i^{tr} + \varepsilon_n, \quad n = 1, \ldots, N, \tag{1}$$

where N_i^{tr} is the true number of events from class i (parameters); N_n^{obs} is the measured number of events in class n (measurements); A_{ni} (element of smearing matrix) is the probability of migration of event from class i into class n; ε_n is the fluctuation of number of events N_n^{obs}. The inverse problem to estimate N_i^{tr} is ill-posed or in other words unstable with respect to noises ε_n.

Nowadays to solve this problem they use the correcting factors method and the method based on the Bayes theorem. It was shown [1], that estimates obtained on the basis of these methods are model dependent, while the goal is to obtain model and detector independent estimates. To obtain estimates with such properties in the class of linear estimates it is natural to use the idea of pseudoinverse operator for the system of linear equations (1), which is defined as [2]

$$A^+ = \begin{cases} A^*(AA^*)^{-1}, N \le I \\ (A^*A)^{-1}A^*, N > I \end{cases} . \tag{2}$$

If one chooses the scalar products in the spaces of parameters $\{N^{tr}\}$ and measurements $\{N^{obs}\}$ respectively

$$(N^{tr}, \tilde{N}^{tr})_{N^{tr}} = \sum_i \tilde{N}_i^{tr} N_i^{tr}, \quad (N^{obs}, \tilde{N}^{obs})_W = \sum_{n,m} \tilde{N}_n^{obs} W_{nm} N_m^{obs}, \tag{3}$$

[1] supported by the RFFR grant No 96-02-16495
E-mail: anikeev@mx.ihep.su

where $W^{-1}_{nm} = \overline{\varepsilon_n \varepsilon_m}$, then $A^* = A^\top W$ and

$$A^+ = \begin{cases} A^\top (AA^\top)^{-1}, & N \leq I \\ (A^\top W A)^{-1} A^\top W, & N > I \end{cases}. \tag{4}$$

The corresponding estimate turns out to be

$$\hat{N}^{tr} = A^+ N^{obs}. \tag{5}$$

Matrices $AA^\top$ and $A^\top WA$ are the Gram matrices respectively for rows and columns of matrix A for the chosen scalar products (3). When essential migration of events from one class to another takes place these matrices will be ill-conditioned. These leads to strong instability of estimates (5) in respect to statistical fluctuations of $N^{obs} - \varepsilon$. This fact becomes evident if we consider spectral representation. In what follows we consider $G, g^{(l)}, \lambda_l$ as one of matrices $AA^\top$ and $A^\top WA$, its eigenvectors and eigenvalues respectively. We introduce new vectors

$$a^{(l)} = \begin{cases} \frac{1}{\sqrt{\lambda_l}} A^\top g^{(l)}, N \leq I \\ \frac{1}{\sqrt{\lambda_l}} A g^{(l)}, \quad N > I \end{cases}, \tag{6}$$

they are orthonormalized and for both cases have the dimension equal to the number of measurements N. Then with the use of spectral representation of matrices G^{-1} the estimate (5) can be written as follows

$$\hat{N}^{tr+} = \begin{cases} \sum_l \frac{1}{\sqrt{\lambda_l}} (N^{obs}, g^{(l)})_{N^{tr}} a^{(l)}, N \leq I \\ \sum_l \frac{1}{\sqrt{\lambda_l}} (N^{obs}, g^{(l)})_W a^{(l)}, \quad N > I \end{cases}. \tag{7}$$

It is clear from (7), that the presence of small λ_l for matrices G leads to big noise component in $\hat{N}^{tr}$.

As the consequence of this fact is the neccessity to regularize the pseudosolution (5). One of the regularization methods is that of spectral widnow. It consists in introduction of the multiplier $f(\lambda)$ under the sum sign in (7). This multiplier is ~ 0 for small λ and ~ 1 for big ones. The simplest choice of $f(\lambda) : f(\lambda) = 1$ for $\lambda > \lambda_{N_R}, N_R < N; \quad f(\lambda) = 0$ for $\lambda < \lambda_{N_R}$. The number of "small" eigenvalues N_R is the regularization parameter.

