M$_{\mathcal{P}}$ estimation applied to platykurtic sets of geodetic observations

Zbigniew Wiśniewski

University of Warmia and Mazury
Institute of Geodesy
1 Oczapowskiego St., 10-957 Olsztyn, Poland
e-mail: zbyszekw@uwm.edu.pl

Abstract: M$_{\mathcal{P}}$ estimation is a method which concerns estimating of the location parameters when the probabilistic models of observations differ from the normal distributions in the kurtosis or asymmetry. The system of Pearson's distributions is the probabilistic basis for the method. So far, such a method was applied and analyzed mostly for leptokurtic or mesokurtic distributions (Pearson's distributions of types IV or VII), which predominate practical cases. The analyses of geodetic or astronomical observations show that we may also deal with sets which have moderate asymmetry or small negative excess kurtosis. Asymmetry might result from the influence of many small systematic errors, which were not eliminated during preprocessing of data. The excess kurtosis can be related with bigger or smaller (in relations to the Hagen hypothesis) frequency of occurrence of the elementary errors which are close to zero. Considering that fact, this paper focuses on the estimation with application of the Pearson platykurtic distributions of types I or II. The paper presents the solution of the corresponding optimization problem and its basic properties.

Although platykurtic distributions are rare in practice, it was an interesting issue to find out what results can be provided by M$_{\mathcal{P}}$ estimation in the case of such observation distributions. The numerical tests which are presented in the paper are rather limited; however, they allow us to draw some general conclusions.

Keywords: M and M$_{\mathcal{P}}$ estimation, platykurtic probabilistic models, Pearson's distributions

1. Introduction

Considering the classical theory of measurement errors, we usually assume that the Gauss distributions (the normal distributions) are their probabilistic models. The family of such distributions corresponds with the hypothesis of the elementary errors given by Hagen and Bessel (e.g. Fischer, 2011). However, the analyses show that the empirical distributions of errors of geodetic, geophysical or astronomical observations might often differ from the normal ones. The basic anomalies in this context concern

Pearson's squared skewness $\beta_1 = \mu_3^2 / \mu_2^3$ and/or the kurtosis $\beta_2 = \mu_4 / \mu_2^2$ (μ_k – the kth central moment). Besides the coefficient β_1, one can also apply the skewness $\gamma_1 = \mu_3 / \mu_2^{3/2} = \mathrm{sgn}(\mu_3)\sqrt{\beta_1}$, which allows us to determine the sign of the asymmetry (positive or negative). Note that for the normal distributions $\beta_1 = 0$ and $\beta_2 = 3$. Due to such a value of the kurtosis, anomalies of other distributions in this context are often described by the excess kurtosis $\gamma_2 = \beta_2 - 3$ (e.g. Dorić et.al., 2009).

Asymmetry might result from the influence of many small systematic errors, which were not eliminated during preprocessing of data. Then, the axiom which concerns the same number of positive and negative errors, and which is given in the classical theory of measurement errors, is not met (e.g. Pearson, 1920; Friori and Zenga, 2009). Kukuča (1967) and Dzhun' (2012) indicated such a reason of asymmetry of the error distributions in the case of geodetic or astronomical observations. If the systematic errors are carefully eliminated, then the skewness usually achieves the small values. For example, in the case of the astrometric observations within the project MERIT, $\beta_1 = 0.0048$ (Dzhun', 2012); for the phase measurements from the SA*POS*®, GNSS observations, $\beta_1 = 0.0121$ (Luo et al., 2011). Similar values for GPS observations were also obtained by Tiberius and Borre (2000).

The excess kurtosis can be related with bigger or smaller (in relations to the Hagen hypothesis) frequency of occurrence of the elementary errors which are close to zero. The surfeit of such errors is the origin of leptokurtic distributions ($\beta_2 > 3$), and the deficiency – platykurtic distributions ($\beta_2 < 3$). Note that distributions are mesokurtic when $\beta_2 = 3$. Romanowski and Green (1983) noted that observation errors have usually symmetric mesokurtic or leptokurtic distributions, which justified application of the modified normal distributions. Except for small asymmetry, such a note is consistent with other empirical analyses. For example, Dzhun' (1992, 2012) showed that the errors of the astronomic observations have usually the kurtosis $\beta_2 = 3.8$ (however, $\beta_2 = 4.858$ was obtained during the project MERIT). Similar values of the kurtosis were obtained by Wassef (1959) and (Kukuča 1967) in the precise leveling. Considering contemporary observations, we should expect a wider range of the kurtosis values. For example, in the case of the observations from Satellite Laser Ranging, the kurtosis ranging $\beta_2 = 2.69 \div 9.46$ (Hu et al., 2001), and for GPS observations $\beta_2 = 2.79 \div 3.29$ (Luo et al., 2011).

One should note that the asymmetric coefficient and the kurtosis are usually estimated based on the sample moments (e.g. Dorić et al., 2009). Another possible approach is to compute all necessary moments during the process of adjustment by the least squares method (Wiśniewski, 1996; Kasietczuk, 1997). The statistic tests are the basis for determining whether the anomalies obtained are relevant, and the empirical distribution cannot still be described by the normal distribution. Here, the Jarque-Bera test or the D'agostino test (D'agostino et.al, 1990; Kayikçi and Sopaci, 2015) can be applied. Note, that the kurtosis estimate might be affected with gross errors (Kukuča, 1967). Thus, before application of the significance testing for β_1 and β_2, it is advisable to detect outliers by applying any of the methods presented in (Barda,1968; Lehmann, 2012).

If the asymmetry and/or kurtosis are significant, then one should decide which theoretical distribution is adequate for the measurement errors. In the case of symmetric leptokurtic distributions, one can apply several different distributions including the modified normal distributions given by Romanowski (1964) or the generalized normal distribution with the shape parameter which is steered by the excess kurtosis. Lehmann (2015) proposed to apply an information criterion, for example the Akaike Information Criterion, when a suitable probabilistic model is selected (besides the statistical hypothesis and test). The author presented pros and cons of such a method, considering the generalized normal distribution and its special cases.

In the case of wide range of asymmetry coefficient and/or kurtosis, the choice of a particular probabilistic model might be a complicated issue. Then the Pearson Distribution system (PD-system), which was proposed by Pearson (1920), seems to be a convenient solution. The distributions that belong to such a system are directly steered by the coefficients β_1 and β_2, and are very stable when approximating empirical distributions. Xi et al. (2012) showed that similar stability concerns also the saddlepoint approximation, the maximum entropy principle or the Johnson system. However, PD-system gives better results for small asymmetry and moderate values of the kurtosis. The general properties of the Pearson distributions are discussed by Elderton (1953) or Friori and Zenga (2009). The several selected distributions of that system were applied in astronomy (Dzhun', 1992, 2012) and in geodesy (Wiśniewski, 1987, 2014).

