APPROXIMATE REPRESENTATION OF THE $p$-NORM DISTRIBUTION

SUN Haiyan

KEY WORDS  $p$-norm distribution; approximate representation of the $p$-norm distribution; skewness; kurtosis

ABSTRACT  In surveying data processing, we generally suppose that the observational errors distribute normally. In this case the method of least squares can give the minimum variance unbiased estimation of the parameters. The method of least squares does not have the character of robustness, so the use of it will become unsuitable when a few measurements inheriting gross error mix with others. We can use the robust estimating methods that can avoid the influence of gross errors. With this kind of method there is no need to know the exact distribution of the observations. But it will cause other difficulties such as the hypothesis testing for estimated parameters when the sample size is not so big. For non-normally distributed measurements we can suppose they obey the $p$-norm distribution law. The $p$-norm distribution is a distributional class, which includes the most frequently used distributions such as the Laplace, Normal and Rectangular ones. This distribution is symmetric and has a kurtosis between 3 and $-6/5$ when $p$ is larger than 1. Using $p$-norm distribution to describe the statistical character of the errors, the only assumption is that the error distribution is a symmetric and unimodal curve. This method possesses the property of a kind of self-adapting. But the density function of the $p$-norm distribution is so complex that it makes the theoretical analysis more difficult. And the troublesome calculation also makes this method not suitable for practice. The research of this paper indicates that the $p$-norm distribution can be represented by the linear combination of Laplace distribution and normal distribution or by the linear combination of normal distribution and rectangular distribution approximately. Which kind of representation will be taken is according to whether the parameter $p$ is larger than 1 and less than 2 or $p$ is larger than 2. The approximate distribution have the same first four order with the exact one. It means that approximate distribution has the same mathematical expectation, variance, skewness and kurtosis with $p$-norm distribution. Because every density function used in the approximate formulae has a simple form, using the approximate density function to replace the $p$-norm ones will simplify the problems of $p$-norm distributed data processing obviously.

1 Introduction

In surveying data processing, it is often supposed that observational errors distribute normally. If observations come from the normal distributional class, the method of least squares can give the minimum variance unbiased estimation of the parameters. Since Gauss put forward the method of least squares, the theory has become mature after 200 hundred years of research and development. It is accepted for a long period that normal distribution can describe statistical character of observational errors perfectly. In addition, the method of least squares is easy to use for computation that makes it take a leading position in the practice of surveying data processing. With the advancement of observation means and instrument, observations tend to be
It is discovered in the practice of surveying data processing that observational errors sometimes do not distribute normally. For example, digital map errors do not distribute normally, but obey $p$-norm distribution law ($p \approx 1.6$)\cite{1}. When observational errors are not normal errors, especially when they distribute with overlapped tails, a few observations with large errors or gross errors will ruin the value of least squares. The limitation that the values of least squares have poor anti-error capability promotes researchers to seek new estimation method. Box had presented the concept of robustness in 1953. Tukey put forward polluted distribution mode in 1960. And in 1953 Huber published his famous paper *Robust Estimation of a Location Parameter*. After that the effects on parameter estimation caused by the inconsistency of actual error distribution and its hypothetical one have been widely researched, and importance has been attached on the theory and method of robust estimation. Polluted distribution mode considers actual distribution of observational errors as a mixture of a standard distribution (or perfect distribution) and polluted distribution, that is, it accepts that most of the observations are normal and come from standard distribution $G(x)$, while a few abnormal observations originate from polluted distribution $H(x)$. The distribution of observational errors can be expressed as\cite{2}:

$$F(x) = (1 - \varepsilon)G(x) + \varepsilon H(x) \quad (1)$$

In practical application, it is difficult to determine the pollution distribution $H(x)$ and the small quantity $\varepsilon$. Actually, robust estimation method just accepts the difference between real distribution and perfect distribution, the focus of which is how to eliminate or weaken the adverse effect, but not discuss and determine the difference in details. The principal part $G(x)$ of the observational distribution $F(x)$ is known, and the pollution distribution $H(x)$ and pollution rate are unknown. Therefore, many conclusions of robust estimation are approximate and suitable to big samples. Thus when the sample size is small, difficulties may occur in application, such as the hypothesis testing for some statistical results.