Another one, it seems to be most popular, regularization method is that of Tikhonov. In this method the estimate is being searched from the minimum of the functional

$$\Phi = ||N^{obs} - AN^{tr}||^2_W + \alpha ||N^{tr}||^2_{N^{tr}} \tag{8}$$

and has the form

$$\hat{N}^{tr} = \begin{cases} A^\top \tilde{G}^{-1} N^{obs}, & N \leq I \\ \tilde{G}^{-1} A^\top W N^{obs}, & N > I \end{cases}, \quad \text{where} \quad \tilde{G} = \begin{cases} G + \alpha W^{-1}, & N \leq I \\ G + \alpha I, & N > I \end{cases}. \tag{9}$$

Substitution of the expression N^{obs} from (1) into (9) gives for the mean of estimate $\bar{\hat{N}}^{tr}$ and its noise component $\hat{N}^{tr}_{\varepsilon}$

$$\bar{\hat{N}}^{tr} = \begin{cases} A^{\top}\tilde{G}^{-1}AN^{tr} \equiv A_R N^{tr}, & N \le I \\ \tilde{G}^{-1}A^{\top}WAN^{tr} \equiv A_R N^{tr}, N > I \end{cases}, \tag{10}$$

$$\hat{N}^{tr}_{\varepsilon} = \begin{cases} A^{\top}\tilde{G}^{-1}\varepsilon, & N \le I \\ \tilde{G}^{-1}A^{\top}W\varepsilon, N > I \end{cases}. \tag{11}$$

Expressions (10–11) permit to interpret the estimate $\hat{N}^{tr}$ as the result of new experiment with better resolution ($A \to A_R$) and increased noise $\hat{N}^{tr}_{\varepsilon}$. The matrix A_R can be called the "residual" apparatus function [3]. The difference of this matrix from unitary one characterizes the systematical error of estimate $\hat{N}^{tr}$, meanwhile expression (11) permits easy calculation of error matrix for $\hat{N}^{tr}$. If one uses the same method, expressions for A_R and error matrix for $\hat{N}^{tr}$ (7) can be obtained for the spectral window method. Several methods used for estimating continous distributions as well as the choice of regularization parameter were considered in our papers [4,5].

Meanwhile in high energy physics $W^{-1}_{nm} = N^{obs}_n \delta_{nm}$ (number of mesurements N is considered to be big enough), so expressions (10–11) for $N > I$ turn out to be approximate because the estimate in this situation is nonlinear. We shall obtain the linear estimate in the following way. Let

$$N^{tr} = RN^{obs}, \quad N^{tr}_{\varepsilon} = R\varepsilon, \tag{12}$$

where R is the matrix being searched for. With the use of Euclidean norms for N^{tr} and N^{obs} and Gilbert-Schmidt norm for R we obtain

$$Sp\Sigma = \overline{\|N^{tr}_{\varepsilon}\|^2} \le \|R\|^2_2 \overline{\|\varepsilon\|^2} = \|R\|^2_2 SpW^{-1}. \tag{13}$$

We look for matrix R from the condition of minimum of $\|R\|^2_2$ under condition of being unbiased by Lagrange method from stationarity of the functional

$$\Phi = SpR^{\top}R + Sp\Lambda^{\top}(RA - I). \tag{14}$$

where Λ is the matrix of Lagrange multipliers. The soultion of these problem leads to the estimate

$$\hat{N}^{tr} = (A^{\top}A)^{-1}A^{\top}N^{obs} = G^{-1}A^{\top}N^{obs}, \tag{15}$$

expressions (10–11) become exact.

Obtained results were proved in numerical tests.

Author thanks DELPHI collaboration at CERN for useful discussions and the Russian Foundation for Fundamental Research for financial support.

References

[1] V.B. Anykeyev, A.A. Spiridonov and V.P. Zhigunov \\ Nucl. Instr. and Meth. A322 (1992) 280.

[2] V.V. Vojevodin, Ju.A. Kuznetsov \\ Matrices and computations, Moscow, "Nauka", 1984.

[3] V.P. Zhigunov et al. \\ Nucl. Instr. and Meth. A273 (1988) 362.

[4] V.B. Anykeyev, A.A. Spiridonov and V.P. Zhigunov \\ Nucl. Instr. and Meth. A303 (1991) 350.

[5] V.B. Anikeev, V.P. Zhigunov \\ Phys. Part. Nucl. **24** (4), July-August 1993 424-452.