One of the main issues of adjustment process is to estimate the parameters of a functional model of observations. Usually, the least squares method (LS-method) is applied in such a case. However, if we know the probability density functions (PDF) of the measurement errors, then application of the maximum likelihood method (ML-method) seems more justified, e.g. Serfling (1980). Considering more general assumptions concerning the probabilistic models, one can also apply M-estimation which is based on a particular influence function or a weight function (Huber, 1964, 1981). Wiśniewski (2009, 2010) proposed another generalization of M-estimation, namely M_{split} estimation, where the main assumption is that there are several competitive functional models which can be related to the particular observation set (see also Duchnowski and Wiśniewski 2011, 2014, 2016).

Disturbances in estimation process that are caused by anomalies of empirical distributions, were discussed in general, for example, in Mooijaart (1985) or Mukhopadhyay (2005), and in the case of geodetic networks and LS-method in Gleinsvik (1971) and Wiśniewski (1985). Wiśniewski (1987) proposed to include such anomalies into the computation by applying and selected Pearson's distributions. Such a method requires the knowledge of the particular PDF, which is the basis for a newly formulated optimization problem. Dzun' (2011) showed that the adjustment that is based on ML-method and Pearson's distributions can be performed in a simpler way. The idea behind such an approach is the application of a weight function and the knowledge that PDFs of Pearson's distributions are solutions of a particular differential equation. The differential expression included in such an equation is proportional to

the influence function, which is very suitable here. Note that the influence function is based on distribution functions for whole PD-system. Considering such assumptions and the main idea proposed by Dzun'a (2011), Wiśniewski (2014) brought and analyzed a new solution called $M_{\mathcal{P}}$ estimation. The paper in question focused on such variants of the method which are referred to mesokurtic or leptokurtic distributions. Actually, such distributions predominate in astronomical and geodetic observations; however, they do not cover all the possible cases (e.g. Hu et al., 2001; Luo et al., 2011). Thus, there is a need for consideration $M_{\mathcal{P}}$ estimation for distributions which excess kurtosis is negative. We should realize that the application of the differential equation Ω_{PD} means that we do not choose any particular probabilistic model in fact (or any particular PDF), which is very important in such a context. We do not refer to the general properties of PDF, but we applied the values of the excess kurtosis and the asymmetry coefficient. Thus, $M_{\mathcal{P}}$ estimation is steered only by those two coefficients and by the standard deviation.

The paper is organized in the following way: Section 2 recalls the general assumptions of M-estimation based on the application of the influence and weight functions; Section 3 presents application of the Pearson system of distributions in $M_{\mathcal{P}}$ estimation. The special attention is paid to the Pearson distributions of types I and II, which are platykurtic. Finally, Section 4 presents results of numerical tests. Although these tests are elementary, they allow some general conclusions to be drawn (Section 5).

2. M-Estimation

The following functional model is usually assumed in theory and practice

$$\mathbf{y} = \mathbf{AX} + \mathbf{v} \tag{1}$$

where $\mathbf{y} \in R^n$ is an observation vector, $\mathbf{A} \in R^{n,r}$ is a known matrix of coefficients (rank $(\mathbf{A}) = r$), $\mathbf{X} \in R^r$ is a vector of unknown parameters and $\mathbf{v} \in R^n$ is a vector of random errors. The elements v_i of the vector $\mathbf{v}$ are assumed to be independent and their distributions P_{θ_i} are indexed with the parameter $\theta \in \Theta$ (Θ is a parameter space). The distributions P_{θ_i} which belong to the family $\mathcal{P} = \{P_{\theta_i} : \theta_i \in \Theta\}$ are regarded as probabilistic models of observation errors. In addition, we assume that the distribution P_X is a probabilistic model of the observation $y_i = \mathbf{a}_i\mathbf{X} + v_i$ ($\mathbf{a}_i - i$th row of the matrix $\mathbf{A}$). Note that such a distribution is indexed with the vector of parameters $\mathbf{X} \in (\Theta = R^r)$.

Consider the classical variant of M-estimation, then one should solve the following optimization problem (Huber,1981; Hampel et al., 1986):

$$\varphi_M(\mathbf{X}) = \sum_{i=1}^{n} \rho(y_i; \mathbf{X}) = \sum_{i=1}^{n} \rho(v_i) = \min \tag{2}$$

where $\rho(y_i; \mathbf{X}) = \rho(y_i - \mathbf{a}_i\mathbf{X}) = \rho(v_i)$. This is a generalization of ML-method, which optimization problem can be written as

$$\varphi_{ML}(\mathbf{X}) = \sum_{i=1}^{n}[-\ln f(y_i;\mathbf{X})] = \sum_{i=1}^{n}[-\ln f(v_i)] = \min \tag{3}$$

where $f(y_i; \mathbf{X}) = f(v_i)$ is PDF (or it is proportional to PDF). Thus, a particular family of distributions indexed with the parameter vector $\mathbf{X}$ should be assumed. In the case of M-estimation, the functions $\rho(v_i)$ are arbitrary; however, considering the following relation

$$\rho(v_i) = -\ln f(v_i) \implies f_\rho(v_i) = c\exp[-\rho(v_i)] \tag{4}$$

($c > 0$ is a normalization parameter) such functions can also be referred to certain distribution families. For example, the Huber method (Huber1964, 1981) assumes that the probabilistic model is defined by the family of the generalized normal distributions with two-segment PDF (Lehmann 2015).

Considering the function $\varphi_M(\mathbf{X})$, one can write the following respective gradient

$$\mathbf{g}(\mathbf{X}) = \frac{\partial\varphi_M(\mathbf{X})}{\partial\mathbf{X}^T} = \frac{\partial}{\partial\mathbf{X}^T}\sum_{i=1}^{n}\rho(v_i) = \sum_{i=1}^{n}\frac{d\rho(v_i)}{dv_i}\frac{\partial v_i}{\partial\mathbf{X}^T} = \sum_{i=1}^{n}\psi(v_i)\frac{\partial v_i}{\partial\mathbf{X}^T} = -\mathbf{A}^T\mathbf{\psi}(\mathbf{v}) \tag{5}$$

where $\mathbf{\psi}(\mathbf{v}) = [\psi(v_i), \cdots, \psi(v_n)]^T$. Thus, M-estimates of the vector $\mathbf{X}$ fulfills the equation (Huber,1981; Hampel et al., 1986)

$$\mathbf{A}^T\mathbf{\psi}(\mathbf{v}) = \mathbf{0} \tag{6}$$

The functions $\psi(v_i)$ are proportional to the influence functions $IF(v_i, F_\mathbf{X})$ which are based on the distribution functions $F(y_i; \mathbf{X}) = F_\mathbf{X}(v_i) \in \mathcal{F}$. Here, $\mathcal{F}$ is a family of distribution functions namely $\mathcal{F} = \{F_\mathbf{X}(v_i) : \mathbf{X} \in R^r\}$, which corresponds with the family of distributions of the observation errors $\mathcal{P} = \{P_{\theta_i} : \theta_i \in \Theta\}$ (Hampel, 1974; Serfling, 1980; Hampel et al., 1986). The functions $\psi(v_i)$ are also often called the influence functions. If the components $\rho(v_i)$ of the objective function in the optimization problem of Eq. (2) are known, then the influence functions $\psi(v_i)$ can be written in the following way

$$\psi(v_i) = \frac{d\rho(v_i)}{dv_i} = \frac{d\rho(v_i)}{d(v_i^2)}\cdot\frac{d(v_i^2)}{dv_i} = w(v_i)\frac{d(v_i^2)}{dv_i} = 2v_iw(v_i) \tag{7}$$

where

$$w(v_i) = \frac{d\rho(v_i)}{d(v_i^2)} = \frac{\psi(v)}{v} \tag{8}$$

is a weight function (Huber, 1981; Yang, 1997). If additionally $\rho(v_i) = \ln f(v_i)$, then one can also write that

$$\psi(v_i) = \frac{d\rho(v_i)}{dv_i} = \frac{d\ln(v_i)}{dv_i} = -\frac{1}{f(v_i)}\frac{df(v_i)}{dv_i} \tag{9}$$