When the error distribution is unimodal and symmetrical, it can be assumed that observational errors obey $p$-norm distribution law. The parameters of this distribution are $\mu$, $\sigma^2$ and $p$. Here $p$ is a positive real number. $\mu$ and $\sigma$ are the mathematical expectation and variance of that distribution respectively. When the parameter $p$ changes from 0 to positive infinitude, $p$-norm distribution will change from degenerate distribution ($p \rightarrow 0$) to rectangular distribution ($p \rightarrow \infty$) through Laplace distribution ($p = 1$) and normal distribution ($p = 2$). The kurtosis can be any value between $+\infty$ and $-6/5$, containing both unimodal symmetrical tail-overlapped distribution and unimodal symmetrical tail-cut distribution. The maximum likelihood adjustment method is used, with $p$ as an estimated parameter, to determine the location and scale parameters as well as the estimated value of $p$ simultaneously, it means on the condition that error distribution is unimodal and symmetrical the distribution is determined thoroughly by observations. Thus we can research the sample distribution. When the sample size is not so big, some statistical results can also be tested.

There are some advantages to use $p$-norm distribution to describe statistical character of observational errors, however, its density function is very complex and some difficulties exist in the theoretical analysis and application. Therefore, it is necessary to seek a simple function to replace the density function of $p$-norm approximately. This paper intends to find this kind of function.

## 2 Approximate representation of the $p$-norm distribution

What can describe the statistical character of a random variable $x$ roundly is the probability distribution function or the density function of it.

$$F(x) = \int_{-\infty}^{x} F(x) dx \quad (2)$$

Here $F(x)$ is the distribution function and $f(x)$ is the density function. As long as $F(x)$ or $f(x)$ can be fixed, the statistical character of the random variable $x$ can be fixed completely. The numeric characteristic of the distribution of the random variable can reflect its statistical character in some
degree. The common-used numeric characteristics are mathematical expectation \[3\]

\[ E(x) = \int_{-\infty}^{\infty} x f(x) dx = \mu \quad (3) \]

variance

\[ D(x) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x) dx = \sigma^2 \quad (4) \]

skewness

\[ \gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - E(x))^3 f(x) dx \quad (5) \]

and kurtosis

\[ \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - E(x))^4 f(x) dx - 3 \quad (6) \]

Mathematical expectation and variance are two common-used numeric characteristics, and the normal distribution can be fully determined by these two parameters. Skewness is a quantity to express the symmetrical character of a distribution, and for symmetrical distribution skewness is zero. Kurtosis expresses the craggedness of the distribution. When a distribution is sharp and the two tails are longer than normal ones, the kurtosis is positive, whereas when a distribution is flat and the two tails are shorter than normal ones, the kurtosis is negative. The kurtoses of Laplace distribution, normal distribution and rectangular distribution are 3, 0 and -6/5 respectively. These four numeric characteristics can describe the distribution almost entirely.

Suppose there are two random variables \( x_1 \) and \( x_2 \), and their density functions are \( f_1(x) \) and \( f_2(x) \) respectively. If the mathematical expectation, variance, skewness and kurtosis of them are all equal, then it is reasonably thought that the random variables \( x_1 \) and \( x_2 \) approximately have the same statistical character, and \( f_1(x) \) is approximately equal to \( f_2(x) \). Their difference arises from the difference between the moments of 5th order and that of higher orders.

