Properties of Majority Transformations under Random Processes Parameters Measurement

Elena Stepanova* and Alexey Liagin and Alla Pletukhina

North-Caucasus Federal University zik@ncstu.ru

Abstract. Statistical inference problem arises when you want to give the best, in some sense, the answers to a limited number of observations. When it comes to problems of statistical inference, it is assumed that it is possible to obtain a random sample, is set consisting of the realizations of independent and identically distributed random variables. The article attempts to assess quality parameters of a random process normally distributed through the application of order statistics (particularly the selected median) with a limited sample size.

In practice it is often required to determine the quantitative attribute (variables) against noise background. Lets assume that its initially known what kind of distribution a sign exactly possesses. There comes the task of assessing parameters that determine this distribution. If its known that the attribute under examination is normally distributed in the entire assembly, which is typical for mixture of signal and Gaussian noise at noise-to-signal ratio greater than unity, it is necessary to evaluate (calculate approximately) mathematical expectation and root-mean-square deviation, as these two parameters completely determine normal distribution.

Usually at a researchers disposal there is only sample data, e.g. quantitative attribute values $x_1, x_2, ..., x_n$ received as a result of N observations (observations assumed to be independent). Test variable is expressed in terms of this data.

When considering $x_1, x_2, ..., x_n$ as independent random variables $X_1, X_2, ..., X_n$, we can state that to find a statistical evaluation of the unknown parameter theoretical distribution means to find a function of the observed random variables, which gives an approximate value of the parameter estimated.

Most frequently in practice, for equally accurate measurements, parameters of random process distribution are evaluated by general average, which is reasonable for a large number of values of random variables observed. However for a small number of measurements, with high degree of unequal accuracy, as is known from mathematical statistics, sample median estimate is more effective than sample average.

Let physical process be described as a function of time $X(t)$. On the signal parameter tester signal $X = X(t) + n(t)$ is applied where $X(t)$ - measuring signal, $n(t)$ - external influence (noise). Signal parameters are evaluated at a certain time interval $\Delta(t)$.

* the presenter, email: cherry-sweet@yandex.ru

In mathematical statistics we know integrated (averaged) evaluation methods $x(\Delta t) = \lim\limits_{n\to\infty} \sum\limits_{i=i}^{n} \frac{x_i}{n}$, where $i = 1...n$ - equal for the majority of equally accurate measurements from the entire assembly $X(t) \subset f(x_1, x_2, ..., x_n)$. However, these techniques make evaluation of the random process parameters consistent, unbiased and effective only under controlled (predictable) changes of the parameters of random process.

When changes of random process parameters are unpredictable nonparametric method for estimation of the process parameters is interesting. Its principle is as follows: in the measurement time interval $\Delta(t)$ several measurements of the random process parameter $x_j = x + x_j$ $(j = 1...k)$ are made. Sensor output signals contain n_j measurement random errors caused by external influences (noise) which properties are characterized by probability densities $(f_j(x))$ Results of measurements (random variables x_j) form set of variate values

$$x_1, x_2, ..., x_k \tag{1}$$

so that $x_1 < x_2 < ... < x_k$. In case of odd number of measurements $(K = 2k+1)$ mean proportional of set of variate values (1.4.2) is taken as valuation of Z parameter of the process measured.

$$Z = X(k+1) \tag{2}$$

For sample size $K = 2k + 1 = 3$ suggested measurement algorithm can be implemented on (max and min) transformations. Then mathematical model and algorithm of measurement can be represented as $Z = med(x_1, x_2, x_3) = \max\{\min(x_1, x_2), \min(x_1, x_3), \min(x_2, x_3)\}$

For symmetrically distributed measurement errors estimate (2) is unbiased. Obviously, the error of estimate (2) is equal to the errors median n_j

$$e = Z - x = n_(k+1) \tag{3}$$

For practice, a case is important when the distribution of measurement errors on the observation interval is different, the differences characteristics are not known beforehand and cannot be used for optimal algorithms measurements apply. This situation occurs while the rapid changes in the measured parameter of physical process on the observation interval, for example, rapid changes in the amplitude of the analog signal, the frequency alteration and phase of harmonic oscillations - the information carriers in a rapidly changing interference intensity on the observation interval and a number of other situations.

Lets compare by statistical efficiency and estimation accuracy of the physical parameter process on sample median with a simple averaging for the case of unequal dimensions.

For practice, a case is interesting when the sample size for the measurement interval K = 3, the measurement errors are normally distributed with zero mean and variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ and it is unknown, what specific measurement corresponds to a certain level of error.