Let us introduce the diagonal weight matrix $\mathbf{w} = \mathrm{diag}\big(w(v_1),..., w(v_n)\big)$, then the vector of the influence function can be written as $\mathbf{\psi(v)} = 2\mathbf{w(v)v}$. This leads to another form of Eq. (6), namely

$$\mathbf{g(X)} = -\mathbf{A}^{\mathrm{T}}\mathbf{w(v)v} = -\mathbf{A}^{\mathrm{T}}\mathbf{w(v)}(\mathbf{y} - \mathbf{AX}) = \mathbf{0} \tag{10}$$

M-estimate which is its iterative solution has the following form

$$\mathbf{\hat{X}} = [\mathbf{A}^{\mathrm{T}}\mathbf{w(\hat{v})A}]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{w(\hat{v})y} \tag{11}$$

For $\mathbf{w(\hat{v})} = \mathbf{P}$, where $\mathbf{P} \propto \mathrm{diag}(\sigma_1^{-2},...,\sigma_n^{-2})$, one can obtained LS-estimate of the parameter vector $\mathbf{X}$. The estimate $\mathbf{\hat{X}}_{LS} = (\mathbf{A}^{\mathrm{T}}\mathbf{PA})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Py}$ is a very convenient and often applied starting point within the iterative process which leads to the solution of Eq. (11).

Wiśniewski (2014) proposed the name $\mathrm{M}_{\mathcal{P}}$ estimates for all the solutions of Eq. (7) which are based on the influence functions $\psi(v_i) \propto IF(v_i, F_{\mathrm{X}})$ and the explicit families of distributions $\mathcal{P} = \{P_{\theta_i} : \theta_i \in \Theta\}$. In such a context, most of the popular M-estimates are also $\mathrm{M}_{\mathcal{P}}$ estimates, including the Huber estimates (the generalized normal distributions) or LS-estimates (the Gauss distributions).

3. $\mathrm{M}_{\mathcal{P}}$ Estimation with PD-SYSTEM

Wiśniewski (2014) proposed and analyzed the case of $\mathrm{M}_{\mathcal{P}}$ estimation, in which $\mathcal{P}_{PD} = \{P_{\theta} : \theta \in \Theta\}$ is a family of Pearson's distributions (PD-system). The origin of PD-system is the following differential equation Ω^* (Pearson, 1920; Elderton, 1953; Dzhun', 2011)

$$\Omega^*(v) = \frac{1}{f(v)}\frac{df(v)}{dv} = -\frac{(c_0 + 3c_2)v - \sigma c_1}{\sigma^2 c_0 - \sigma c_1 v + c_2 v^2} \tag{12}$$

where: $c_0 = 4\beta_2 - 3\beta_1$, $c_1 = \gamma_1(\beta_2 + 3)$, $c_2 = 2\beta_2 - 3\beta_1 - 6$, $\gamma_1 = \mathrm{sgn}(\mu_3)\sqrt{\beta_1}$. However, there might be some problems in direct application of that equation in $M_{\mathcal{P}}$ estimation. Note that in such a case, the expected value $E(v)$ lays at the origin of the coordinate system. If we assume that the estimate $\hat{\mathbf{X}}$ should minimize the amount of information of a particular observation set, then the mode M_0 should lay at the origin of the coordinate system. In the case of asymmetric distribution, the mode M_0 does not coincide with the expected value $E(v)$. Considering such a requirement, Wiśniewski (2014) proposed to modify the differential equation of PD-system in the following way

$$\Omega(v) = \frac{1}{f(v)}\frac{df(v)}{dv} = -\frac{(c_0 + 3c_2)v}{\sigma^2 c_0 - \sigma c_1(v-s) + c_2(v-s)^2} = -\psi(v) \tag{13}$$

where $s = M_0 - E(v) = \sigma c_1/(c_0 + 3c_2)$. Taking in account the properties of the influence function $\psi(v) = -\Omega(v)$ in the context of $M_{\mathcal{P}}$ estimation, it is worth considering two versions of that function, namely

$$\psi(v) = \begin{cases} \psi_{lep}(v), & c_1 = \gamma_1(\beta_2 + 3) & \text{for } \beta_2 \geq 3 \quad (\gamma_2 \geq 0) \\ \psi_{plt}(v), & c_1 = -\gamma_1(\beta_2 + 3) & \text{for } \beta_2 < 3 \quad (\gamma_2 < 0) \end{cases} \tag{14}$$

wherein, if $\beta_2 = 3$ then $\psi_{lep}(v) = \psi_{mez}(v)$, and $\psi_{mez}(v)$ is the influence function given for mesokurtic distributions (*lep* – leptokurtic distributions, *plt* – platykurtic distributions).

For $k = c_1^2 - 4c_0c_2 < 0$ (the Pearson distributions of types IV and VII), the influence function $\psi_{lep}(v)$ and the corresponding weight function

$$w_{lep}(v) = \frac{\psi_{lep}(v)}{v} \tag{15}$$

have unlimited range. $M_{\mathcal{P}}$ estimation with such a weight function was discussed in detail in (Wiśniewski, 2014). Such a property, which is advisable from the practical point of view, does not apply to the influence functions in the case of platikurtic distributions. For $k = c_1^2 - 4c_0c_2 > 0$ (the Pearson distributions of types I and II), the influence function $\psi_{plt}(v)$ and the corresponding weight function

$$w_{plt}(v) = \frac{\psi_{plt}(v)}{v} \tag{16}$$

have limited domain $\delta v = (a_1, a_2)$, where

$$a_1 = \frac{\sigma m_1 z}{2(m_1 + m_2)}, \quad a_2 = -\frac{m_2}{m_1} a_1, \quad z = \frac{2\sqrt{c_1^2 - 4c_0 c_2}}{c_2}$$

$$m_1 = -2m + \vartheta, \quad m_2 = -2m - \vartheta, \quad m = \frac{c_0 + 3c_2}{2c_2}, \quad \vartheta = \frac{\gamma_1(c_0^2 - c_2^2)}{z c_2^2}$$
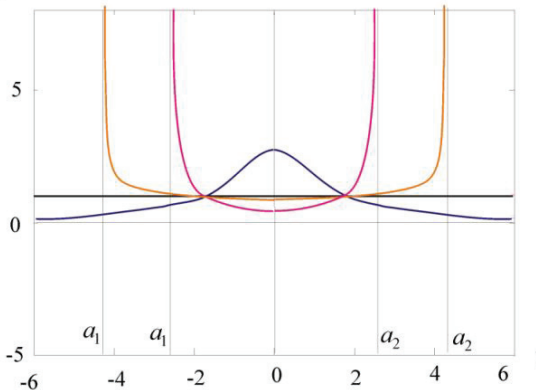
(17)

For $\beta_1 = 0$, it holds that $-a_1 = a_2$, where $a_1 = -8\sqrt{\beta_2 / (6 - 2\beta_2)}$, $a_2 = 8\sqrt{\beta_2 / (6 - 2\beta_2)}$. If $\beta_2 \to 3_-$, then $a_1 \to -\infty$ and $a_2 \to +\infty$.