The density function of \( p \)-norm distribution is \[4\]

\[ f_p(x) = \frac{2\lambda^\frac{1}{p}}{2a\Gamma(\frac{1}{p})} \exp\left[-\left(\frac{x - \mu}{\sigma}\right)^{p}\right] \quad (7) \]

Its mathematical expectation is \( \mu \), variance is \( \sigma^2 \), skewness is 0 and kurtosis is \( \gamma_2 \), where

\[ \lambda = \sqrt{\frac{\Gamma(\frac{2}{p})}{\Gamma(\frac{1}{p})}} \quad (8) \]

\[ \gamma_2 = \frac{\Gamma(\frac{1}{p})\Gamma(\frac{5}{p})}{\Gamma(\frac{3}{p})} - 3 \quad (9) \]

If \( 1 \leq p \leq 2 \), then \( 3 \geq \gamma_2 \geq 0 \), the following formula can be used to express \( f_p(x) \) approximately:

\[ f(x) = (1 - \varepsilon)f_N(x) + \varepsilon f_L(x) \quad (10) \]

Here \( f_N(x) \) and \( f_L(x) \) are density functions of normal distribution and Laplace distribution respectively.

\[ f_N(x) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (11) \]

\[ f_L(x) = \frac{1}{\sqrt{2\sigma}}\exp\left[-\frac{1}{\sigma}(x - \mu)\right] \quad (12) \]

It is obvious that \( f(x) \) has the same mathematical expectation \( \mu \), and variance \( \sigma^2 \) with \( f_N(x) \) and \( f_L(x) \) and its skewness is zero. While the kurtosis of \( f(x) \) is

\[ \gamma_2 = \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - \mu)^4 f(x) dx - 3 \quad (13) \]

Therefore, let

\[ \varepsilon = \frac{1}{3} \left[ \frac{\Gamma(\frac{1}{p})\Gamma(\frac{5}{p})}{\Gamma(\frac{3}{p})} - 3 \right] \]

then \( f_p(x) \) and \( f(x) \) will have the same mathematical expectation, variance, skewness and kurtosis. That is, if \( f(x) \) can be used to replace \( f_p(x) \), it will be exact to the first four order moments.

Similarly, if \( 2 \leq p \leq \infty \), the kurtosis of \( f_p(x) \) is between those of normal distribution and rectangular distribution; thus, the combination of normal distribution and rectangular distribution can represent it approximately, that is

\[ f(x) = (1 - \varepsilon)f_N(x) + \varepsilon f_R(x) \quad (14) \]

where

\[ f_R(x) = \begin{cases} \frac{1}{2\sigma\sqrt{3}} \exp\left[-\frac{|x - \mu|}{\sigma}\right] & \frac{|x - \mu|}{\sigma} \leq \sqrt{3} \\ 0 & \frac{|x - \mu|}{\sigma} > \sqrt{3} \end{cases} \quad (15) \]

It is easy to know that the mathematical expectation of \( f_R(x) \) is 0, the variance is \( \sigma^2 \), the skewness is \( \gamma_1 = 0 \) and the kurtosis is \( -\frac{6}{5} \). The kurtosis of
distribution \( f(x) \) is

\[
\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \\
\frac{1}{\sigma^4} \left[ (1 - \varepsilon)\int_{-\infty}^{\infty} (x - \mu)^4 f_N(x) \cdot dx + \varepsilon \int_{-\infty}^{\infty} (x - \mu)^4 f_R(x) dx \right] - 3 = -\frac{6}{5} \varepsilon.
\]

Let

\[
\varepsilon = \frac{5}{6} \left[ 3 - \frac{\Gamma(\frac{1}{p})\Gamma(\frac{2}{p})}{\Gamma^2(\frac{3}{p})} \right]
\]

(16)

then \( f_p(x) \) and \( f(x) \) have the same mathematical expectation, variance, skewness and kurtosis. That is, if \( f(x) \) is used to replace \( f_p(x) \), it will also be exact to first four order moments.

### 3 The determination of \( \varepsilon \)

It is shown in Eqs. (10) and (14) that when \( p > 1 \), \( p \)-norm distribution can be expressed approximately by normal distribution and Laplace distribution or by normal distribution and rectangular distribution. In another word, a sample from \( p \)-norm distributional class can be considered as a mixture of samples from normal distribution and Laplace distribution or as a mixture of samples from normal distribution and rectangular distribution. \( 1 - \varepsilon \) and \( \varepsilon \) can reflect the proportion of each distribution. In order to express it clearly, some simple computations are performed as follows.

