Improving the accuracy of estimates of the pulse sequence period using the methodology of complete sufficient statistics

Konstantin Zhuchkov\textsuperscript{1}, Mikhail Vasilchenko\textsuperscript{2}, Anna Zagrebneva\textsuperscript{3} & Alexey Zavyalov\textsuperscript{1}

This paper is devoted to the synthesis of new signal processing algorithms based on the methodology of complete sufficient statistics and the possibility of using the Lehmann–Scheffe theorem. Using the example of a sequence of quasi-rectangular pulses, an approach to estimating their period was illustrated, taking into account the duty-off factor and the pulse squareness coefficient. A mathematical model was developed, on the basis of which, estimates of the potential accuracy of the methods were carried out. It is established that for the sample size value \((n > 8)\), the relative root-mean-square error of estimating the repetition period using the methodology of complete sufficient statistics is lower than that of the traditional estimate. In addition to theoretical calculations, simulation results confirming the achieved effect are presented. The results obtained have a wide range of applicability and can be used in the design of control and measuring equipment in the oil and gas industry, in the development of medical equipment, in the field of telecommunications, in the design of pulse-Doppler radars, etc.

The development of microelectronic components and computing platforms in the rapidly developing modern world creates new challenges for the development of algorithms and new approaches in the processing of real-time information. The synthesis of information processing algorithms has to be carried out in the presence of a complex set of interfering factors: temporal and spatial heterogeneity of the characteristics of useful signals and interference, sensitivity heterogeneity and sensor defects, pulse interference, etc. Even if all the interfering factors could be fully taken into account, the signal and noise environment model would be very complex, and signal processing algorithms based on such a model would be unrealizable in real time. In addition, these factors are inherently random and there may be a priori uncertainty in the description of real signals. Thus, the synthesis of algorithms for estimating signal parameters, as a rule, has to be carried out under conditions of multiple of interfering factors and a priori uncertainty.

In the condition of uncertainty in the initial data, attempts are made to solve the problem of evaluating the efficiency of the algorithm for measuring signal parameters against the background of noise using the classical Bayesian criterion\textsuperscript{1}. However, the complexity of these attempts is associated with the requirement for a priori probabilities of the presence and absence of a signal in the observed sample and risk estimation.

At the same time, there is a growing demand for the synthesis of new algorithms for solving problems of estimating signal parameters in a wide range of applications: radars\textsuperscript{2}, communication systems\textsuperscript{3,4}, electronic intelligence\textsuperscript{5}, X-Ray absorption spectra\textsuperscript{6}, Bragg scattering\textsuperscript{7}, plasma parameters estimation\textsuperscript{8,9}, assessment of the stability of information systems\textsuperscript{10}, the solution of the problem of parameter estimation for in-tube flaw detection\textsuperscript{11}, the development of optoelectronic displacement sensors\textsuperscript{12}, etc. The requirements for the accuracy of the assessment and its reliability are increasing, which determines both the use of more productive computing resources and determines the relevance of the search for new information processing algorithms.

The purpose of this work is an attempt to synthesize an algorithm for estimating the period of quasi-rectangular pulses under conditions of possible loss of a part of the sequence. The construction of a signal model was also part of this work. The novelty of the problem statement and the above solution concludes in the use of methods of complete sufficient statistics based on the application of the Rao–Blackwell estimate for the Lehman–Scheffe
theorem in the synthesis of the estimation algorithm. If the conditioning statistics are complete and sufficient, and the initial estimate is unbiased, then the Rao-Blackwell estimate is the best unbiased estimate.

At the same time, it is well known that Gaussian noise maximum likelihood estimation (MLE) is asymptotically optimal, i.e., the stability of its operations and the efficiency of the estimates are achieved at high signal/noise ratios and an infinite sample size. In the context of the problems of estimating the parameters of real signals, there is a need to search for algorithms that can work no worse with finite samples and low signal/noise ratios. That way, for example, in order to perform the parameter estimation for the sine combination signals and periodic signals the multi-innovation stochastic gradient algorithm based on the gradient optimization principle is presented. A closed-form maximum likelihood amplitude estimator is obtained. The extended noise-resistant correlation method to identify the period of pseudo-periodic signals is derived. Attempts are being made to use Fourier and wavelet transformation with windows of various kinds to increase the localization of signal energy or multi-threshold circuits are used, which give a gain in the accuracy of the estimate in particular cases. However, in general, they carry the risk of instability. It is also possible to use neural network methods to solve the problem of estimating signal parameters, but the main disadvantage here is a multiple increase in the sample size for training, even with a slight change to the signal–noise environment.

