19. QUASI-INVERSION METHOD FOR SOLVING G-EQUATIONS

Suppose that $f(x)$ is a Borel function in R^{m_n} having partial derivatives of the third order. Let $\vec{x}_1, \dots, \vec{x}_n$ be independent observations of an m_n-dimensional vector $\vec{\xi}$, $\mathbf{E}\,\vec{\xi} = \vec{a}$. We need a consistent estimator of the value $f(\vec{a})$. Many problems of multivariate statistical analysis can be formulated in these terms. If f is a continuous function we take

$$\hat{\vec{a}} = n^{-1} \sum_{i=1}^{n} \vec{x}_i$$

as the estimator of $\vec{a}$. Then, obviously, for fixed m, $p\lim_{n\to\infty} f(\hat{\vec{a}}) = f(\vec{a})$. But the application of this method in solving practical problems is unsatisfactory due to the fact that the number of observations n necessary to solve the problem with a given accuracy increases sharply with m. It is possible to reduce significantly the number of observations n by making use of the fact that under some conditions, including $\lim_{n\to\infty} mn^{-1} = c$, $0 < c < \infty$, the relation

$$p \lim_{n\to\infty} [f(\hat{\vec{a}}) - \mathbf{E}\, f(\hat{\vec{a}})] = 0 \tag{19.1}$$

holds. We call (19.1) and similar identities the basic relations of the G-analysis of large dimensional observations. The methods of estimating functions of some characteristics of random vectors would be studied by this method.

19.1. G-equations for estimators of differentiable functions of unknown parameters

Suppose that vector $\vec{\xi}$ has a Normal distribution $N(\vec{a}, R_{m_n})$ and consider the functions

$$u\,(t, \vec{z}) = \mathbf{E}\, f\left(\vec{z} + \vec{a} + \vec{\nu} t^{1/2} n^{-1/2}\right), \tag{19.2}$$

where $t > 0$ is a real parameter, $\vec{z} \in R^{m_n}$, and $\vec{\nu}$ is a Normal $N(0, R_{m_n})$ random vector.

Suppose that the integrals

$$\mathbf{E}\, \frac{\partial^2}{\partial z_i \partial z_j} f\left(\vec{z} + \vec{a} + \vec{\nu} t^{1/2} n^{-1/2}\right)$$

exist. Let us find the differential equation for the function $u(t, \vec{z})$. We note that $\vec{\nu}(t + \Delta t)^{1/2} \approx \vec{\nu} t^{1/2} + \vec{\nu}_1 (\Delta t)^{1/2}$, where $\Delta t \geq 0$, $\vec{\nu}_1$ is a random vector which does not depend on the vector $\vec{\nu}$ and $\vec{\nu} \approx \vec{\nu}_1$. Then

$$\begin{aligned} \frac{\partial}{\partial t} u\,(t, \vec{z}) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbf{E} & \left[f\left(\vec{z} + \vec{a} + n^{-1/2}\left(\vec{\nu} t^{1/2} \quad + \vec{\nu}_1 (\Delta t)^{1/2}\right)\right) \right. \\ & \left. - f\left(\vec{z} + \vec{a} + n^{-1/2} \vec{\nu} t^{1/2}\right)\right]. \end{aligned}$$

Then, by using the expansion of the function f in a Taylor series

$$f\left(\vec{a} + \vec{h}\right) - f(\vec{a}) = \sum_{k=0}^{s} \left(\sum_{i=1}^{m_n} \frac{\partial}{\partial a_i} h_i \right)^k f(\vec{a}) + o\left(\left\| \vec{h} \right\| \right)$$

we obtain that the functions $u\,(t, \vec{z})$ satisfy the equation

$$\frac{\partial}{\partial t} u\left(t, \vec{z}\right) = A u\left(t, \vec{z}\right); \quad A = \frac{1}{2n} \sum_{i,j=1}^{m_n} r_{ij} \frac{\partial^2}{\partial z_i \partial z_j} \tag{19.3}$$