Some variants of the influence functions and the corresponding weight functions, which are obtained for several values of the kurtosis and for the asymmetry coefficient equal to zero, are presented in Figure 1.
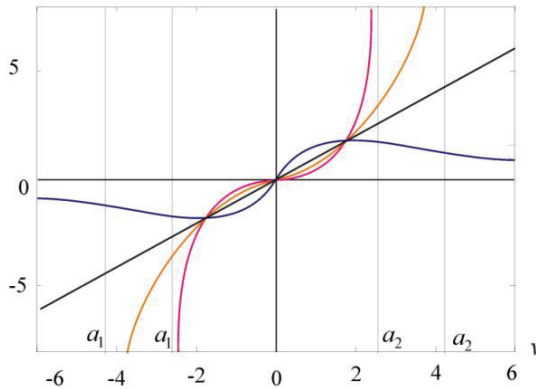
Fig. 1. Influence and weight functions in the case of symmetric probabilistic models

The weight functions $w_{lep}(v)$ have unlimited range and are bell-shaped (symmetric for $\beta_1 = 0$ or asymmetric for $\beta_1 > 0$). What is more, $\sup_v w_{lep}(v) = w(M_0)$. Considering the general classification of M-estimates (Kadaj, 1988; Wiśniewski, 2014), one can say that the following $M_\mathcal{P}$ estimate

$$\hat{\mathbf{X}}_{lep} = [\mathbf{A}^T \mathbf{w}_{lep}(\hat{\mathbf{v}})\mathbf{A}]^{-1} \mathbf{A}^T \mathbf{w}_{lep}(\hat{\mathbf{v}})\mathbf{y} \qquad (18)$$

satisfies the condition $\mathcal{K}_-$ (it is a robust estimate). If $\beta_1 = 0$, $\beta_2 = 3$, then $c_0 = 2$, $c_1 = 0$, $c_2 = 0$. Thus, we obtain $\psi_{lep}(v) = \psi_{mez}(v) = \psi_{LS}(v) = v/\sigma^2$ and $w_{LS}(v) = \psi_{LS}(v_1)/v_1 = 1/\sigma^2$, which are the influence function and the weigh function of LS-method, respectively (and the method itself satisfies the condition $\mathcal{K}_0$ – neutral estimation).

If $\beta_2 < 3$, then the weight functions $w_{PD}(v) = w_{plt}(v)$ are U-shaped within the interval $\delta v = (a_1, a_2)$. Therefore, they have two upper bounds within the interval $\Delta v = \langle a_1 + e, a_2 - e \rangle = \langle q_1, q_2 \rangle$, $e > 0$, namely

$$\sup_{q_1 \le v < M_0} w_{plt}(v) = w_{plt}(q_1) = \lambda_{w,1}^* \ , \qquad \sup_{M_0 < v \le q_2} w_{plt}(v) = w_{plt}(q_2) = \lambda_{w,2}^* \qquad (19)$$

wherein, if $e \to 0$, then $\lambda_{w,1}^* \to \infty$ and $\lambda_{w,2}^* \to \infty$. Since $w(M_0) < \min(\lambda_{w,1}^*, \lambda_{w,2}^*)$, then $M_\mathcal{P}$ estimates $\hat{\mathbf{X}}_{plt}$, which solves the equation $\mathbf{A}^T \mathbf{w}_{plt}(\mathbf{v})\mathbf{v} = \mathbf{0}$, satisfy the condition $\mathcal{K}_+$ within the interval δv (weak estimation). As a consequence, especially when the number of observations is low, the following iterative process

$$\mathbf{X}^{j+1} = [\mathbf{A}^T \mathbf{w}_{plt}(\mathbf{v}^j)\mathbf{A}]^{-1} \mathbf{A}^T \mathbf{w}_{plt}(\mathbf{v}^j)\mathbf{y} \ , \qquad \mathbf{v}^{j+1} = \mathbf{y} - \mathbf{A}\mathbf{X}_{plt}^{j+1} \qquad (20)$$

might be divergent. This results from the fact that the vector $\mathbf{X}^{j+1}$ is computed by applying the weight matrix, $\mathbf{w}_{plt}(\mathbf{v}^j)$, which depends on the residuals from the previous iterative step. Because the weight function is convex, the estimates $\mathbf{X}^j$ and $\mathbf{X}^{j+1}$ move away from each other and tend to the respective boundary points of the interval δv. This would be an interesting property of the method; however, this also raises a problem. While the estimates approach the boundary points a_1 or a_2, respectively, some iterative residuals might be out of the interval $\delta v = (a_1, a_2)$. The weight functions do not exist outside such interval (or they achieve unrealistic values), thus the iterative process cannot stabilize (without any special intervention, like for example, introduction of external artificial weigh functions). One should note that the cross weighting leads to interesting results in the case of a weak estimation, but only if the interval δv is infinite. A good example of such an approach is M_{split} estimation which is based on a split of M-estimate with the application of the weight function $w(v) = v^2$ (Wiśniewski, 2009, 2010; Cellmer, 2014; Wiśniewski and Zienkiewicz, 2016).

It is noteworthy that for the growing kurtosis, the weight function flattens within the certain interval $\Delta v \subset \delta v$. Considering a particular interval $\Delta v = \langle -k\sigma, k\sigma \rangle$, such flattening might be analyzed by application of the following differences (for $k > 0$)

$$r_L = w_{plt}(-k\sigma) - w_{plt}(M_0), \qquad r_R = w_{plt}(k\sigma) - w_{plt}(M_0) \tag{21}$$

Table 1 presents the values of r_L and r_R for several values of β_1 and β_2 (the skewness is positive) and under the assumption that $k = 2.5$ and $\sigma = 1$.