An important point for solving this problem is to ensure the stability of the proposed algorithm under conditions of a high level of a priori uncertainty. Also, when synthesizing an algorithm, it is necessary to establish the complexity class and estimate the runtime realizability limitations in the context of current computing hardware capabilities.

In radio engineering applications, the distributions of samples and observed technical processes often have complete sufficient statistics, which creates the prerequisites for the application of the Lehmann–Scheffe theorem in finding effective estimates of signal parameters. The application of the estimates obtained can find a place not only in traditional problems of estimating the repetition period, but also in several problems that reduce the problem of finding the repetition period. Two examples are: when implementing a clock recovery circuit for a demodulator operating on short signal samples, and when the phase-locked loop capture time is significantly longer than the signal sample duration.

Mathematical signal model
Consider the process \( x(t) \) on the segment \([0, \Delta T]\), which is an additive mixture of signal and noise:

\[
x(t) = A_m s(t) + \eta(t)
\]

where \( A_m \) is the signal amplitude, \( s(t) \) is the normalized signal, \( \eta(t) \) is the differentiable stationary Gaussian noise characterized by the variance \( \sigma^2 \) and the normalized correlation function \( r(t) \). The distribution of the energy parameters of the signal and noise will be characterized by the signal/noise ratio \( q = A_m/\sigma \).

Let the leading edges of the signal pulse \( s(t) \) intersect the level \( H \) at time \( t_{0i}^+ \), and the trailing at time \( t_{0i}^- \):

\[
t_{0i}^+ = t_{00}^+ + iT, \quad t_{0i}^- = t_{00}^- + iT/v, \quad i = 0, 1, \ldots,
\]

where \( T \) is the signal period; \( v \) is the off-duty factor; \( t_{00}^+ \) is the first crossing of \( H \) level. Then, for large signal/noise ratios, \( q >> 1 \), the process \( x(t) \) will cross the \( H \) level at times

\[
t_i^+ = t_{0i}^+ + \Delta t_{0i}^+ \quad \text{and} \quad t_i^- = t_{0i}^- + \Delta t_{0i}^-,
\]

where \( \Delta t_{0i}^+ \) and \( \Delta t_{0i}^- \) are small random additions, due to the effect of noise \( \eta(t) \). Figure 1 schematically shows the process and the designations adopted in the work.
It is assumed that the pulses have a quasi-rectangular shape and can be approximated according to\(^{28}\) by the following function

\[
s(x) = \begin{cases} 
\exp\left\{ -\frac{\pi}{2} \frac{(2x-1)^2}{(k-1)^2} \right\}, & x > \frac{1}{2k}, \\
1, & |x| \leq \frac{1}{2k}, \\
\exp\left\{ -\frac{\pi}{2} \frac{(2x+1)^2}{(k+1)^2} \right\}, & x < -\frac{1}{2k}.
\end{cases}
\] (3)

Here \(k\) is the coefficient of the pulse squareness, it equals the ratio of pulse duration at the intersection of H level \(t_0\) to the pulse duration \(t_1, k = \tau_0 / \tau_1\) (see Fig. 1).

The objective of this research is to synthesize the most effective estimation of the repetition period of quasi-rectangular pulses against background noise and interference, under conditions of a priori uncertainty of its repeatability (period and phase). An additional complication of the model is the possible loss of a pulse in the sequence according to a pseudo-random law, which is not known a priori. The known parameters are: off-duty factor (excluding loss), squareness coefficient and the times when the signal leading and trailing edges cross the H level.

