ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 194, 2005

Janusz Wywiał[*]

ON ESTIMATION OF DOMINANT OF MULTIDIMENSIONAL RANDOM VARIABLE

Abstract

The problem of estimation of the mode of a continuous distribution function of multidimensional random variable is considered. The estimator of the mode is the vector of means from appropriately truncated sample. The truncation sample is obtained through rejecting the observation in such a way that the measure of skewnees of multidimensional variable takes value as close zero as possible. We can expect that through successful truncation of the sample the vector of sample means approach to vector of modes of multidimensional variable. The estimator constructed in such a way is usually biased estimators of the mode. Moreover, the biased estimators of values of modal regressions are proposed. The well known "jackknife" procedure is proposed to evaluate the mean square errors of the estimators.

Key words: moments of truncated distribution, estimation of distribution mode.

I. ESTIMATION OF MODE IN ONE DIMENSIONAL CASE

Let $X_{(1)} \leqslant X_{(2)} \leqslant ... \leqslant X_{(n)}$ be the sequence of the order statistics from a simple sample drawn from one-dimensional distribution. The third central moment from the right hand truncated simple sample is defined by the expression:

$$M_3(X_{(k)}) = \frac{1}{k} \sum_{i=1}^{k} (X_{(i)} - \overline{X}_k)^3, \quad 1 < k \leqslant n$$

where: $\overline{X}_k = \frac{1}{k} \sum_{i=1}^{k} X_{(i)}$. Particularly, $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n$.

[*] Professor, Department of Statistics, University of Economics in Katowice.

Let us assume that $M_{(3)}(X_{(n)}) = M_3 > 0$. The sample quasi-mode D_q is defined as follows:

$$G_e = \overline{X}_e, \text{ if } M_3(X_{(e)}) \leqslant 0 \text{ and } M_3(X_{(k)}) > 0 \text{ for } k = e + 1, \ ..., \ n.$$

The statistic Dq will be usually a biased estimator of the mode γ.

Theorem 1 (Johnson and Rogers, 1951). Let the density function of a random variable X be one-modal end it is concave from the left side of the dominant γ and it is convex from the right side of the dominant γ. Moreover, let $D^2(X) > 0$. There exists one dimensional probability distribution if and only if $(E(X) - \gamma)^2 \leqslant 3 \ D^2(X)$.

Hence, the well known Pearson coefficient of skewnees is bounded.

Let us assume that probability distribution of X is right side truncated in the point a and $E(X|a) = \gamma(a)$, $D^2(X|a)$ and $\eta_3(X|a)$ *be the expected value, the variance and the third central moment of the truncated distribution, respectively*. The theorem 1 lead to the following. If $\eta_3(X|a) \to 0$ and $D^2(X|a)$ do not increases then the $(E(X|a) - \gamma)^2$ do not increase. Hence, the appropriate truncation of the probability distribution can lead to decreasing the distance between mode and the expected value of the truncated probability distribution. This lead to the conclusion that the estimator G_e is closer to the mode than the common mean $\overline{X}$ from non-truncated simple sample. That is why the mean from the truncated sample can be used to estimation the mode.

II. ESTIMATION OF MODE IN MULTI-DIMENSIONAL CASE

Let $f(x_1, \ ..., \ x_k) = f(\mathbf{x})$ where $\mathbf{x} = [x_1...x_k]$, be a density function of an k-dimensional random variable defined in R^k. The vectors of expected values and variances are denoted by $\boldsymbol{\mu} = [\mu_1...\mu_k]$ and $\boldsymbol{\sigma}^2 = [\sigma_1^2...\sigma_k^2]$. The mode of the k-dimensional random variable is denoted by $\gamma = [\gamma_1...\gamma_k]$ and $f(\gamma) = \text{maximum}$. The central moments of the order 3 of the r-dimensional variable are denoted by:

$$\eta_{uwz}(X_i, X_j, X_g) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} (x_i - \mu_i)^u (x_j - \mu_j)^v (x_g - \mu_g)^g f(x_1, ..., x_k) dx_1 ... dx_k,$$

$$\eta_u(X_i) = \eta_{u00}(X_i, X_j, X_g)$$

Let us consider the truncated multidimensional variable of the following density function:

$$f(x_1, \ldots, x_k|A) = \frac{f(x_1, \ldots, x_k)}{P([X_1, \ldots, X_k] \in \mathrm{A})}$$

