Methods of Estimating the Form of the Probability Distribution Density in Tasks of Processing Measurement Results

V M Artyushenko¹ and V I Volovach²
¹Technological University, 42, Gagarin str., Korolev, 141070, Russia
²Volga Region State University of Service, 4, Gagarin str., Togliatty, 445017, Russia

E-mail: volovach.vi@mail.ru

Abstract. Issues associated with methods for estimating the shape of the probability distribution density curve are analyzed in order to class them when processing measurement results. For example, such nonparametric methods as the method of histograms and frequency polygon, as well as the method of classification of distributions, are considered. It is shown that the values of the anticurtosis and entropy coefficient can be taken as independent features of the form of symmetric distributions. For probability distribution densities that have a one-sided character, such as multiplicative noise, a skewness coefficient should be added to the parameters to consider. Recurrent procedures for obtaining current estimates of numerical characteristics of analyzed random processes are given. The results of processing a random process based on recurrent procedures are presented. It is shown that when the number of samples increases, the estimates obtained by using recurrent and non-recurrent procedures converge. The scattering of estimates of probability distribution density parameters, such as variance, relative mean square error, and entropy error, is determined.

1. Introduction
It is known, that the most important characteristics for statistical processing of observation results are the probability distribution function (PDF) and the probability distribution density [1-5, etc.]. Many papers are devoted to practical issues related to estimating these functions and their use in problems of processing observations results [6-9, etc.].

As a rule, two approaches are used in PDF estimating. In the first case, it is assumed that the PDF is described by a function of a known type that depends on unknown parameters that define this function. This method of estimation is called parametric and, in essence, means estimating unknown parameters of PDF [10-12].

However, a priori information about PDF is quite often completely absent or insufficient when processing the results of observations. It is often possible to define only the most general information, such as smoothness, concentration within a limited area, etc. In these cases non-parametric methods are used [12-14].

Interest in these methods is caused by the rapid development of learning systems, the development of non-parametric algorithms for solving identification problems, non-linear filtering algorithms, adaptive filtering algorithms, etc.
Given that the issues of processing experimental data of observation results by one or another method are widely known, for example [10, 13], we will consider the method of histograms and the frequency polygon in more detail.

The purpose of this work is analysis of methods of estimating the shape of a probability distribution density curve for their classification when processing signal parameter measurement results in radio engineering systems.

2. The method of histogram and the polygon of frequency

We assume, that a random variable \( \xi \) which has a PDF \( W_\xi(\xi) \) can be characterized by a sequence of independent observations \( \{\xi_h, h = 1, H\} \).

A sample PDF \( W_\xi(\xi) \) or a histogram is often used to characterize a random variable. As a rule, histograms are constructed in the following sequence.

Having determined the domain of a random variable \([a, b]\), one should divide it into \( N \) non-intersecting intervals \( \Delta_1, \Delta_2, \ldots, \Delta_N \) of equal length \( \delta = (b - a)/N \) and combine the results \( n_i \) of observations that fall into each \( i \)-th interval. Then, on these intervals, rectangles high are constructed, from which a step function \( W_n(\xi) = n_i/H\delta \) is obtained; \( \xi \in \Delta_i \) is called a histogram.

The resulting histogram is compared with the theoretical curve:

\[
W(\xi) = \delta^{-1} \int_{\xi_0}^{\xi} W_n(\xi) d\xi, \quad \xi \in \Delta_i.
\]

Moreover, if \( W_\xi(\xi) \) is continuous on the interval \([a, b]\), then the probability is

\[
P\left( \sup_{\xi \in [a,b]} |W_n(\xi) - W(\xi)| > \varepsilon \right) \rightarrow 0,
\]

when \( H \rightarrow \infty, \ N \rightarrow \infty \),

where sup is a mathematical abbreviation meaning: the exact upper bound and exact lower bound of the set (lub in English); \( \varepsilon \) is the value that determines the limit of the error.