**Methodology**

**Description of maximum likelihood estimation method.** The measurement process of the pulse repetition period is derived from the process of measuring the time of the beginning (end) of the pulse. Taking into account that the parameter (classical in the theory of maximum likelihood estimates) is non-energetic, the lower boundary of the estimated variance of the arrival time for each pulse as defined by Eq. (5), is presented as:

\[
D\{\hat{t}|\hat{t}\} = -1/q^2 \Psi''(0)
\] (5)

where \(\hat{t}\) is the MLE of the pulse arrival time. Assuming that the uncertainty function is stationary with respect to the parameter \(t - \hat{t}\) it gives the following expression:

\[
\Psi(t - \hat{t}) = \frac{1}{2E} \int_{-\infty}^{\infty} s(t - \hat{t}) s^*(t) dt
\] (6)

where \(E\) is a signal energy. After applying Parseval theorem and double differentiation (3) dispersion of MLE for the parameter \(t\) is bounded by the expression below:

\[
D\{\hat{t}|\hat{t}\} = \frac{\int_{-\infty}^{\infty} |S(f)|^2 df}{4\pi^2 q^2 \int_{-\infty}^{\infty} f^2 |S(f)|^2 df}
\] (7)

where \(S(f)\) is the Fourier transformation of the complex envelop of the signal \(s(t)\). Taking into account the approximation in which the pulse is rectangular\(^{30}\) with duration \(t_0\) (passed through a perfect filter with passband \(\Delta f\), where \(\Delta f t_0 = 1\)), the following assumption can be used (\(S(f) = |\sin(\pi f t_0)| / \pi t_0\) and normalization condition for \(S(f)\)\(^{14}\)):

\[
\int_{-\infty}^{\infty} f^2 |S(f)|^2 df \approx \frac{1}{\pi^2 t_0^2}.
\] (8)

In this case, if a sequence of pulses with an off-duty factor \(v = 1/t_0 = 2\) (square wave) is an input of the meter, the lower boundary of the estimated variance of the arrival time for each pulse as defined by Eq. (5), is presented as:

\[
D\{\hat{t}|\hat{t}\} = \frac{T^2}{16\pi^2 q^2}
\] (9)

The most common method for estimating the period is based on measuring the difference between the start times of adjacent pulses. If the assessment is made by \(n\) adjacent periods, then its relative error decreases \(n\) times
If the period is estimated using the information contained in the statistics of the pulse end times (crossing threshold by the trailing edge), then the estimate can be improved by $\sqrt{2}$. Thus, we can assume that for traditional estimates the video pulse repetition with an off duty factor $= 2$, the relative root-mean-square (RMS) error is

$$\delta_0 = \frac{1}{4\sqrt{2\pi} qn}$$

The estimate (11) is the asymptotic accuracy bound for the method (the limiting value of the Cramer-Rao inequality). An example of the traditional estimation of the pulse repetition period by the MLE method is the expression:

$$\tilde{T} = \frac{1}{2n} \sum_{k=1}^{n} \left( \frac{t_k^+ - t_{k-1}^+ + t_k^- - t_{k-1}^-}{i_k - i_{k-1}} \right)$$

where, $t_k^+$, $t_k^-$ are the times of intersection of the leading and trailing pulse edges (10), and $i_k$ is the number of the pulse in the sequence. When determining the numbers $i_k$, all pulses are considered, even those that were missed and did not fall into the sample.

**Description of complete sufficient statistics estimation method.** Consider the statistics $t^* = [t_0^+, ..., t_n^+]$ and $\mathbf{r} = [r_0, ..., r_n]$, where $t_i^+$ and $t_i^-$ are given (2) to synthesize the estimate of the period $T$. To find the joint probability density of the statistics $t^*$ and $\mathbf{r}$, the results of13–15,26,27 are used. Suppose that with a probability close to one, the crossing of the level $H$ will occur in disjointed bounds ($t_{\min}^*, t_{\max}^*$) and ($t_{\min}^{-}, t_{\max}^{-}$) at points $t_{0i}^*$ and $t_{0i}$ (Fig. 1). If the probabilities of false crossings of the level $H$ are small and the condition below is satisfied:

$$\Phi \left( \frac{U_m|s'(t_{0i})|}{\sigma_i} \right) + \frac{\sigma_i}{U_m|s'(t_{0i})|} \exp \left\{ \frac{U_m^2|s'(t_{0i})|^2}{\sigma_i^2} \right\} \approx 1$$

where $\Phi(x)$ is the Laplace integral, $\sigma_i^2 = \sigma^2 r''(0)$ is the variance of the derivative of the process $\eta(t)$, and $r''(0)$ is the value of the second derivative of the normalized correlation function $r(t)$ at $t = 0$, then the one-dimensional probability distribution densities at level $H$ is as follows:

$$\rho(t_i^+) = \frac{1}{\sqrt{2\pi} \sigma_i^2} \exp \left\{ -\frac{(t_i^+ - t_{\min}^+ - iT)^2}{2\sigma_i^2} \right\},$$

$$\rho(t_i^-) = \frac{1}{\sqrt{2\pi} \sigma_i^2} \exp \left\{ -\frac{(t_i^- - t_{\min}^+ - T/v - iT)^2}{2\sigma_i^2} \right\},$$

where $\sigma_+ = 1/[q|s'(t_{0i})]|$, $\sigma_- = 1/[q|s'(t_{0i})]|$.

Then, under the assumption that the processes within the intervals of the leading and trailing edges of the pulse will be statistically independent, the joint probability density distribution is obtained for the sample vectors $t^*$ and $\mathbf{r}$:

$$\rho(t^+, \mathbf{r}) = \prod_{i=0}^{n} \rho(t_i^+) \rho(t_i^-) = \left( \frac{1}{\sqrt{2\pi} \sigma_+ \sigma_-} \right)^{n+1} \exp \left\{ -\sum_{i=0}^{n} \frac{(t_i^+ - t_{\min}^+ - iT)^2}{2\sigma_+^2} - \sum_{i=0}^{n} \frac{(t_i^- - t_{\min}^+ - T/v - iT)^2}{2\sigma_-^2} \right\}.$$
\[ \mu = g(Y) = \sum_{i=1}^{M} \sum_{j=0}^{N} a_{ij}m_i^j(Y), \quad -\infty < a_{ij} < +\infty \]  

(16)

Then the only effective estimate of the parameter \( \mu \) is given\textsuperscript{13,29} by the expression:

\[ \mu^* = \sum_{i=1}^{M} \sum_{j=0}^{N} a_{ij}Y_i^j \]  

(17)

Here, \( M \) is the dimension of the complete sufficient statistics \( Y = \{Y_1, Y_2, \ldots, Y_m\} \), \( N \) is the maximum number of the moment order, \( m_i^j(Y) = M\{Y_i^j\}, i = 1, \ldots, M; j = 0, \ldots, N; m_i^0(Y) = 1. \)

Let’s define complete sufficient statistics \( X = \{X_1, X_2, \ldots, X_m\} \).

Considering the conditions of the problem \( (\sigma_+ = \sigma_-) \), the joint probability distribution density \( \rho(t^+, t^-) \) has the following form:

\[ \rho(t^+, t^-) = \left( \frac{1}{2\pi \sigma_+^2} \right)^{n+1} \exp \left\{ -\frac{1}{\sigma_+^2} \sum_{i=0}^{n} \left( t_i^+ - t_{i00}^+ - iT \right)^2 - \frac{1}{\sigma_-^2} \sum_{i=0}^{n} \left( t_i^- - t_{i00}^- - T/v - iT \right)^2 \right\}. \]  

(18)

A practical way to find sufficient statistics \( Y(x) \) is to use the factorization theorem\textsuperscript{28,31,32}, according to which the conditional probability density distribution \( \rho_\theta(x) = P(X = x|\theta) \) must be represented as:

\[ \rho_\theta(x) = h(x)g(\theta, Y(x)) \]

The resulting expression for \( Y(x) \) is a sufficient statistic.