where A is a convex region and $\mathrm{A} \subseteq \mathrm{R}^k$. Hence, $f(x_1, \ldots, x_k|\mathrm{A})$ is density function of truncated distribution. The vectors of expected values and variances of the truncated distributions are as follows: $\boldsymbol{\mu}(\mathrm{A}) = [\mu_1(\mathrm{A}) \ldots \mu_k(\mathrm{A})]$, $\boldsymbol{\sigma}^2(\mathrm{A}) = [\sigma_1^2(\mathrm{A}) \ldots \sigma_k^2(\mathrm{A})]$, where:

$$\mu_i(\mathrm{A}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_i f(x_1, \ldots, x_k|\mathrm{A}) dx_1 \ldots dx_k,$$

$$\sigma_i^2(\mathrm{A}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_i - \mu_i(\mathrm{A}))^2 f(x_i, \ldots, x_k|\mathrm{A}) dx_1 \ldots dx_k,$$

$$\eta_{uwz}(X_i, X_j, X_g|\mathrm{A}) =$$

$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_i - \mu_i(\mathrm{A}))^u (x_j - \mu_j(\mathrm{A}))^v (\mathrm{x}_g - \mu_g(\mathrm{A}))^g f(x_1, \ldots, x_k|\mathrm{A}) dx_1 \ldots dx_k.$$

Let us introduce the following notation:

$$\theta = [\theta_1 \ \theta_2 \ \theta_3] \text{ is of dimensions } 1 \times (k^2 + k(k-1)(k-2)/6),$$

where

$$\theta_1 = [\eta_3(X_1) \ \eta_3(X_1) \ \ldots \ \eta_3(X_k)] \text{ of dimensions } 1 \times k,$$

$$\theta_2 = [\eta_{12}(X_1, X_2) \ \eta_{21}(X_1, X_2) \ \ldots \ \eta_{21}(X_{k-1}, X_k) \ \eta_{21}(X_k, X_{k-1})]$$
$$\text{of dimensions } 1 \times k(k-1),$$

$$\theta_3 = [\eta_{111}(X_1, X_2, X_3) \ \eta_{111}(X_1, X_2, X_4) \ \ldots \ \eta_{111}(X_{k-3}, X_{k-1}, X_k)$$
$$\eta_{111}(X_{k-2}, X_{k-1}, X_k)] \text{ of dimensions } 1 \times \frac{1}{6} k(k-1)(k-2),$$

The vector can be estimated by the vector $\mathbf{L} = [\mathbf{L}_1 \, \mathbf{L}_2 \, \mathbf{L}_3]$, where:

$$\mathbf{L}_1 = [C_{30}(X_1, \ X_2) \ C_{03}(X_2, \ X_1) \ \ldots \ C_{03}(X_{k-1}, X_k)],$$

$$\mathbf{L}_2 = \ [C_{12}(X_1, X_2) \ C_{21}(X_1, X_2) \ \ldots \ C_{21}(X_{k-1}, X_k) \ C_{21}(X_k, X_{k-1})],$$

$$\mathbf{L}_3 = [C_{111}(X_1, X_2, X_3) \ C_{111}(X_1, X_2, X_4) \ \ldots \ C_{111}(X_{k-3}, X_{k-1}, X_k)$$
$$C_{111}(X_{k-2}, X_{k-1}, X_k)].$$

Similarly, we define the vectors of moments of truncated distribution:

$$\boldsymbol{\theta}(\mathrm{A}) = [\boldsymbol{\theta}_1(\mathrm{A})\ \boldsymbol{\theta}_2(\mathrm{A})\ \boldsymbol{\theta}_3(\mathrm{A})],$$

where

$$\boldsymbol{\theta}_1(\mathrm{A}) = [\eta_i(X_1|\mathrm{A})],\ \boldsymbol{\theta}_2(\mathrm{A}) = [\eta_{12}(X_i, X_j|\mathrm{A})]$$

and

$$\boldsymbol{\theta}_3(\mathrm{A}) = [\eta_{111}(X_i, X_j, X_t|\mathrm{A})].$$

It is well known that if a distribution function is symmetric, then all central moments of the order 3 of the marginal one or two dimensional distributions are equal to zero and $\boldsymbol{\theta} = 0$. In the case when $\boldsymbol{\theta} \neq 0$ we can find such set $\mathrm{A} = \{\mathrm{A}: \mathrm{A} \subseteq \mathrm{R}^k$ and $\boldsymbol{\theta}(\mathrm{A}) = 0\}$. The vector of mean values $\boldsymbol{\mu}(\mathrm{A}) = [\mu_1(\mathrm{A})...\mu_k(\mathrm{A})]$ from truncated distribution we define as the A-mode of this distribution and it will be denoted by $\gamma(\mathrm{A})$. It is obvious that the A-mode can not necessary be equal to the mode of the entire distribution. The conditions of such equality can be stated quite simple only in the case of one-dimensional distribution, see Wywiał (2000a–2000b). That is why the parameter $\gamma(\mathrm{A})$ can be called a quasi-mode as in one-dimensional case.