According to [1], the probability density estimation errors algorithm sample median:

$$f_2(x) = \sum_{i=1}^{3} f_i(x)\{F_j(x)[1 - F_b(x)] + F_b(x)[1 - F_j(x)]\}$$
$$(i, j, e = 1, 2, 3; i \neq j \neq e) \tag{4}$$

At a normal distribution inaccuracy, the formula looks like this:

$$F_i(x) = 0,5\left[1 + erf\left(\frac{x}{\sqrt{2}\sigma_i}\right)\right], where\ erf(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-t^2}dt.$$

Error variance σ_2^2 for the sample median, the corresponding distribution (5) turns out to be [1]

$$\sigma_{(2)}^2 = \sigma_1^2\left\{\frac{\lambda_1^2+\lambda_2^2}{2} - \frac{1}{\pi}\left[\arctan\frac{1}{A} + \frac{A^2+\lambda_1^2\lambda_2^2}{A(1+A^2)}\right] - \frac{\lambda_1^2}{\pi}\left[\arctan\frac{\lambda_1^2}{A} + \right.\right.$$
$$\left.\left. + \frac{\lambda_1^2(A^2+\lambda_2^2)}{A(A^2+\lambda_1^4)}\right] - \frac{\lambda_2^2}{\pi}\left[\arctan\frac{\lambda_2^2}{A} + \frac{\lambda_2^2(A+\lambda_1^2)}{A(A^2+\lambda_2^4)}\right]\right\} \tag{5}$$

From (6) follows that if the error variance of any two measurements are limited and the errors of the third dimension are infinitely large variance, and the measurement results practically unreliable, error variance is found to be $\sigma_2^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2)$.

This means that the assessment on sample median virtually eliminated false data.

The other situation is observed at an average measurement results. The calculation is defined as:

$$\sigma^2 = \frac{1}{9}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2) \tag{6}$$

That is: an unlimited increase of dispersion errors in one of the measurements leads to unlimited increase of error variance estimates.

At symmetric distribution laws $F_i(x)$ of errors and when they do not contain systematic components, the probability $P(\varDelta)$ in the case of unequal probability measurement is defined as:

$$P(\varDelta) = 1 - 2\left[\int_0^{\varDelta}\{f_1(x)(F_2(x)[1-F_3(x)] + F_3(x)[1-F_2(x)]) + f_2(x)(F_1(x)[1-\right.$$
$$\left.-F_3(x)] + F_3(x)[1-F_1(x)]) + F_3(x)(F_1(x)[1-F_2(x)] + F_2(x)[1-F_1(x)])\}dx\right]$$

As a result of integration, we obtain the following:

$$P(\varDelta) = F_1(\varDelta)F_2(\varDelta) + F_1(\varDelta)F_3(\varDelta) + F_2(\varDelta)F_3(\varDelta) - 2F_1(\varDelta)F_2(\varDelta)F_3(\varDelta) - \frac{1}{2} \tag{7}$$

In case of a normal error distribution at $\sigma = \sigma_2 = \sigma_3$ and $\sigma_3 \neq \sigma$ probability $P(\Delta)$ is:

$$P(\Delta) = 1 - \frac{1}{2}\Phi\left(\frac{\Delta}{\sqrt{2}\sigma\delta}\right)\left[1 - \Phi^2\left(\frac{\Delta}{\sqrt{2}\sigma}\right)\right] - \Phi\left(\frac{\Delta}{\sqrt{2}\sigma}\right), \tag{8}$$

where $\delta = \frac{\sigma_3}{\sigma}$.

when unequal measurements, the sample median has the best performance in terms of quality considered criterion than the sample mean an order of magnitude or more. This demonstrates the feasibility of applying the algorithm sample median for measuring the parameters of stochastic processes on the background noise and the impact of external influences, both on a physical process, and the measuring device. Practical confirmation the latest are researches for example, in [2,3,4].

References

[1] Signal processing based on an ordered choice. Guilbaud E. P., Chelpanov I.B. - Moscow: Soviet Radio, 1975.

[2] Properties of order statistics in the measurement of non-energy parameters of harmonic oscillations. Lyagin A.M., Pletuhina A.A., Pletuhina E.P. Topical issues and innovation in the economy, governance, education, information technology. Proceedings of the international conference. Issue. 6. T1. Stavropol.

[3] Majoritarian transformation algorithm for signal processing RPM at low sample sizes. Lyagin A.M., Pletuhina A.A., Pletuhina E.P. Stavropol, Kislovodsk, 2012. Fifth International Scientific Conference "Information and Communication Technologies in science, business and education" (Infocom-5). Collection of scientific papers. Part 1.

[4] Device distinguish discrete phase-modulated signals based on an ordered choice. Lyagin A.M., Pletuhina A.A., Pletuhina E.P., Mashchenko A.A. Herald, North - Caucasus University, the scientific journal, 2013, 1 (34), pages 39 - 42