$$u\left(1, \vec{z}\right) = \mathbf{E}\, f\left(\vec{z} + \hat{\vec{a}}\right), \quad u\left(0, \vec{z}\right) = f\left(\vec{z} + \vec{a}\right),$$

where r_{ij} are the entries of the matrix R_{m_n}. Suppose that the random vector $\vec{\xi}$ has arbitrary distribution with $R_{m_n} = \mathbf{E}\left(\vec{\xi} - \vec{a}\right)\left(\vec{\xi} - \vec{a}\right)^T$. Let

$$\alpha_n\left(kn^{-1}, \vec{z}\right) = \mathbf{E}\, f\left\{\vec{z} + \vec{a} + n^{-1} \sum_{p=1}^{k} \left(\vec{x}_p - \mathbf{E}\, \vec{x}_p\right)\right\},$$

$$u_n\left(t, \vec{z}\right) = \alpha_n\left(kn^{-1}, \vec{z}\right), \quad kn^{-1} \leq t < (k+1)n^{-1}; \;\; k = 1, \ldots, n,$$

$$\lim_{n\to\infty} n\mathbf{E} \int_0^1 (1 - t^2) \left[\frac{1}{n} \sum_{i=1}^{m_n} \left(\vec{x}_i - \vec{a}_i\right) \left(\frac{\partial}{\partial z_i}\right)\right]^3$$
$$\times f\left\{\vec{z} + \vec{a} + \frac{1}{n} \sum_{i=1}^{k-1} \left(\vec{x}_i - \vec{a}_i\right) + \frac{t}{n}\left(\vec{x}_k - \vec{a}_k\right)\right\} \mathrm{d}t = 0.$$

Then, by using the expansion of the function f in a Taylor series, we obtain

$$\begin{aligned} & n\left[\alpha_n\left(\frac{k}{n}, \vec{z}\right) - \alpha_n\left(\frac{k-1}{n}, \vec{z}\right)\right] \\ & = \frac{1}{2n} \sum_{i,j=1}^{m_n} r_{ij} \frac{\partial^2}{\partial z_i \partial z_j} \alpha_n\left(\frac{k-1}{n}, \vec{z}\right) + \varepsilon_n, \end{aligned} \tag{19.4}$$

where $\lim_{n\to\infty} \varepsilon_n = 0$.

From equation (19.4) we have

$$u_n\left(t, \vec{z}\right) = u_n\left(0, \vec{z}\right) + \int_0^t \frac{1}{2n} \sum_{i,j=1}^{m_n} r_{ij} \frac{\partial^2}{\partial z_i \partial z_j} u_n\left(y, \vec{z}\right) \mathrm{d}y + \varepsilon_n. \tag{19.5}$$

19.2. ***G*-equation of higher orders**

Let $f(\vec{x}),\ \vec{x}\in R^{m_n}$ be the Borel function with mixed particular derivatives of order p inclusively; let $\vec{\xi},\ \mathbf{E}\,\vec{\xi}=\vec{a}$ be a certain m_n-dimensional random vector and let $\vec{x}_1,\dots,\vec{x}_n$ be independent observations of the vector $\vec{\xi}$.

If, for every $\vec{z}\in R^{m_n}$ and $k=1,\dots,n$

$$\lim_{n\to\infty} n\mathbf{E}\int_0^1 \frac{(1-t)^{p-1}}{(p-1)!}\left(\frac{1}{n}\sum_{i=1}^{m_n}(x_{ik}-a_i)\frac{\partial}{\partial z_i}\right)^p \times f\left(\vec{z}+\vec{a}+\frac{1}{n}\sum_{i=1}^{k-1}(\vec{x}_i-\vec{a})+\frac{t}{n}(\vec{x}_k-\vec{a})\right)\mathrm{d}t=0,$$

$$\sup_{\vec{z}\in R^{m_n}}\mathbf{E}\left|f\left(\vec{z}+\vec{a}+\frac{1}{n}\sum_{i=1}^{k-1}(\vec{x}_i-\vec{a})\right)\right|<\infty,$$