Table 1. Differences r_L and r_R in relation to asymmetric coefficient β_1 and kurtosis β_2

		$\beta_2 = 2.9999999$	$\beta_2 = 2.99$	$\beta_2 = 2.90$	$\beta_2 = 2.70$	$\beta_2 = 2.30$
$\beta_1 = 0.00$	r_L	0.000	0.010	0.114	0.443	10.568
	r_R	0.000	0.010	0.114	0.443	10.568
$\beta_1 = 0.01$	r_L	0.212	0.027	0.131	0.474	20.253
	r_R	0.187	0.002	0.099	0.415	7.102
$\beta_1 = 0.04$	r_L	0.226	0.051	0.149	0.507	201.285
	r_R	0.176	0.002	0.086	0.389	5.519

Kutterer (1999) considered the problem of how LS-estimation is influenced by disturbances of the weighs (herein, such disturbances correspond with the values of the differences r_L and r_R which decrease with the excess kurtosis tending to zero from the left-hand side). Generally, such influence should be analyzed separately from the theoretical point of view. However, for moderate negative values of the excess kurtosis and small values of the skewness, one can expect that $\hat{\mathbf{X}}_{plt} = \hat{\mathbf{X}}_{LS}$. If the anomalies of the empirical distribution are bigger, then such equality might not be true. In such a case, one can get satisfactory results if the observations are concentrated around the mode in a sufficient way or, in other words, if the errors of the observations which are grouped around the mode have the decisive influence on value of the expression $\sum_{i=1}^{n} w(v_i) y_i$. Although the weights of those observations are the smallest, the great number of the errors which are close to the mode might make the iterative process of Eq. (20) to converge. The chance for a satisfactory solution increases with the growing number of observations.

If the number of observations is small, then the empirical distribution might have some local modes (apart from the global one). Then, in the final steps of the iterative procedure of Eq. (20), the process might stabilize and generate sequence of some repeated values. The estimate that we are interested in, and which zeroes the gradient $\mathbf{g}(\mathbf{X}) = -\mathbf{A}^T \mathbf{w}_{plt}(\mathbf{v})\mathbf{v}$ might be among such values. However, such estimation way is not convincing from the theoretical as well as practical point of view. Generating of the sequence of some repeated values usually results from unsuitable starting point

or relatively big iterative increases (especially in the very first iterative steps). If the skewness is moderate, the estimate $\hat{\mathbf{X}}_{LS}$ is usually close to the mode. Thus, if one assumes that $\mathbf{X}^0 = \hat{\mathbf{X}}_{LS}$, then it might happen that after the first iterative step the process "skips" the global mode and tends to one of the local ones. To avoid such a situation, one can apply the Newton method with the correction which forces the reduction of the iterative increases between subsequent iterative steps. Hessian of the objective function $\varphi_M(\mathbf{X})$, which is necessary in such an approach, can be written as follows

$$\mathbf{H}(\mathbf{X}) = \frac{\partial^2 \varphi_M(\mathbf{X})}{\partial \mathbf{X} \partial \mathbf{X}^{\mathrm{T}}} = \frac{\partial g(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}} \mathbf{w}_{plt}(\mathbf{v}) \mathbf{A} \tag{22}$$

Thus, the iterative process can be written in the following way ($j = 1,..., m$)

$$\begin{aligned} d\mathbf{X}^{j+1} &= -[\mathbf{H}(\mathbf{X}^j)]^{-1}\mathbf{g}(\mathbf{X}^j) = [\mathbf{A}^{\mathrm{T}}\mathbf{w}_{plt}(\mathbf{v}^j)\mathbf{A}]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{w}_{plt}(\mathbf{v}^j)\mathbf{v}^j \\ \mathbf{X}^{j+1} &= \mathbf{X}^j + \tau d\mathbf{X}^{j+1} \\ \mathbf{v}^{j+1} &= \mathbf{y} - \mathbf{A}\mathbf{X}^{j+1} \end{aligned} \tag{23}$$

where $\tau < 1$ is a reduction coefficient of the iterative increase. One should assume the value of τ so that the iterative process of Eq. (23) is convergent and ends up at the point $\hat{\mathbf{X}} = \mathbf{X}^{j=m}$, $\hat{\mathbf{v}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{X}}$, for which $\mathbf{g}(\hat{\mathbf{X}}) = -\mathbf{A}^{\mathrm{T}}\mathbf{w}_{plt}(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mathbf{0}$.

4. Numerical examples

Let us assume the following functional model of the observations $v_i = y_i - X$, $i = 1,...,n$, wherein X is an estimated parameter. From the geodetic point of view, such a mode might relate, for example, to a leveling network with one unknown point W and some fixed points P_j, $j = 1,...,k$ (see, Figure 2). Let h_j be a height difference between the points P_j and W, and let each h_j be measured s_j times, then $n = \sum_{j=1}^{k} s_j$. In the context of the model in question, measurements of h_j are observations y_j, and the height of the point W is the parameter X.

The observations are simulated by applying the Gaussian generator $randn(n,1)$ of the system MatLab. For small numbers of the observations n, the sample distributions often differ from the given normal distribution in the skewness or the kurtosis (Wiśniewski 2014). Such sets usually have some local aggregations of the observations. For bigger values of n, observation sets generated by $randn(n,1)$ have the assumed theoretical properties. For that reason, to obtain sets with negative excess kurtosis or significant skewness, one can combine normal distributions with different expected values (but with the control of the assumed standard deviation of the whole

set). Sets of the observations y_i, $i = 1,...,n$, will be denoted as ω, where l is the number of a set.
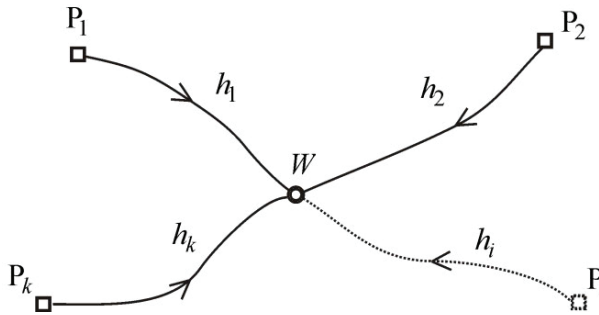

Fig. 2. Simulated, elementary leveling network

Consider some sets ω_1 which were generated under the assumption that $X = 0$. Let the empirical moments be computed for each of such sets. Hence σ, γ_1 and β_2 can also be computed. Thus, one can compute the following statistic

$$JB = n\left[\frac{\beta_1}{6} + \frac{(\beta_2 - 3)^2}{24}\right] \qquad (24)$$

which is the basis for the Jarque-Bera test (Bera and Jarque, 1980). Such a statistic is distributed according to the χ_α^2 distribution with two degrees of freedom if only the null hypothesis $H_0 : y \sim N(0,\sigma^2)$ is true (N - normal distribution, α – significance level). If $JB < \chi_\alpha^2$, then the null hypothesis is not rejected (the observations are from a normal distribution). Table 2 presents some critical values of the variable χ_α^2 for given significance levels.