Along with the histogram, the concept of a frequency polygon \( \varphi_n(\xi) \) is used, which is obtained by smoothing histograms, so that

\[
\varphi_n(\xi) = \frac{\left[ n + n_{k+1} \right]}{2H\delta + (\xi - a_k)(n_{k+1} - n_k)/H\delta^2}; \quad \xi \in [x_k, x_{k+1}].
\]

where \( x_k, k = 1, N \) is the middle of the intervals \( \Delta_i \); \( a_k \) is the right end of this interval.

Note, that

\[
\varphi_n(x_k) = \frac{n_k}{H\delta}; \quad \varphi_n(x_{k+1}) = \frac{n_{k+1}}{H\delta}.
\]

The frequency polygon \( \varphi_n(\xi) \) is a consistent estimate \( W(\xi) \), which means, that

\[
P\left( \max_{\xi \in [a,b]} |\varphi_n(\xi) - W(\xi)| > \varepsilon \right)_{H \rightarrow \infty} \rightarrow 0,
\]

if \( W_\xi(\xi) \) is continuous on the interval \([a, b]\).
3. Methods of distributions classification

Experimentally obtained histograms are often approximated by a suitable analytical expression. As a measure of the proximity of the true PDF \( W(\xi) \) and the approximating PDF \( W_{\text{apr}}(\xi) \) belonging to a certain parametric class selected in a certain way,

\[
\{W_{\text{apr}}(\xi, \lambda), \lambda = \{\lambda_1, \ldots, \lambda_m\}\},
\]

where \( \lambda \) are unknown parameter values, the minimum of the Kullback measure \( J_k \) can be used [1].

The use of this measure provides a minimum loss of information due to approximation and means that in the approximating PDF \( W_{\text{apr}}(\xi, \lambda) \), the parameters \( \lambda \) are selected in such a way as to minimize the interval:

\[
J_k(\lambda) = \int W(\xi) \ln \frac{W_0(\xi, \lambda)}{W(\xi)} d\xi,
\]

where \( W_0(\xi, \lambda) \) is the initial PDF \( \xi \) and \( \lambda \), so \( \lambda \) is selected according to the condition

\[
\lambda = \max^{-1} \left\{ \int W(\xi) \ln W_0(\xi, \lambda) d\xi \right\}.
\]

There are also other ways to classify distributions.

For instance, R. Pearson suggested using the coefficients of kurtosis and skewness of the PDF as coordinates of features. However, this turned out to be not viable, since with the skewness equal to zero, all the symmetrical PDFs were located along the axis of the kurtosis.

The concept of entropy coefficient, which varies for any PDF in the range from 0 to 2.066 appeared to be more viable

\[
k = \Delta_e / \sigma,
\]

where \( \sigma \) is root-mean-square deviation (RMSD) of a random variable \( \xi \); \( \Delta_e = 0.5 \exp H \) is an entropic error value [2];

\[
H = -\int_{-\infty}^{\infty} W(\xi) \ln W(\xi) d\xi
\]

is the PDF entropy.

Thus, for a uniform PDF \( \Delta_e = \sqrt{3} \sigma = 1.73 \sigma \) and, consequently, \( k = 1.73 \); for a normal PDF \( \Delta_e = \frac{\sigma \sqrt{2 \pi e}}{2} = \sqrt{0.5 \pi e} \sigma \approx 2.066 \sigma \) and \( k = 2.066 \); for Laplace PDF \( k = 1.93 \); for are sine PDF \( k = \frac{\pi}{\sqrt{8}} \approx 1.11 \) and so on.

As a second independent feature that characterizes the form of the PDF, one can take the coefficient of kurtosis

\[
k_e = \mu_4 / \sigma^4,
\]

which is equal to the relation of the fourth central moment \( \mu_4 \) to the squared variance \( \sigma^2 \). However, \( k_e \) can vary from zero to infinity for different PDFs, so for convenience, its nonlinear transformation is made: \( 1 / \sqrt{k_e} \). The value
\[ \mu = 1/\sqrt{k_e} \]

it is called a antikurtosis and can change from 0 to 1.