The simple sample of size n is denoted by $\mathbf{X} = [\mathbf{X}*_1, ..., \mathbf{X}*_n]$, where: $\mathbf{X}*_t = \begin{bmatrix} X_{1t} \\ ... \\ X_{kt} \end{bmatrix}$, $t = 1, ..., n$. The sample moments are as follows:

$$C_{uvz}(X_i, X_j, X_g) = \frac{1}{n}\sum_{t=1}^{n}(X_{it} - \overline{X}_i)^u(X_{jt} - \overline{X}_j)^v(X_{gt} - \overline{X}_g)^z, \quad \overline{X}_i = \frac{1}{n}\sum_{t=1}^{n} X_{it}$$

Let $\mathbf{X}_{*(h)}$ $(h = n, ..., 1)$ be such vector that

$$d(\mathbf{X}*_{(h)}, \overline{\mathbf{X}}*_{(h)}) = \underset{i \in F_{(h)}}{\mathrm{maximum}}\{Q_{i,h}\}$$

where:

$$F_{(h)} = \{t : t = 1, ..., n \quad \text{and} \quad \mathbf{X}*_t \neq \mathbf{X}*_{(p)}, \quad \text{for} \quad p > h]$$

$$Q_{i(h)} = (\mathbf{X}*_1 - \overline{\mathbf{X}}*_{(h)})^T(\mathbf{X}*_i - \overline{\mathbf{X}}*_{(h)}),$$

$$\overline{\mathbf{X}}*_{(h)} = \frac{1}{\mathbf{n}_{(h)}}\sum_{t \in F_{(h)}} \mathbf{X}*_t, \quad n_{(h)} = \mathrm{Card}\{F_{(h)}\}$$

$$\overline{X}*_{(h)} = [\overline{X}_{1(h)}...\overline{X}_{k(h)}], \quad \overline{\mathbf{X}}*_{(n)} = \overline{\mathbf{X}} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}*_t.$$

A truncated sample is identified by the set $F_{(h)}$.

Let

$$C_{uv}(X_i,X_j,X_g|h)=\frac{1}{n_{(h)}}\sum_{t\in F_{(h)}}(X_{it}-\overline{X}_{i(h)})^u(X_{jt}-\overline{X}_{j(h)})^v(X_{gt}-\overline{X}_{g(h)})^z,$$

$$\mathbf{L}_{(h)}=[\mathbf{L}_{1(h)}\ \mathbf{L}_{2(h)}\ \mathbf{L}_{3(h)}]=[L_{1(h)}\ L_{2(h)}\ \ldots\ L_{w(h)}],\ \text{where}$$

$$w=1\times(\mathrm{k}^2+\mathrm{k}(\mathrm{k}-1)(\mathrm{k}-2)/6)\quad \text{and}$$

$$\mathbf{L}_{1(h)}=[C_{30}(X_1,\ X_2|h)\ \ C_{03}(X_2,\ X_1|h)\ \ldots\ C_{03}(X_{k-1},\ X_k|h)],$$

$$\mathbf{L}_{2(h)}=[C_{12}(X_1,\ X_2|h)\ \ C_{21}(X_1,\ X_2|h)\ \ldots\ C_{21}(X_{k-1},\ X_k|h)\ \ C_{21}(X_k,\ X_{k-1}|h)],$$

$$\mathbf{L}_{3(h)}=[C_{111}(X_1,\ X_2,\ X_3|h)\ \ C_{111}(X_1,\ X_2,\ X_4|h)\ \ldots$$
$$C_{111}(X_{k-3},\ X_{k-1},\ X_k|h)\ C_{111}(X_{k-2},\ X_{k-1},\ X_k|h)].$$

The vector parameter γ can be estimated bz means of the statistic $\mathbf{G}_1(u)$ determined bz the following procedure:.