Let \( 1 \leq p \leq 2 \), Table 1 can be obtained from Eq. (13).

| \( p \) | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| \( \varepsilon \) | 1.00 | 0.76 | 0.58 | 0.45 | 0.34 | 0.25 | 0.18 | 0.13 | 0.08 | 0.04 | 0.00 |

Table 1 Relationship between \( p \) and \( \varepsilon \) (\( 1 \leq p \leq 2 \))

| \( p \) | 2.0 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 2.6 | 2.7 | 2.8 | 2.9 | 3.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| \( \varepsilon \) | 0.00 | 0.08 | 0.15 | 0.21 | 0.26 | 0.31 | 0.35 | 0.39 | 0.42 | 0.45 | 0.48 |

Table 2 Relationship between \( p \) and \( \varepsilon \) (\( 2 \leq p \leq \infty \))

| \( p \) | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| \( \varepsilon \) | 0.68 | 0.77 | 0.83 | 0.87 | 0.90 | 0.92 | 0.93 | 0.98 | 0.99 | 0.99 | 1.00 |

One point should be noticed that the approximate degree of the approximate expression of \( p \)-norm distribution in this paper is expressed by how many orders of moments are the same. As we can see, the density function of \( p \)-norm distribution has nearly the same shape with the approximate density function. As regards the difference of the value of density function between \( p \)-norm distribution and its approximate density function at every point, further research is needed.

### 4 Conclusion

When the distribution of observational errors is unimodal and symmetrical, it can be assumed that the observational errors obey \( p \)-norm distribution law. The method of maximum likelihood estimate can be used to determine location parameters, scale parameters and the value of the parameter \( p \). Because that the density function of \( p \)-norm distribution is comparatively complex, some difficulties will be caused in the theoretical analysis and application. Therefore, a simple function needs to be found to replace that of \( p \)-norm distribution. The analysis of this paper indicates that according to the value of \( p \) (\( 1 \leq p \leq 2 \) or \( 2 \leq p \leq \infty \)), \( p \)-norm distribution can be represented by the linear combination of Laplace distribution and normal distribution or by the linear combination of normal distribution and rectangular distribution approximately. The density
functions of Laplace distribution, normal distribution and rectangular distribution are all simple; therefore, the approximate representation is useful for both theoretical analysis and application.

Reference

1. Liu D J, Shi W Z, Tong X H, et al. Precision analysis and quality control of GIS spatial data. Shanghai: Shanghai Publishing House of Scientific Documents, 1999

2. Chen X R, Fang Z B, Li G Y, et al. Non-parameter statistics. Shanghai: Shanghai Publishing House of Science and Technology, 1989

3. Li Q H, Tao B Z. Application of probability statistical theory in surveying. Beijing: Beijing Publishing House of Surveying and Mapping, 1982

4. Sun H Y. p-norm distribution theory and its application in surveying data processing; [Ph. D Thesis]. Wuhan: Wuhan Technical University of Surveying and Mapping, 1995

STATEMENT ON JOINING

“WANFANG DATA—DIGITIZED PERIODICAL GROUP”

In order to electronize the process of editing, publishing and issuing scientific and technological periodicals as well as to advance the networking process of scientific and technological information interchange, we have now joined “Wanfang Data—Digitized Periodical Group”. Therefore, all contributions having been accepted to our publication will be brought into “Wanfang Data-Digitized Periodical Group” by the editorial office to offer information service on the Internet. Those who disagree, please contribute to other periodicals. The remuneration that our publication pays includes that of service on net of the content of the publication, so no extra payment will be further made.

“Wanfang Data—Digitized Periodical Group” is one of the key scientific and technological projects in the nation’s “the Ninth Five-Year Plan”. All Content in the periodical will be in uniform format. Readers can inquire and browse our publication, and solicit for subscriptions on the net.

Editorial Office of Geo-spatial Information Science