When these independent features are used, any PDF can be represented as a point in the coordinate system \((k, \mu)\) (Fig. 1) [1].

For example, an exponential PDF described by the expression

\[ W(\xi) = \left\{ \frac{\alpha}{2\Gamma(\alpha^{-1})} \right\} \exp\left\{ -|\xi|^\alpha \right\}, \]

where \(\alpha\) is the exponent parameter of the distribution; \(\Gamma(\cdot)\) is a Gamma function for which

\[ k_e = \frac{\Gamma(1/\alpha)\Gamma(5/\alpha)}{[\Gamma(3/\alpha)]^2}, \quad \mu = \frac{\Gamma(3/\alpha)}{[\Gamma(1/\alpha)\Gamma(5/\alpha)]^{1/2}}, \]

\[ k = \alpha^{-1} \exp\left\{ \alpha^{-1} \right\} \frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}. \]

corresponds to a curve passing through the points 2-6-7-5-8.

\[\begin{align*}
  & 0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 & 1.0 \\
  & 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1.0 & 1.2 & 1.4 & 1.6 & 1.8 & 2.0 \\
  \hline \\
  k & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 12 & 13 \\
  \end{align*}\]

\[\begin{align*}
  2 & 3 & 4 & 5 & 6 & 7 & 8 & 12 & 13 \\
  \end{align*}\]

**Figure 1.** The plane of features characterizing the form of the PDF.

Point 2 represents \(\alpha \to 0, \mu \to 0, k \to 0\); point 6 represents \(\mu = 0.199, k = 1.35, \alpha = 0.5\); point 7 represents a PDF with \(\mu = 0.408, k = 1.92, \alpha = 1\), which is a Laplace PDF. Point 5 represents Gaussian PDF.

Point 8 with coordinates \((\mu = 0.745, k = 1.73)\) represents a uniform distribution, when \(\alpha \to \infty\).

The family of Student's PDF
When observing samples of a random process $x$, if the number of degrees of freedom is $v = 1$, then the estimated PDF is one-sided, such as the one of multiplicative noise, then the skewness coefficient $k$ must be added to other considered parameters. As a result, this PDF is presented not as a form of PDF will be valid if the form of the PDF is symmetrical, and, as a rule, two-sided.

Thus, the considered features of PDF form, presented as points in the plane with coordinates $(\mu, k)$ allow us to visually estimate how close or different from each other different analytical models of the PDF, as well as their points corresponding to a particular model of experimental PDFs. The considered method has a significant drawback. It is related to the fact that if the values of the entropy coefficient $k$ and antikurtosis $\mu$ of the considered PDF are determined unambiguously, then the reverse transition is not unambiguous. This is due to the fact that through the point with the specified coordinates $(\mu, k)$ can pass not one, but a set of curves corresponding to PDFs of different classes.

It should be particularly noted that the considered method of determining the characteristics of the form of PDF will be valid if the form of the PDF is symmetrical, and, as a rule, two-sided.

If the estimated PDF is one-sided, such as the one of multiplicative noise, then the skewness coefficient $k$ must be added to other considered parameters. As a result, this PDF is presented not as a point in the feature space $(\mu, k, k)$, but as a point in the feature space $(\mu, k, k)$. Nowadays, the so-called recurrent procedures are widely used, which allow us to obtain current estimates of the numerical characteristics of the processes having a random character. In contrast to the posteriori algorithms, their implementation requires a much smaller memory capacity of the processing computer.