Let

$$\mathbf{Z}_{(h)}=[L_{1(h)}^2\ L_{2(h)}^2\ldots L_{w(h)}^2]$$

$$\mathbf{L}*_{(h)}=\underset{i=1,\ldots,w}{\text{maximum}}\,\{L_{i(h)}^2\}$$

$$\mathbf{G}_{1(u)}=\mathbf{X}*_{(u)}:\quad u=\max\{h:\ h=n,\ldots,1\quad\text{and}\quad L*_{(h)}\leqslant \mathrm{e}\}$$

where $\mathrm{e}\geqslant 0$. It seems that e should be assigned in such a way that the size $n_{(u)}$ of the truncated sample is sufficiently large.

The next estimator is based on measure of multivariate skewness defined by Mardia (1970). Let $\sigma^{ij}(\mathrm{A})$ be an $(i,\,j)$ element of the matrix $\mathbf{\Sigma}^{-1}(\mathrm{A})$ where $\mathbf{\Sigma}(\mathrm{A})=[\sigma_{ij}(\mathrm{A})]$, $\sigma_{ij}(\mathrm{A})=\eta_{11}(X_i,X_j|\mathrm{A})$, is the variance-covariance matrix of the truncated distribution. In the case of truncated distribution the skewness coefficient can be written in the following way.

$$\beta(\mathrm{A})=\sum_{\{i_1,i_2,i_3\}}\ \sum_{\{j_1,j_2,j_3\}}\sigma^{i_1j_1}(\mathrm{A})\sigma^{i_2j_2}(\mathrm{A})\sigma^{i_3j_3}(\mathrm{A})\eta_{111}(X_{i_1},X_{i_2},X_{i_3}|\mathrm{A})\eta_{111}$$
$$(X_{j_1},X_{j_2},X_{j_3}|\mathrm{A})$$

where $\{i_1,\ i_2,\ i_3\}$ and $\{j_1,\ j_2,\ j_3\}$ are variations with replications. Each variation is determined on the basis of sequence: 1, 2, ..., k.

The coefficient of skewness from the truncated sample $F_{(h)}$ is as follows:

$$\mathrm{B}_{(h)} = \frac{1}{n_{(h)}^2} \sum_{i,j \in F_{(h)}} ((\mathbf{X}*_i - \overline{\mathbf{X}}*_{(h)})^T \mathbf{S}_{(h)}^{-1} (\mathbf{X}*_i - \overline{\mathbf{X}}*_{(h)}))^3$$

where:

$$\mathbf{S}_{(h)} = [C_{11(h)}(X_i, X_j)],\ \mathrm{C}_{11(h)}(X_i, X_j) = \frac{1}{n_{(h)}} \sum_{t \in F_{(h)}} (X_{it} - \overline{X}_{i(h)})(X_{jt} - \overline{X}_{j(h)}).$$

The second estimator of quasi-mode is as follows:

$$\mathbf{G}_{2(v)} = \mathbf{X}*_{(v)}: \quad v = \max\{h: h = n, ..., 1 \quad \text{and} \quad \mathrm{B}_{(h)} \leqslant \mathrm{e}\}$$

The statistics $\mathbf{G}_{1(u)}$ and $\mathbf{G}_{1(u)}$ can be biased estimators of the mode γ. Their variances can be estimated using the well known method of jackknife.

Let us assume that $\boldsymbol{\theta}_1 = [\eta_3(X_1)\ \eta_3(X_1)\ ...\ \eta_3(X_k)] > \mathbf{0}$ for the values $\mathbf{x} \in \mathbf{B} \subseteq \mathbf{R}^k$ of a multidimensional random variable X. Let $\mathrm{A}_{\#} \subset \mathrm{B}$ be such the sup-set that $\boldsymbol{\theta}_1(\mathrm{A}_{\#}) = 0$. If $\mathrm{D}^2(X_i) \geqslant \mathrm{D}^2(X_i|A_{\#})$ for each $i = 1, ..., k$ and $\mathrm{D}^2(X_j) > \mathrm{D}^2(X_j|\mathrm{A}_{\#})$ for at least one index j where $j = 1, ..., k$ then

$$\sum_{i=1}^{k} (\mathrm{E}(X_i|\mathrm{A}_{\#}) - \gamma_i)^2 \leqslant \sum_{i=1}^{k} (\mathrm{E}(X_i) - \gamma_i)^2.$$

It means that the appropriate truncation of the multivariate probability distribution can lead to decreasing the distance between mode and the expected value of the truncated probability distribution. This leads to the conclusion that in special cases the expected values of the above introduced estimators can be closer to the mode than the vector of means from non-truncated simple sample.