We factorize the conditional probability density \( \rho(t^+, t^-) \) (18) for fixed parameters \( \sigma_+, t_{00}^+, T \). We denote:

\[ \theta = \{\theta_1, \theta_2, \theta_3\} = \left\{ \frac{t_{00}^+}{\sigma_+^2}, \frac{T}{\sigma_+^2}, \frac{1}{\sigma_+^2} \right\} \]  

(19)

Considering the first sum in (18) and collecting the terms containing random variables with the parameters \( \theta_1, \theta_2, \theta_3 \), the remaining terms are assigned to the function \( C_2(\theta_1, \theta_2, \theta_3) \):

\[ -\sum_{i=0}^{n} \frac{(t_i^+ - t_{i00}^+ - iT)^2}{2\sigma_+^2} = -\frac{1}{\sigma_+^2} \sum_{i=0}^{n} (t_i^+)^2 - 2t_{i00}^+ - 2iTt_i^+ + 2it_{i00}^+ + (iT)^2 + (t_{00}^+)^2 \]

\[ = -\frac{1}{\sigma_+^2} \sum_{i=0}^{n} (t_i^+)^2 + \frac{t_{00}^+}{\sigma_+^2} \sum_{i=0}^{n} t_i^+ + \frac{T}{\sigma_+^2} \sum_{i=0}^{n} t_i^+ i \]

\[ = \frac{t_{00}^+}{\sigma_+^2} \frac{n(n+1)}{2} - \frac{(t_{00}^+)^2}{2\sigma_+^2} - \frac{T^2}{\sigma_+^2} \frac{n(n+1)(2n+1)}{6} \]

\[ \Rightarrow \theta_1 \sum_{i=0}^{n} t_i^+ + \theta_2 \sum_{i=0}^{n} t_i^+ i + \theta_3 \sum_{i=0}^{n} (t_i^+)^2 + C_1(\theta_1, \theta_2, \theta_3), \]

\[ C_2(\theta_1, \theta_2, \theta_3) = \frac{(n+1)\theta_1^2}{4\theta_3} + \frac{(n+1)\theta_1\theta_2}{2\theta_3} + \frac{\theta_2^2}{4\theta_3} n(n+1)(2n+1) \]  

(20)

Similarly, an expression for another sum (18) gets:

\[ -\sum_{i=0}^{n} \frac{(t_i^- - t_{i00}^- - T/v - iT)^2}{2\sigma_+^2} = \theta_1 \sum_{i=0}^{n} t_i^- + \theta_2 \sum_{i=0}^{n} t_i^- (i + 1/v) + \theta_3 \sum_{i=0}^{n} (t_i^-)^2 + C_2(\theta_1, \theta_2, \theta_3), \]

\[ C_2(\theta_1, \theta_2, \theta_3) = \frac{(n+1)\theta_1^2}{4\theta_3} + \frac{(n+1)\theta_1\theta_2}{2v\theta_3} + \frac{\theta_2^2}{4\theta_3} \frac{(n+1)(nv+1)}{\nu^2} \]  

(21)

After factorization, the probability density distribution \( \rho(t^+, t^-) \) (18), taking into account (19) and the transformations performed above, takes the form

\[ \rho(t^+, t^-) = C(\theta) \exp \{ \theta_1 Y_1 + \theta_2 Y_2 + \theta_3 Y_3 \} \]  

(22)

where

\[ Y_1 = \sum_{i=0}^{n} (t_i^+ + t_i^-) ; \quad Y_2 = \sum_{i=0}^{n} \left[ it_i^+ + (i + 1/v)t_i^- \right] ; \quad Y_3 = \sum_{i=0}^{n} \left[ (t_i^+)^2 + (t_i^-)^2 \right] \]  

(23)
\[ C(\theta) = \left( \frac{\partial_1}{\pi} \right)^{n+1} \exp \left\{ \frac{(n+1)\sigma_2^2}{2\sigma_3} + \frac{(n+1)(n+1/v)\partial_1\partial_2}{2\sigma_3} + \frac{\partial_2^2}{4\sigma_3} \left[ \frac{n(n+1)(2n+1)}{6} + \frac{(n+1)(nv+1)}{v^2} \right] \right\} \]  

(24)