then

$$\begin{aligned}&\varphi_n(t,\vec{z})=f(\vec{z}+\vec{a})+\int_0^t B\varphi_n(y,\vec{z})\,\mathrm{d}y+\varepsilon_n;\\ &\varphi_n(1,\vec{z})=\mathbf{E}\,f\left(\vec{z}+\hat{\vec{a}}\right),\end{aligned}\tag{19.6}$$

where

$$\varphi_n(t,\vec{z})=\mathbf{E}\,f(\vec{z}+\vec{a}+\vec{\nu}_k),\ \ \frac{k}{n}\le t<\frac{k+1}{n};\ \ k=1,\dots,n-1,$$

$$\vec{\nu}_k=\frac{1}{n}\sum_{i=1}^{k}(\vec{x}_i-\vec{a}),$$

$$B=\sum_{l=1}^{p-1}\frac{1}{l!}\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{m_n}(x_{i1}-a_i)\frac{\partial}{\partial z_i}\right)^l.$$

19.3. ***G*-equation for functions of the empirical vector of expectations and the covariance matrix**

Let us find the G-equations for the differentiable functions $\varphi_n\left(\hat{\vec{a}},\hat{R}_{m_n}\right)$ of the empirical vector $\hat{\vec{a}}$ and the covariance matrix $\hat{R}_{m_n}$ which are obtained by independent normally distributed $N(\vec{a},R_{m_n})$ observations $\vec{x}_1,\dots,\vec{x}_n$.

Consider the functions

$$u_n(t,\vec{z},X_{m_n})=\varphi\left\{\vec{a}+\vec{z}+R_{m_n}^{1/2}\vec{\eta}_n n^{-1/2},R_{m_n}+X_{m_n} +R_{m_n}^{1/2}\sum_{s=1}^{k}\left(\frac{1}{n-1}\vec{\eta}_s\vec{\eta}_s^{\,T}-I\right)R_{m_n}^{1/2}\right\},$$

where $\vec{\eta}_s$ are independent m_n-dimensional random Normal law $N(0,I)$ vectors , and $X_{m_n}=(x_{ij})$ is a matrix of the parameters of the same order as the matrix R_{m_n}.

If the functions $u_n(t,\vec{z},X_{m_n})$ can be represented as

$$u_n\left(\frac{k}{n},\vec{z},X_{m_n}\right)-u_n\left(\frac{k-1}{n},\vec{z},X_{m_n}\right)=Au_n\left(\frac{k-1}{n},\vec{z},X_{m_n}\right)+\frac{\varepsilon_n}{n},$$

where

$$A=\frac{1}{2n}\sum_{i,j,p,l=1}^{m_n}\mathbf{E}\left(R_{m_n}^{1/2}\frac{\vec{\eta}_s\vec{\eta}_s^T-I}{n-1}R_{m_n}^{1/2}\right)_{ij}$$
$$\times\left(R_{m_n}^{1/2}\frac{\vec{\eta}_s\vec{\eta}_s^T-I}{n-1}R_{m_n}^{1/2}\right)_{pl}\frac{\partial^2}{\partial x_{ij}\partial x_{pl}}+\frac{1}{2n}\sum_{i,j=1}^{m_n}r_{ij}\frac{\partial^2}{\partial z_i\partial z_j};$$

then we obtain the equation

$$\psi_n\left(t,\vec{z},X_{m_n}\right)=\varphi\left(\vec{z}+\vec{a},X_{m_n}+R_{m_n}\right)$$
$$+\int_0^t A\psi_n\left(y,\vec{z},X_{m_n}\right)\mathrm{d}y+\varepsilon_n,$$
$$\psi_n\left(1,\vec{z},X_{m_n}\right)=\mathbf{E}\,\varphi\left(\vec{z}+\hat{\vec{a}},X_{m_n}+\hat{R}_{m_n}\right)$$

for the functions

$$\psi_n\left(t,\vec{z},X_{m_n}\right)=u_n\left(\frac{k}{n},\vec{z},X_{m_n}\right);\;\;\frac{k}{n}\le t<\frac{k+1}{n}.$$