Table 2. Critical vales of χ_α^2 in the Jarque-Bera test

α	0.025	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
χ_α^2	7.378	5.991	4.605	3.219	2.408	1.833	1.386	1.022	0.713	0.446	0.211

The basic analysis of the estimation for platikurtic distributions is based on the observation sets for which $n = 300$. The empirical distributions of two example sets are presented in Figure 3. The results obtained by applying the iterative process of Eq. (20) are presented in Table 3 (all the iterative processes were convergent). Table 3 presents the empirical distribution parameters, namely σ, γ_1, β_2, the boundary points of the interval $\delta v = (a_1, a_2)$, and finally $M_\mathcal{P}$ and LS estimates of the parameter X. The last column shows the significant levels α^*, for which the null hypothesis about normality of the distribution should be rejected.
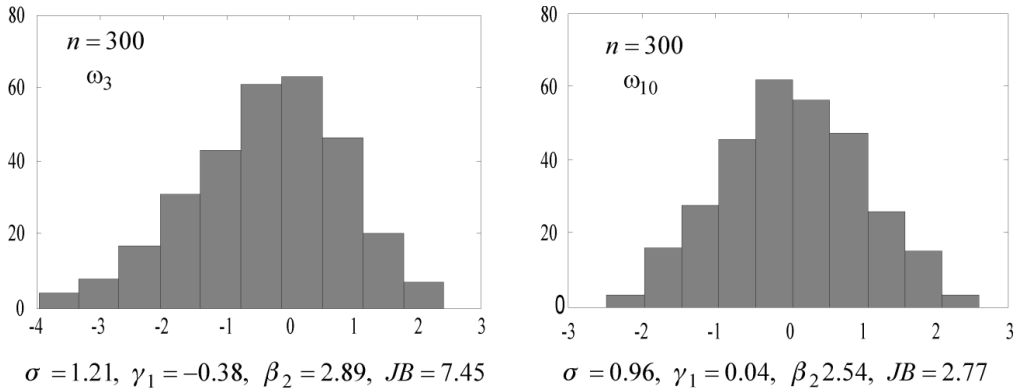
$\sigma = 1.21, \ \gamma_1 = -0.38, \ \beta_2 = 2.89, \ JB = 7.45$ $\qquad$ $\sigma = 0.96, \ \gamma_1 = 0.04, \ \beta_2 2.54, \ JB = 2.77$

Fig. 3. Example empirical distributions for $n = 300$

Table 3. $M_{\mathcal{P}}$ and LS estimates of $X = 0$ for platikurtic distributions ($n = 300$)

Sets	Empirical parameters			Acceptable interval		Estimators		JB	α^*
	σ	γ_1	β_2	a_1	a_2	$M_{\mathcal{P}}$	LS		
ω_1	1.130	-0.395	2.985	-8.994	3.097	-0.104	-0.369	7.818	0.025
ω_2	1.094	-0.384	2.818	-6.088	2.476	-0.036	-0.291	7.808	0.025
ω_3	1.206	-0.382	2.893	-7.785	3.023	-0.088	0.397	7.452	0.025
ω_4	1.068	0.348	2.964	-3.285	8.811	-0.004	0.231	6.071	0.05
ω_5	1.203	0.298	2.689	-2.689	5.699	-0.006	0.289	5.664	0.10
ω_6	1.178	0.077	2.405	-3.042	3.616	0.192	0.293	4.726	0.10
ω_7	1.022	0.249	2.768	-3.037	5.620	-0.003	0.174	3.774	0.20
ω_8	1.060	0.041	2.499	-3.042	3.616	0.082	0.120	3.208	0.30
ω_9	1.042	0.132	2.593	-3.082	4.174	0.126	0.224	2.994	0.30
ω_{10}	0.959	0.044	2.537	-3.011	3.323	0.009	0.044	2.774	0.30
ω_{11}	1.039	-0.158	2.709	-5.096	3.455	-0.024	-0.129	2.293	0.40
ω_{12}	0.947	0.006	2.644	-3.624	3.680	0.042	0.050	1.581	0.50
ω_{13}	1.024	-0.024	2.653	-4.122	3.883	0.042	0.021	1.534	0.50
ω_{14}	1.006	-0.065	2.961	-14.049	9.598	-0.046	-0.080	0.228	0.90

Similar results were also obtained for other tests when $n = 300$. The difference between $M_{\mathcal{P}}$ and LS estimates of the parameter X usually decreases with decreasing values of the statistic JB. However, it is not a general rule since the statistic JB can achieve the similar values for different skewness and excess kurtosis. The asymmetry (skewness) has the biggest influence on the difference between the estimates in question. If it is small, the difference is also not so significant (see, the sets ω_{11}, ω_{12}, ω_{13}); if the skewness is very close to zero, then usually the $M_{\mathcal{P}}$ and LS estimates are equal to each other. On the contrary, if the negative excess kurtosis is very small and the skewness is large (deviation from the normal distribution is very significant), then the differences between the estimates are distinct (see, the sets from ω_1 to ω_4).

For the big observation sets, the results of $M_{\mathcal{P}}$ estimation obtained for platikurtic distributions are similar to those for leptokurtic distributions (if only the observation sets are free of outlying observations). If the number of observations is smaller, then some local modes might occur. In the case of leptokurtic distributions, the weight functions are concave; hence, such local modes do not result in any serious optimization problems. On the other hand, some problems with finding the minimum of the objective function might happen for platykurtic distributions, for which the weight functions are convex (this was mentioned in the previous section). Table 4 presents the estimates of the parameter X, for the observation sets for which $n = 32$. The empirical distributions of two of them are presented in Figure 4. Note that for small observation sets and a reasonable significance level, the values of the statistics JB do not suggest that the hypothesis about the normality of observation distributions should be rejected (even if excess kurtosis or asymmetry is significant). However, this does not mean that $M_{\mathcal{P}}$ estimation cannot provide better results than the conventional LS estimation.

$\sigma = 1.02,\ \gamma_1 = 0.19,\ \beta_2 = 2.20,\ JB = 1.04$ $\sigma = 0.84,\ \gamma_1 = -0.297,\ \beta_2 = 2.47,\ JB = 0.84$
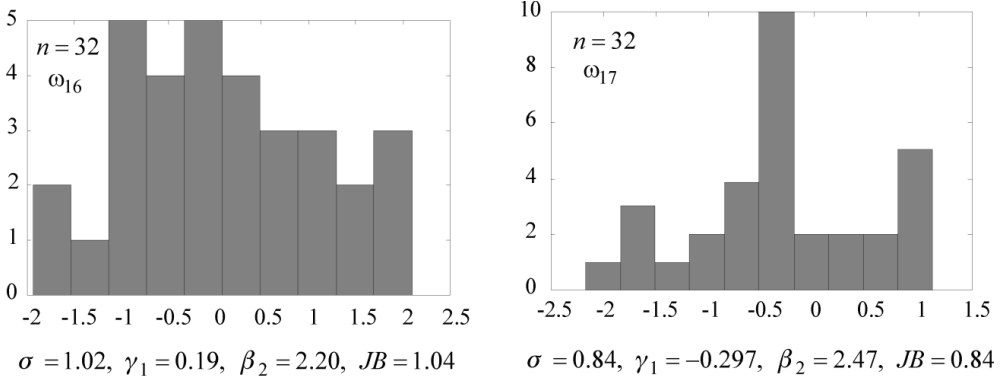
Fig. 4. Empirical distributions for $n = 32$

Table 4. $M_\mathcal{P}$ and LS estimates of parameter $X = 0$ for platykurtic distributions ($n = 32$)