Here the parameter \( C(\theta) \) is obtained as \( C(\theta) = (-\theta/n)^{n+1} \exp[C_0(\theta) + C_2(\theta)] \), where \( C_0(\theta) \) and \( C_2(\theta) \) are represented by formulas (20) and (21). Thus, the conditional probability density \( \rho(t^+, t^-) \) takes the form:

\[ \rho(t^+, t^-) = h(t^+, t^-) g(\theta, Y(t^+, t^-)) \]

where \( h(t^+, t^-) \equiv 1 \) and \( g(\theta, Y) = C(\theta) \exp \{ Y \cdot \theta \} \) and according to the factorization theorem, the statistics \( Y \) expressed by formula (23) are sufficient statistics.

Distribution density \( \rho(t^+, t^-) \) belongs to the exponential distribution with sufficient statistics \( Y = \{ Y_1, Y_2, Y_3 \} \) and parameter \( \theta \). For unknowns \( t^+_0, T, \sigma \), the parameter takes values from the region \( (0, \infty) \times (0, \infty) \times (-\infty, 0) \). We use the completeness theorem, according to which if the probability distribution density belongs to the class of exponential distributions and the set of parameter values contains an \( n \)-dimensional interval, then the statistics \( Y \) is complete. Then we can apply the theorem on the uniqueness of the effective estimate.

Let us express the period in terms of the mathematical expectation of the complete sufficient statistic \( Y \). We calculate the mathematical expectation of \( M(Y_1) \) and \( M(Y_2) \), and obtain the system of equations for \( t^+_0, T, \sigma \). We get expressions for \( Y_1 \) and \( Y_2 \) giving similar ones:

\[ Y_1 = \sum_{i=0}^{n} (\Delta t^+_i + \Delta t^-_i) + 2(n+1)t^+_0 + (n+1) \left( n + \frac{1}{v} \right) T, \]

\[ Y_2 = \sum_{i=0}^{n} (\Delta t^+_i + \Delta t^-_i \left( i + \frac{1}{v} \right)) + (n+1) \left( n + \frac{1}{v} \right) t^+_0 + (n+1) \left( \frac{3nv + v^2n(2n+1) + 3}{3v^3} \right) T. \]

(25)

Considering that the mathematical expectation of small random additions due to the action of noise with a normal distribution is equal to zero, we obtain the following complete system of linear equations with two unknowns \( t^+_0 \) and \( T \):

\[
\begin{align*}
M(Y_1) &= 2(n+1)t^+_0 + (n+1) \left( n + \frac{1}{v} \right) T, \\
M(Y_2) &= (n+1) \left( n + \frac{1}{v} \right) t^+_0 + (n+1) \left( \frac{3nv + v^2n(2n+1) + 3}{3v^3} \right) T.
\end{align*}
\]

(26)

Having solved the system by Cramer’s method, we obtain the required representation of the period in terms of the first moments of the complete statistic \( Y \):

\[ T = \frac{6v^2}{(n+1)(n^2v^2 + 2nv^2 + 3)} \left[ M(Y_2) - \frac{nv + 1}{2v} M(Y_1) \right] \]

(27)

Thus, the estimate of the pulse repetition period (27) is presented in the form (16). According to the uniqueness theorem for the effective estimate, the only effective estimate of the pulse repetition period is:

\[ \hat{T} = \frac{6v^2}{(n+1)(n^2v^2 + 2nv^2 + 3)} \left[ Y_2 - \frac{nv + 1}{2v} Y_1 \right] \]

(28)

Taking into account the expressions for \( Y_1 \) and \( Y_2 \), the effective estimate of a pulses period is:

\[ \hat{T} = \frac{3v}{(n+1)(n^2v^2 + 2nv^2 + 3)} \sum_{i=0}^{n} [(2iv - nv - 1)t^+_i + (2iv - nv + 1)t^-_i]. \]

(29)