19.4. *G*-equation for functions of empirical expectations

Let

$$u_n\left(kn^{-1},\vec{z}\right)=\mathbf{E}\,f\left(\vec{z}+\vec{a}+n^{-1}\sum_{p=1}^{k}\left(\vec{x}_p-\mathbf{E}\,\vec{x}_p\right)\right),$$

$$\psi_n\left(t,\vec{z}\right)=u_n\left(\frac{k}{n},\vec{z}\right),\;\;\frac{k}{n}\le t<\frac{k+1}{n};\;\;k=1,\dots,n.$$

If the limit exists,

$$\lim_{n\to\infty}\left\{n\left[u\left(\frac{k}{n},\vec{z}\right)-u\left(\frac{k-1}{n},\vec{z}\right)\right]-\theta\left(u\left(\frac{k}{n},\vec{z}\right)\right)\right\}=0,$$

where $\theta(y)$ is a certain continuous function on [0,1], then for the functions $\psi_n(t,\vec{z})$ we have

$$\psi_n\left(t,\vec{z}\right)=\varphi\left(\vec{z}+\vec{a}\right)+\int_0^t\theta\left\{\psi_n\left(y,\vec{z}\right)\right\}\mathrm{d}y+\varepsilon_n.$$

We deduce the finding of G-estimators of the functions $f\left(\vec{a}\right)$ to solution of the inverse problem for equation (19.5). The latter consists of finding $\alpha_n\left(0,z\right)$ by the function $\alpha_n\left(1,z\right)$, which is replaced by the function $f\left(\vec{z}+\hat{\vec{a}}\right)$ based on observations of the random vector $\vec{\xi}$. Of course, the solution of the inverse problem with such a replacement cannot exist in the class of functions $W_2^{(0,2)}$. Therefore, it appears expedient to find a generalized solution of the estimation problem of function $f\left(\vec{a}\right)$.

Let $\psi\left(\vec{x}\right)\in L_2$ and let the functional

$$I(\varphi) = \int_D |\alpha_n(1, \vec{x}, \varphi(\cdot)) - \varphi(\vec{x})| \mathrm{d}\vec{x} \tag{19.7}$$

be determined by the functions $\varphi(\vec{x}) \in W_2^{(0,2)}$. Here D is a domain on m-dimensional Euclidean space, which is bounded by the piecewise smooth surface S, and $\alpha_n(1, \vec{x}, \varphi(\cdot))$ is the solution of the equation

$$\alpha_n(t, \vec{x}, \varphi(\cdot)) = \varphi(\vec{x}) + \int_0^t \frac{1}{2n} \sum_{i,j=1}^m r_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \alpha_n(u, \vec{x}, \varphi(\cdot)) \, \mathrm{d}u + o(1),$$

at the point $t = 1$. The function $\hat{\varphi}(\vec{x})$ is the solution of the inverse problem if

$$\inf_{\varphi(\cdot) \in W_2^{(0,2)}} I(\varphi) = I(\hat{\varphi}).$$

To solve this problem, we proceed as follows. First, we solve the direct problem

$$\alpha_n(t, \vec{x}, \varphi(\cdot)) = \varphi(\vec{x}) + \int_0^t A\alpha_n(u, \vec{x}, \varphi(\cdot)) \, \mathrm{d}u + o(1),$$

where

$$A = \frac{1}{2n} \sum_{i,j=1}^m r_{ij} \frac{\partial^2}{\partial x_i \partial x_j}, \;\; \alpha_n(u, \vec{x}, \varphi(\cdot)) = 0, \;\; \in S.$$

Here S is the piecewise smooth boundary of a connected domain D and

$$\alpha_n(1, \vec{x}, \varphi(\cdot)) = \psi(\vec{x})$$

is a given function. Then we have an approximate value for the initial condition of the function $\varphi(x)$. It is quite possible that, in general, such a problem has no solution for the given function. Therefore, it is appropriate to solve the inverse problem approximately with the help of the so-called quasi-inversion method. Thus, we consider the following equation

$$\frac{\partial u(t, \vec{z})}{\partial t} = A_\delta u(t, \vec{z}), \;\; u(1, \vec{z}) = \alpha_n(1, \vec{z}) \tag{19.8}$$