Sets	Empirical parameters			Estimates		JB	Comments about estimation process of Eqs. (20 or 23)
	σ	γ_1	β_2	$M_\mathcal{P}$	LS		
ω_{15}	0.915	0.053	2.010	0.065	0.116	1.321	Process (20) divergent Process (23) convergent for $\tau = 0.1$
ω_{16}	1.034	0.191	2.202	0.042	0.109	1.044	Process (20) divergent Process (23) convergent for $\tau = 0.2$
ω_{17}	0.842	-0.297	2.471	-0.037	-0.249	0.842	Process (20) convergent

For the sets ω_{15}, ω_{16}, the iterative processes of Eq. (20), which solve the equation $\mathbf{A}^T\mathbf{w}_{plt}(\mathbf{v})\mathbf{v} = \mathbf{0}$, were divergent. Thus in such cases, the Newton method with a reduction coefficient τ was also applied. The iterative process for the observation set ω_{16} is presented in Table 5 ($\tau = 0.2$). As for the observation set ω_{15}, the gradient was zeroed (with the assumed tolerance) for $\tau = 0.1$, which make the iterative process much longer. In the case of the set ω_{17}, the iterative process of Eq. (20) was convergent and ended up after 15 iterative steps (similar number of the iterative steps was obtained when the gradient method for $\tau = 1$ was applied).

Table 5. Iterative process (the Newton method) for set ω_{16}

Steps j	$g(X^j)$	$H(X^j)$	dX^j	X^{j+1}
0			$X^0 = X_{LS} = 0.1092$	
1	7.126	19.401	-0.0735	0.0356
2	-0.480	15.693	0.0061	0.0418
3	-0.005	15.909	6.6×10^{-5}	0.0418
4	6.2×10^{-5}	15.910	7.9×10^{-5}	0.0418
5	7.5×10^{-7}	15.910	9.4×10^{-9}	0.0418

5. Conclusions

Although platykurtic distributions are rare in practice, it was an interesting issue to find out what results can be provided by $M_\mathcal{P}$ estimation in the case of such observation distributions. The numerical tests which are presented in the paper are rather limited; however, they allow us to draw some general conclusions.

The weight function in $M_\mathcal{P}$ estimation is convex if one applies the platykurtic Pearson distributions of types I or II. For big observation sets, which excess kurtosis

is small and negative, such property of the weight function does not result in any serious problems in searching for the estimate that minimizes the objective function. In such a case, the iterative process is very similar to $M_{\mathcal{P}}$ estimation with the application of leptokurtic distributions and hence concave weight functions. Thus, the iterative process described by Eq. (20) can be applied to solve the equation $\mathbf{A}^T\mathbf{w}_{plt}(\mathbf{v})\mathbf{v} = \mathbf{0}$ directly. On the other hand, some problems might occur for small observation sets. Within such sets there might be some local modes which do not necessarily result from the combination of various random variables. Thus, the iterative process of Eq. (20) might be divergent or might stabilize at the sequence of some repeated values. Note that in the case of a concave weight function (like for the Pearson distributions of types IV or VII), the empirical local modes generally have little chance to dominate the iterative process. Here, the significance of observations which lay around the global mode is strengthened by the biggest values of the weight function within the whole interval $\delta v = (-\infty, \infty)$. The situation is much different for convex weight functions for which local modes can destabilize the iterative process. Note that the weight functions does not exist beyond the interval $\delta v = (a_1, a_2)$ or, like for the Pearson distributions of types I or II, they achieve the unrealistic values hence they cannot be regarded as real weight functions. Then, it is necessary to control the value of the gradient for all the final values of the iterative process of Eq.(20). Usually, none of such values zeroes the gradient. The iterative process may succeed when one applies the Newton method with the correction $\tau < 1$ which forces the reduction of the iterative increase. However, one should assume that there is no local "peak" between the starting point and the global mode. Such condition is usually met if $\mathbf{X}^0 = \hat{\mathbf{X}}_{LS}$ (except some extremely unfavorable conditions when the empirical distribution has big asymmetry and large negative excess kurtosis). If for the given starting point the solution $\hat{\mathbf{X}}$ for which $\mathbf{g}(\hat{\mathbf{x}}) = \mathbf{0}$ cannot be obtained, then the iterative process can be repeated for smaller value of the coefficient τ. It might happen that the starting point should be changed too.

Generally speaking, $M_{\mathcal{P}}$ estimation for platykurtic distributions provides similar results like for leptokurtic ones (if only the observation set is free of outlying observations). Both of versions of such a method, which are presented herein, can reduce the influence of the anomalies of empirical distributions on the final estimate. Note that asymmetry of the empirical distribution disturbs the estimate in most significant way. The influence of that anomaly increases with grooving absolute value of the excess kurtosis. In the absence of asymmetry, the mode is equal to the expected value (for moderate negative excess kurtosis), and $M_{\mathcal{P}}$ and LS estimates are also equal to each other; however, some numerical problems with computation of $M_{\mathcal{P}}$ estimate may also occur in such a case.

The value of the test statistic *JB* depends most of all on the number of observations. If such number is small, the null hypothesis (observations are normally distributed) is rejected only for large anomalies (measured by the skewness and the excess kurtosis). The empirical distribution can then be described by the normal one (at the reasonable

significance level). However, this does not mean that M_P and LS estimates are always equal to each other.

Acknowledgments

References

Baarda, W. (1968). A test procedure for use in geodetic networks. *Neth. Geod. Comm. Publ. Geod., New Ser.* Vol. 2(5): 27-55.

Bera, A. K. and Jarque, C. (1980). Efficient Test for Normality, Heteroscedasticity and Serial Independence of Regression Residuals. *Economic Letters*, Vol. 6: 255-259.

Cellmer, S. (2014). Least fourth powers: optimisation method favouring outliers. *Survey Review*, Vol. 47(345): 411-417. DOI: 10.1179/1752270614Y.0000000142

D'agostino, R., Belanger, A. and D'agostino, R.A. (1990). A suggestion for using powerful and informative tests of normality. The American Statistician, Vol. 44(4): 316–321. DOI:10.1080/0003 1305.1990.10475751

Dorić, D., Nikolić-Dorić, E., Jeveremović, V. and Mališić, J. (2009). On measuring skewness and kurtosis. *Quality and Quantity*, Vol. 43(3): 481-493. DOI: 10.1007/s11135-007-9128-9

Duchnowski, R. and Wiśniewski, Z. (2011). Estimation of the shift between parameters of functional models of geodetic observations by applying M_{split} estimation. *Journal of Surveying Engineering* Vol. 138(1): 1-8. DOI: 10.1061/(ASCE)SU.1943-5428.0000062

Duchnowski, R. and Wiśniewski, Z. (2014). Comparison of two unconventional methods of estimation applied to determine network point displacement. *Survey Review*, Vol. 46(339): 401-405. DOI: 10.1179/1752270614Y.0000000127

Duchnowski, R. and Wiśniewski, Z. (2016). Accuracy of the Hodges–Lehmann estimates computed by applying Monte Carlo simulations. *Acta Geod Geophys*. DOI 10.1007/s40328-016-0186-0

Dzhun', I.V. (1992). Pearson distribution of type VII used to approximate observation errors in astronomy. *Measurement Techniques*, Vol. 35(3): 277-282.