The variance of the estimate \( \hat{T} \) is given by

\[
\begin{align*}
D(\hat{T}) &= \frac{(3v)^2}{(n+1)^2(n^2v^2 + 2nv^2 + 3)^2} \sum_{i=0}^{n} [(2iv - nv - 1)^2 D(t^+_i) + (2iv - nv + 1)^2 D(t^-_i)] \\
&= \frac{(3v)^2\sigma_2^2}{(n+1)^2(n^2v^2 + 2nv^2 + 3)^2} \sum_{i=0}^{n} [(2iv - nv - 1)^2 + (2iv - nv + 1)^2], \\
D(\hat{T}) &= \frac{6v^2}{(n+1)(n^2v^2 + 2nv^2 + 3)} \left[ \frac{q(s_0)}{q(s^+_0)} \right]^2
\end{align*}
\]

(30)

Here we assume that \( D(t^+_i) = D(t^-_i) = \sigma_2^2 = 1 \left( \frac{q(s^+_0)}{q(s^+_0)} \right)^2. \)

In order to exclude the parameter \( s'(t^+_0) \) from the number of a priori unknowns, its estimate in the approximation of a quasi-rectangular pulse shape is obtained. Based on the criterion of maximum likelihood, the variance \( D(\hat{T}) \) will be minimal for such values of the threshold \( H \), at which the steepness of the signal front at
the point of intersection is maximum, i.e. we investigate the derivative of the function \( s'(x) \) to the maximum absolute value. Therefore, we find the second derivative \( s''(x) \) (31), equate it to zero and obtain Eqs. (32) and (33)

\[
s''(x) = \begin{cases} 
\frac{\pi k^2}{4(k-1)^4} \exp\left\{ -\frac{\pi (2xk-1)^2}{8(k-1)^2} \right\} (4\pi x^2 k^2 - 4\pi xk - 4k^2 + 8k - 4 + \pi), & x > \frac{1}{2k}, \\
0, & |x| \leq \frac{1}{2k}, \\
\frac{\pi k^2}{4(k-1)^4} \exp\left\{ -\frac{\pi (2xk+1)^2}{8(k-1)^2} \right\} (4\pi x^2 k^2 + 4\pi xk - 4k^2 + 8k - 4 + \pi), & x < -\frac{1}{2k},
\end{cases}
\]

(31)

4\pi x^2 k^2 - 4\pi xk - 4k^2 + 8k - 4 + \pi = 0

(32)

4\pi x^2 k^2 + 4\pi xk - 4k^2 + 8k - 4 + \pi = 0

(33)

Each of these equations has two solutions, but we will only consider roots that are greater than \( 1/2k \) for (32) and less than \(-1/2k\) for (33) for \( k > 1 \), since the other roots have no physical meaning. The selected solutions corresponding to Eqs. (32) and (33) have the following form:

\[
x_{00}^+ = \frac{\pi + 2\sqrt{\pi} k^2 - 2\pi k + \pi}{2\pi k}
\]

(34)

\[
x_{00}^- = \frac{-\pi - 2\sqrt{\pi} k^2 - 2\pi k + \pi}{2\pi k}
\]

(35)

The derivative of the dimensionless signal \( s'(x) \) takes the minimum value at \( x = x_{00}^+ \), and the maximum value at \( x = x_{00}^- \), but in absolute value they are equal and maximum

\[
|s'(x_{00}^+)| = |s'(x_{00}^-)| = \frac{\sqrt{\pi} k}{4\sqrt{\pi} (k-1)}
\]

(36)

The physical meaning of the obtained variables can be stated as follows: \( x_{00}^+ \)—the initial phase of the quasi-rectangular signal \( s(x) \), i.e. the first crossing of the \( H \) level by the signal leading edge, at which the variance of the signal period estimate \( T \) reaches a minimum value; \( x_{00}^- \)—similarly, only for the trailing edge.