instead of equation (19.5); here A_δ is some operator similar in some sense, to the operator A and such that the solution of equation (19.8) is stable. We can choose

$$A_\delta = A + \delta A^2, \;\; \delta > 0.$$

19.5. Estimator G_{19} of regularized function of unknow parameters

By obtaining the solution of equation (19.6), we can apply the spectral theory of the operator A_δ. Its spectrum is, however, continuous. Therefore, it would be better to replace operator A by an operator A_ε, such that its spectrum is discrete and whose eigenfunctions form the complete orthonormal basis in the Hilbert space L_2. For example, instead of such an operator A, we can choose

$$A_\varepsilon = A + \varepsilon q(\vec{z}) + \delta \left[A + \varepsilon q(\vec{z})\right]^2, \ \ \varepsilon, \delta > 0,$$

where $q(\vec{z})$ is any measurable function such that the operator $A + \varepsilon q(\vec{z}), \ \vec{z} \in R^m$ satisfies the above mentioned condition. From the operator spectral theory, it follows that instead of function $q(\vec{z})$ we can choose any measurable function such that

$$\lim_{\|\vec{z}\| \to \infty} q(\vec{z}) = \infty.$$

Let $\lambda_k(\varepsilon)$ and $\varphi_{k\varepsilon}(\vec{z}), \ k = 1, 2, ...$ denote the eigenvalues and eigenfunctions of the operator $A + \varepsilon q(\vec{z}), \ \vec{z} \in R^m$, respectively Now we can give the main form of G_{19}-estimators of function $f(\vec{a})$;

$$\begin{aligned} G_{19} &= \exp\left\{A_\delta - \varepsilon A_\delta^2\right\} f\left(\hat{\vec{\alpha}} + \vec{z}\right)_{\vec{z}=0} \\ &= \sum_{k=0}^{\infty} \exp\left\{\lambda_k(\varepsilon) - \delta \lambda_k^2(\varepsilon)\right\} \int f\left(\hat{\vec{\alpha}} + \vec{z}\right) \varphi_k(\vec{z}) \, \mathrm{d}\vec{z} \varphi_k\left(\vec{0}\right), \end{aligned}$$

where

$$A_\varepsilon = \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2}{\partial z_i \partial z_j} \mathbf{E} \left(\hat{\vec{\alpha}} - \vec{\alpha}\right)_i \left(\hat{\vec{\alpha}} - \vec{\alpha}\right)_j + \varepsilon q(\vec{z}); \quad \varepsilon > 0, \quad \delta > 0,$$

and $q(\vec{z})$ is any continuous function satisfying the condition

$$\liminf_{n \to \infty} \lim_{\|\vec{z}\| \to \infty} q(\vec{z}) = \infty.$$

Random Matrices. Kiev University Publishing, Ukraine, 1975, 448pp. (in Russian).

Theory of Random Determinants. Kiev University Publishing, Ukraine, 1975, 368pp. (in Russian).

Limit Theorems for Functions of Random Variables. "Higher School" Publishing, Kiev, Ukraine, 1983, 207pp. (in Russian).

Multidimensional Statistical Analysis. "Higher School" Publishing, Kiev, Ukraine, 1983, 320pp. (in Russian).

Spectral Theory of Random Matrices. "Science" Publishing, Moscow, Russia, 1988, 376pp. (in Russian).

Theory of Random Determinants. Kluwer Academic Publishers, The Netherlands, 1990, 677pp.

Theory of Systems of Empirical Equations. "Lybid" Publishing, Kiev, Ukraine, 1990, 264pp. (in Russian).

Statistical Analysis of Observations of Increasing Dimensions. Kluwer Academic Publishers, The Netherlands, 1995, 286pp.

Theory of Linear Algebraic Equations with Random Coefficients. Allerton Press, Inc, New York, U.S.A. 1996, 320pp.

An Introduction to Statistical Analysis of Random Arrays. VSP, Utrecht, The Netherlands. 1998, 673pp.

Theory of Stochastic Canonical equations. Volume I and II. Kluwer Academic Publishers, (xxiv + 497, xviii + 463).(2001).