Some simulation methods should be used to analysis of the accuracy of the proposed estimators. The similar estimators can be constructed on the basis of other coefficients of multivariate skewness e.g. like those proposed by Wywiał (1983, 1985).

III. MODAL REGRESSION

Let us consider the following regression model:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}$$

where $\mathbf{Y}^T = [Y_1\ ...\ Y_n]$ is the vector of independent random variables. The matrix of non-random values of explanatory variables is denoted by $\mathbf{X}$ and it is of dimension $n \times r$. The vector of parameters $\mathbf{B}$ has dimensions $r \times 1$.

Moreover, E(U) = 0 and variance-covariariance matrix of **U** is diagonal and $D^2(U_i) = \sigma^2$ and $\eta_3(U_i) > 0$ for each $i = 1, ..., n$. The modal values of the variables $\mathbf{Y}^T = [Y_1 ... Y_n]$ will be denoted by $\gamma^T = [\gamma_1 ... \gamma_n]$ and the modal values of the variables $\mathbf{U}^T = [U_1 ... U_n]$ are the same and equal to κ. The modal regression is defined in the following way:

$$\gamma = \mathbf{XB} + \kappa \mathbf{J}_n$$

where $\mathbf{J}_n$ is the vector of dimensions $n \times 1$.

Under these assumption $E(Y) = \mathbf{XB}$ can be estimated by means of the statistic:

$$\hat{\mathbf{Y}} = \mathbf{XB}$$

where $\hat{\mathbf{B}}$ is unbiased estimator of **B** obtained by the well known method of least squares and

$$\hat{\mathbf{B}} = (\mathbf{X}^T \ \mathbf{X})^{-1} \ \mathbf{X}^T \mathbf{Y}$$

It is obvious that $E(\hat{\mathbf{Y}}) = E(\mathbf{Y})$

The vector of residual is as follows:

$$\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{MY}$$

where

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \ \mathbf{X})^{-1} \ \mathbf{X}^T, \ \mathbf{I}_n \text{ is unit matrix of degree } n.$$

The problem is determining the vector γ of modal values. Firslty, as suggested Pawłowski (1973), the mode κ should be estimated by means of a statistic $\hat{\mathbf{G}}$ which is a function of the residuals $\hat{\mathbf{U}}$. This leads to the following estimator of the regression

$$\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} + \hat{\mathbf{G}}\mathbf{J}_n = \mathbf{X}\hat{\mathbf{B}} + \hat{\mathbf{G}}\mathbf{J}_n$$

where $\mathbf{J}_n$ is the vector of dimensions $n \times 1$. In order to simplify the consideration let us assume that $\kappa > 0$. The statistic $\hat{\mathbf{G}}$ can take form of the estimator of mode proposed in the first. Let $\hat{U}_{(1)} \leqslant \hat{U}_{(2)} \leqslant ... \leqslant \hat{U}_{(n)}$ be the sequence of the order statistics section. The third central moment from the right hand truncated sample is defined by the expression:

$$M_3(\hat{U}_{(k)}) = \frac{1}{k}\sum_{i=1}^{k}(\hat{U}_{(i)} - \hat{U}_{\#k})^3, \quad 1 < k \leqslant n$$

where:

$$\hat{U}_{\#k} = \frac{1}{k}\sum_{i=1}^{k}\hat{U}_{(i)}$$

When we assume that $M_3(\hat{U}_{(n)}) = M_3(\hat{U}) > 0$, the sample quasi-mode $\hat{G}$ is defined as follows:

$$\hat{G} = \hat{U}_{\#e}, \text{ if } M_3(\hat{U}_{\#e}) \leqslant 0 \text{ and } M_3(\hat{U}_{\#e}) > 0 \text{ for } k = e + 1, \ldots, n$$

The statistic $\tilde{Y}$ is biased estimator of γ because $\hat{G}$ is usually biased estimator of κ as it was discussed in the previous paragraph.