4. Recurrent estimate of initial moments of the $i$-th order

When observing samples of a random process $\xi$, $h = 1/H$, recurrent estimates of the initial moments of the $i$-th order $m$ have the following form

$$m_i = m_{i-1} + h^{-1}(\xi_i - m_{i-1}); \quad m_0 = 0; \quad h = 1/H. \quad (1)$$

When the mathematical expectation $m$ is known, the estimate of the variance of a random process (the 2-nd central moment $M_2$) can be found based on the expression

$$M_2 = M_{2, i-1} + h^{-1}\left[(\xi_i - m)^2 - M_{2, i-1}\right]. \quad (2)$$
In this case, the 3rd and 4th central moments will be determined based on the expressions:

\[ \dot{M}_{3,h} = \dot{M}_{3,h-1} + h^{-1} \left( (\hat{z}_h - m_h)^3 - \dot{M}_{3,h-1} \right) \]  \hspace{1cm} (3)

\[ \dot{M}_{4,h} = \dot{M}_{4,h-1} + h^{-1} \left( (\hat{z}_h - m_h)^4 - \dot{M}_{4,h-1} \right) \]  \hspace{1cm} (4)

Using (1)–(4), we obtain the recurrent relations that allow us to determine, respectively, the coefficients of skewness and kurtosis

\[ k_{a,h} = \dot{M}_{3,h} \dot{M}_{2,h}^{-1.5} \]  \hspace{1cm} (5)

\[ k_{c,h} = \dot{M}_{4,h} / \dot{M}_{2,h}^2 \]  \hspace{1cm} (6)

To complete the calculation procedure, i.e. to determine the required number of measurements \( H \), a criterion is usually used that characterizes the required measurement accuracy or the permissible measurement error \( \varepsilon \), obtained from the inequality \( |h - \lambda_{h-1}| \leq \varepsilon \).

Figure 2 shows the results of processing a random process \( \{y_h\} \) whose instantaneous values are bimodal, where \( r \) is the correlation coefficient; \( m_{H} = \frac{\sum_{h=1}^{H} y_{h}}{H} \).

Figure 3 shows the results of the numerical simulation.

**Figure 2.** Results of processing a fragment of a random process \( \{y_h\} \): (a) is the dependence of the mathematical expectation and variance on the iteration step; (b) is the histogram of the PDF of the instantaneous values.

Here, curves calculated using formulas (5) and (6) are shown as solid lines, curves calculated using non-recurrent formulas and the final sample are shown as dotted lines. As the presented graphs show, both these curves converge with an increase in the number of samples, which indicates that estimation procedures for recurrent and non-recurrent algorithms are equivalent.

Note that the entropy coefficient can be numerically determined from the histogram based on the expression:
\[ k = \frac{dn}{2\sigma} 10^{-\frac{1}{n-1} \sum_{j=1}^{n} \frac{1}{\sigma^2}}. \]

**Figure 3.** The results of numerical simulation in an iteration step: \( a \) – a skewness coefficient; \( b \) – a kurtosis coefficient.

Here the variable \( n \) determines a sample size; \( d \) and \( l \), respectively, determine the width and number of columns of the histogram; \( n_j \) is the number of observations in the \( j \)-th column \( (j = 1, m) \). A straight line at the top means averaging over the set.

5. **Estimating the variance of estimates of PDD parameters**

In practice, it is very important to obtain an estimate of the scattering of the estimates \( \sigma, \omega, \) and \( k \) depending on the values \( n \) and \( k_e \) of the PDF. This can be done by using the following formulas.

Thus, the variance \( D \) of the sample variance \( D^* \) when \( n > 20 \) with an error of 10\% can be found using the expressions

\[ D[D^*] = \frac{\mu_4 - \sigma^4}{n}; \quad \sigma[D^*] = \frac{\sqrt{\mu_4 - \sigma^4}}{2\sigma\sqrt{n}}, \]

where \( \sigma[D^*] \) is an RMS of the sample RMS \( \sigma \).