Dzhun', I.V. (2011). Method for diagnostics of mathematical models in theoretical astronomy and astrometry. *Kinematics and Physics of Celestial Bodies, Mathematical Processing of Astronomical Data*, Vol. 27(5): 260-264.

Dzhun', I.V. (2012). What should be the observation-calculation residuals in modern astrometric experiments. *Kinematics and Physics of Celestial Bodies, Mathematical Processing of Astronomical Data*, Vol. 28(1): 43-47. DOI: 10.3103/S0884591312010096

Elderton, W.P. (1953). *Frequency curves and correlation*. Cambridge University Press.

Fischer, H. (2011). A history of central limit theorem. From classical to modern probability theory. *Sources and Studies in the History of Mathematics and Physical Sciences*, Springer New York-Dordrecht-Heidelberg-London, Book Chapter: 75-137.

Friori, A.M. and Zenga, M. (2009). Karl Pearson and the origin of kurtosis. *International Statistical Review*, Vol. **77**(1): 40-50. DOI: 10.1111/j.1751-5823.2009.00076.x

Gleinsvik, P. (1971). Zur Leistungsfähigkeit der Methode der kleinsten Quadrate bei Ausgleichung nicht normalverteilter Beobachtungen. Theoretische Untersuchungen. *Zeitschrift für Vermessungswesen*, Vol. 6: 224-233.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, Vol. 69(346): 383-397.

Hampel, F.R., Ronchetti, E.M., Rousseuw, P.J. and Stahel, W.A. (1986). *Robust statistics. The approach based on influence functions*. John Wiley & Sons, New York.

Hu, X., Huang, C. and Liao, X. (2001). A new solution assessment approach and its application to space geodesy data analysis. *Celestial Mechanics and Dynamical Astronomy*, Vol. 81(4): 265-278. DOI: 10.1023/A:1013204418865

Huber, P.J. (1964). Robust estimation of location parameter. *The Annals of Mathematical Statistics*. Vol. 35(1): 73-101. DOI:10.1214/aoms/1177703732

Huber, P.J. (1981). Robust statistics. The approach based on influence functions. John Wiley & Sons, New York.

Kadaj, R. (1988). Eine verallgemeinerte Klasse von Schätzverfahren mit praktischen Anwendungen. *Zeitschrift für Vermessungswesen*, Vol. 113: 157-166.

Kasietczuk, B. (1997). Estimation of asymmetry and kurtosis coefficients in the process of geodetic network adjustment by the least-squares method. *Journal of Geodesy*, Vol. 71(3): 131-136.

Kayikçi, E.T. and Sopaci, E. (2015). Testing the normality of the residuals of surface temperature data at VLBI/GPS co-located sites by goodness of fit tests. *Arabian Journal of Geosciences*, Vol. 8(11): 10119–10134. DOI: 10.1007/s12517-015-1911-7

Kukuča, J. (1967). Some problems in estimating the accuracy of a measuring method. *Studia Geophysica et Geodaetica*, Vol. 11(1): 21-33.

Kutterer, H. (1999). On the sensitivity of the results of least-squares adjustments concerning the stochastic model. *Journal of Geodesy*, Vol. 73(7): 350 -361.

Lehmann, R. (2012). Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *Journal of Geodesy*, Vol. 86(12): 1137–1146. DOI: 10.1007/s00190-012-0569-0

Lehmann, R. (2015). Observation error model selection by information criteria vs. normality testing. *Studia Geophysica et Geodaetica*, Vol. 59(4): 489-504. DOI: 10.1007/s11200-015-0725-0

Luo, X., Mayer, M. and Heck, B. (2011). On the probability distribution of GNSs carrier phase observations. *GPS Solutions*, Vol. 15(4): 369-379. DOI: 10.1007/s10291-010-0196-2

Mooijaart, A. (1985). Factor analysis for non-normal variables. *Psychometrika*, Vol. 50(3): 323-342. DOI: 10.1007/BF02294108

Mukhopadhyay, N. (2005). Dependence or independence of the sample mean and variance in non -IID or non-normal cases and the role of some tests of independence. *Recents Advances in Applied Probability*. Springer Science + Business Media, Inc. Book Chapter: 397-426.

Pearson, K. (1920). The fundamental problem of practical statistics. *Statistics*, Vol. 13: 1–16.

Romanowski, M. (1964). On the normal law of errors. *Bulletin Géodésique*, Vol. 73(1): 195-215. DOI: 10.1007/BF02528935

Romanowski, M. and Green, E. (1983). Reflexions on the kurtosis of samples of errors. *Bulletin Géodésique*, Vol. 57(1): 62-82. DOI: 10.1007/BF02520912

Serfling, R. (1980). Approximation theorems of mathematical statistics. John Wiley & Sons (Polish edition, PWN, 1991).

Tiberius, C.C.J.M. and Borre, K. (2000). Are GPS data normally distributed. *In*: *Schwarz K.P.* (Ed.) *Geodesy Beyond 2000. International Association of Geodesy Symposia*, Vol. 121: 243-248.

Wassef, A.M. (1959). Note of the application of mathematical statistics to the analysis of levelling errors. *Bulletin Géod*ésique, Vol. 52(1): 19-26.

Wiśniewski, Z. (1985). The effect of the asymmetry of geodetic observation error distribution on the results of adjustment by least squares method. *Geodezja i Kartografia*, Vol. 34(1): 11-21.

Wiśniewski, Z.(1987, Method of geodetic network adjustment in extend to probabilistic measurement error properties. *Zeszyty Naukowe Akademii Górniczo-Hutniczej, Geodezja*, 95, No. 1127, 73-88.

Wiśniewski, Z. (1996). Estimation of the third and fourth order central moments of measurement errors from sums of powers of least squares adjustment residuals. *Journal of Geodesy*, Vol. 70(5): 256-262.

Wiśniewski, Z. (2009). Estimation of parameters in a split functional model of geodetic observations (M$_{split}$ estimation). *Journal of Geodesy*, Vol. 83(2): 105-120. DOI: 10.1007/s00190-008-0241-x

Wiśniewski, Z. (2010). M$_{split(q)}$ estimation: estimation of parameters in a multi split functional model of geodetic observations. *Journal of Geodesy*, Vol. 84(6): 355-372. DOI: 10.1007/s00190-010-0373-7

Wiśniewski, Z. (2014). M-estimation with probabilistic models of geodetic observations. *Journal of Geodesy*, Vol. 88(10): 941–957. DOI: 10.1007/s00190-014-0735-7.

Wiśniewski, Z. and Zienkiewicz, M.H. (2016). Shift-M$^*_{split}$ estimation in deformation analyses. *Journal of Surveying Engineering*, DOI: 10.1061/(ASCE)SU.1943-5428.0000183.

Xi, Z., Hu, C. and Youn, B.D. (2012). A comparative study of probability estimation methods for reliability analysis. *Structural and Multidisciplinary Optimization*, Vol. 45(1): 33-52. DOI: 10.1007/s00158-011-0656-5

Yang, Y. (1997). Estimators of covariance matrix at robust estimation based on influence functions. *Zeitschrift für Vermessungswesen*, Vol. 122: 166-174.