The idea of construction of the next estimator is based on the truncated least-square method proposed by Ruppert and Carrolla (1980). Let $\mathbf{Y}_{(e)}$ is consisted of those variables of the vector $\mathbf{Y}$ which indexes are equal to appropriate indexes of residuals in the set $\{\hat{U}_i: i = 1, \ldots, n; \hat{U}_i \leqslant \hat{U}_{(e)}\}$ where evaluation of $\hat{U}_{(e)}$ was showed above. In the same way the submatrix $\mathbf{X}_{(e)}$ of the observations matrix $\mathbf{X}$ is determined. Next, we again evaluate the parameters of the linear regression:

$$\mathbf{B}_{(e)} = (\mathbf{X}_{(e)}^T\mathbf{X}_{(e)})^{-1}\mathbf{X}_{(e)}^T\mathbf{Y}_{(e)}$$

If $e > r$, the next estimator of the modal regression is as follows:

$$\tilde{\mathbf{Y}}_{(e)} = \mathbf{X}_{(e)}\mathbf{B}_{(e)}$$

In general case the statistic $\tilde{\mathbf{Y}}_{(e)}$ is not unbiased estimator of the modal regression.

Let us note that estimators $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Y}}_{(e)}$ can be easy adopted to evaluation conditional prognosis of the variable under study for some fixed values of auxiliary variables. Let $\mathbf{x}$ be row vector (of dimensions $r \times 1$) of values of auxiliary variables. The two predictors of the value $\gamma(\mathbf{x})$ of modal regression for given vector x are as follows:

$$\tilde{\mathbf{Y}}(\mathbf{x}) = \mathbf{x}\hat{\mathbf{B}} + \hat{\mathbf{G}} \quad \text{or} \quad \tilde{\mathbf{Y}}_{(e)}(\mathbf{x}) = \mathbf{x}\mathbf{B}_{(e)}$$

The mean square error of the estimators or predictors of the modal regression values and the parameters can be evaluated on the basis of the jackknife method. The accuracy of the both estimators or predictors of the modal regression can be analysed on the basis of appropriate simulation studies.

REFERENCES

Johnson N.L., Rogers C.A. (1951), The moment problem for unimodal distributions, *Annals of Mathematical Statistics*, **22**, 433-439.

Mardia K.V. (1970), Measures of multivariate skewneess and kurtosis with application, *Biometrika*, **57**, 3, 519-530.

Pawłowski Z. (1973), *Prognozy ekonometryczne*, PWN, Warszawa.

Ruppert D., Carroll R.J. (1980), Trimmed least squares estimation in the linear model, *Journal of the American Statistical Association*, **75**, 828-838.

Wywiał J. (1983), Normalized coefficients of deviation from multinormal distribution (in Polish). *Przegląd Statystyczny*, **30**, 77-83.

Wywiał J. (1985), *Test normalności dla wielowymiarowej zmiennej losowej dla dużych prób*, (A test for normality of a multidimensional random variable in the large sample case; in Polish), *Przegląd Statystyczny*, **32**, 355-364.

Wywiał J. (2000a), Estimation of distribution function mode on the basis of sample moment or sample median, *Badania Operacyjne i Decyzje*, no 2, 89-98.

Wywiał J. (2000b), Estimation of mode on the basis of a truncated sample, *Acta Universitatis Lodziensis. Folia Oeconomica*, **152**, 73-81.

Janusz Wywiał

O ESTYMACJI DOMINANTY WIELOWYMIAROWEJ ZMIENNEJ LOSOWEJ

Streszczenie

Praca dotyczy problemu estymacji dominanty rozkładu prawdopodobieństwa wielowymiarowej zmiennej losowej. Zajmowano się problemem estymacji dominanty zmiennej losowej ciągłej. Analizowano jednomodalne rozkłady prawdopodobieństwa.

W niniejszej pracy analizowano głównie klasę rozkładów prawdopodobieństwa zmiennych losowych charakteryzującą się tym, że zachodzą dla nich pewne nierówności między wartościami oczekiwanymi i dominantą rozkładu funkcją jego trzecich momentów centralnych. Dla takiej klasy rozkładów są wprowadzone dwa estymatory dominanty, których wyznaczanie w praktyce ma charakter iteracyjny. Z grubsza rzecz biorąc, wyliczenie wartości estymatora dominanty wiąże się z sukcesywnym obcinaniem obserwacji próby do chwili, gdy zostaną spełnione pewne warunki, a w szczególności, że pewna funkcja trzech mieszanych momentów centralnych rozkładu uciętego osiągnie wartość zero. Wówczas oceną punktu będącego dominantą rozkładu wielowymiarowego są właśnie średnie z uciętych rozkładów brzegowych z próby. Proponowane estymatory mogą być obciążone. W niektórych przypadkach dało się oszacować maksymalny poziom takiego obciążenia. Proponowane są również dwa estymatory regresji modalnej.