In this case, a relative mean square error of \( \sigma \) will be determined as

\[ \delta(\sigma^*) = \frac{\sigma(\sigma^*)}{\sigma} = \frac{\sqrt{k_e - 1}}{2\sqrt{n}}. \]

The variance of the estimate \( \omega \) with an error of 8...10\% can be:

\[ \delta(\omega) = \frac{\sigma(\omega^*)}{\omega} = \frac{4(k_e^2 - 1)}{\sqrt{29n}}. \]
6. Conclusions

Thus, the method for determining the form of symmetrical two-sided PDF is considered and analyzed. It is shown that the values of the antikurtos is \( \mu \) and entropy coefficient \( k \) can be taken as independent features of the PDF form. If the estimated PDF is one-sided, then the skewness coefficient \( k \) must be added to the considered features. Then the estimated PDF is represented by a point not on the plane of features \((\mu, k)\), but by a point in the space of features \((\mu, k, k_a)\). Recurrent procedures for obtaining current estimates of their numerical characteristics are studied.

The scientific novelty consists in the analysis of the method for determining the features of the form of one- and two-sided PDFs using asymmetry, counter-echo and entropy coefficients as independent features. The practical significance is that the methods given can be used to process the measurement results when a priori information about the PDF is completely or partially absent.

7. References

[1] Damelin S B and Miller W Jr 2012 The Mathematics of Signal Processing (Cambridge: Cambridge University Press)

[2] Bentsman J 2016 Introduction to Signal Processing, Instrumentation, and Control. An Integrative Approach University of Illinois at Urbana-Champaign, USA, 2016 (Urbana-Champaign: University of Illinois)

[3] Kumar A and Arya N A study relation between energy spectral density and probability density function with impulse response first order control system 2017 Proc. Int. conf. of Electronics, Communication and Aerospace Technology (ICECA) 20-22 April 2017, Coimbatore, India

[4] Silva H S, Queiroz W J L, Fonseca I E An approximation to the exponentiated weibull phase probability density function 2020 Proc. Int. Symp. on Antennas and Propagation and North American Radio Science Meeting 5-10 July 2020 Montreal, QC, Canada

[5] Artyushenko V M and Volovich V I Mathematical Methods for Describing a non-Gaussian Probability Density Function 2020 Proc. International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon 2020) 6-9 Oct. 2020, Vladivostok, Russia

[6] Bouhel N and Dziri A Kullback-Leibler Divergence Between Multivariate Generalized Gaussian Distributions 2019 IEEE Signal Processing Letters 26(7) 1021-5

[7] Tekel J and Cohen L Constructing and Estimating Probability Distributions from Moments 2012 Proc. of SPIE - The International Society for Optical Engineering (Proc. of SPIE) 8391 83910E 23-27 April 2012, Baltimore, Maryland, United States

[8] Kropotov Yu A, Belov A A and Proskuryakov A Yu Estimation of the distribution probability density acoustic signals and interferences, the reconstruction methods Proc. 2018 Dynamics of Systems, Mechanisms and Machines (Dynamics) (Omsk, Russia)

[9] Krylov V A, Moser G, Serpico S B and Zerubia J On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation 2019 IEEE Transactions on Image Processing 22(10) 3791-06

[10] Alghalith M A New Parametric Method of Estimating the Joint Probability Density 2017 Physica A: Statistical Mechanics and its Applications (PHYSICA A) 471(1) 799-3

[11] Madan D Estimating Parametric Models of Probability Distributions 2015 Methodology and Computing in Applied Probability 17 823-1

[12] Klejnova E Comparison of parametric and non parametric methods for signal time-frequency description 2015 Proc. 2015 25th Int. Conf. Radioelektronika (RADJOELEKTRONIKA) 21-22 April 2015, Pardubice, Czech Republic

[13] Fadda D, Sleza E and Bijaoui A Density Estimation with non-Parametric Methods 1998 Astronomy and Astrophysics Supplement Ser. 127(2) 335-52

[14] Pana C, Siven S, Thaddeu G and de Abreu An adaptive approach to non-parametric estimation of dynamic probability density functions 2016 Proc. 2016 13th Workshop on Positioning, Navigation and Communications (WPNC) 19-20 Oct. 2016, Bremen, Germany