By inverting the variables \( t_0 = \sqrt{T} x_0 \), we pass to the signal \( s(t) \), where \( t \) is no longer a dimensionless quantity, and time has the dimension of time. Then the desired values \( t_{00}^+ \) and \( s'(t_{00}^+) \) will take the following form:

\[
t_{00}^+ = t_0 x_{00}^+ = t_0 \frac{\pi + 2\sqrt{\pi} k^2 - 2\pi k + \pi}{2\pi k}
\]

(37)

\[
|s'(t_{00}^+)| = \frac{\sqrt{\pi} k}{t_0 \sqrt{\pi} (k-1)}
\]

(38)

Taking into account the assumption of the signal off-duty factor \( v = T/t_0 = 2 \), we present \( s'(t_{00}^+) \) in terms of the period:

\[
|s'(t_{00}^+)| = \frac{2\sqrt{\pi} k}{T \sqrt{\pi} (k-1)}
\]

(39)

Substituting the obtained derivative value and the signal off-duty factor value, \( v = 2 \), in (30), for the variance of the signal period estimate, we obtain:

\[
D(\hat{T}) = \frac{6}{(n+1)(4n^2 + 8n + 3)} \frac{eT^2(k-1)^2}{q^2 \pi k^2}
\]

(40)

**Comparison and analysis of methods**

We find the relative error in estimating pulse repetition period obtained by the method of complete sufficient statistics. The relative error for \( n \) measurements is determined as follows:

\[
\delta_1 = \frac{\sqrt{D(\hat{T})}}{T} = \frac{\sqrt{6e}}{\sqrt{\pi} \frac{1}{(n+1)(4n^2 + 8n + 3)}} \frac{(k-1)}{q \pi k}
\]

(41)

Consider the relative error \( \delta_0 \), which depends on the signal-to-noise ratio \( q \) and the number of measurements \( n \). Assume that the squareness coefficient of the pulse \( k \) is \( 1.2 \) (\( k = 1.2 \)). In practice, a differential scheme may be used to estimate \( k \), which allows to conclude about the steepness of the leading and trailing edges of the pulse:
Figure 2 shows the ratio of the relative errors $\delta_0 / \delta_1$, where $\delta_0$ and $\delta_1$ are relative errors of the MLE and CSS methods respectively. The figure shows that this ratio is greater than 1 for $n > 8$ and increases with rise in the sample size. This indicates that the estimate obtained by CSS method is more efficient than the MLE estimate for $n > 8$.

The greatest gain from using the CSS method can be obtained at low values of the signal-to-noise ratio, when the accuracy of both methods is low. Let us fix the number of measurements (for example, $n = 16$) to estimate the energy gain of the CSS methodology in comparison with traditional approaches for the finite and small sample size. The dependencies $\delta_0$ and $\delta_1$ on $q$ in the range of 0–10 dB will give a difference in the achieved signal-to-noise ratio of the order of 2 dB in favor of an estimate using the CSS methodology (Fig. 3).

**Numerical modeling, results and discussion**

Expressions (11) and (42) only allow us to theoretically estimate the energy gain of the CSS method in comparison with the MLE when solving the problem of estimating the pulse repetition period. At the same time, to verify such estimates, numerical simulations were performed in the MATLAB environment. For this purpose, a sequence of pulses were generated against a background of additive white Gaussian noise with intra-pulse quadrature phase shift keying (QPSK). The order of occurrence of the pulse at a given time position and its carrier frequency in the receiving band were determined by a pseudo-random sequence generator.

Two alternative algorithms for estimating the pulse repetition period were applied to the resulting implementation: based on the MLE and CSS methods, computed according to the formulas (12) and (29). The number of pulses in the implementation $n$ ranged from 10 to 50, and the signal-to-noise ratio $q$ ranged from 6 to 22 dB.

Figure 3 shows the ratio of the relative errors $\delta_0 / \delta_1$ obtained by the MLE $\delta_0$ and CSS $\delta_1$, calculated using formulas (11) and (42) for a different number of pulses in the sample $n$. The figure shows that this ratio is greater than 1 for $n > 8$ and increases with rise in the sample size. This indicates that the estimate obtained by CSS method is more efficient than the MLE estimate for $n > 8$.

The signal-to-noise ratio (SNR) is plotted along the X axis.
The results of the estimation for \( n = 16 \) as a function of \( q \) are shown in Fig. 3. The relative errors of the two alternative algorithms are presented. It is easy to see that the energy gain of the CSS method is more than 2 dB in the \( q \) interval up to 12 dB. The difference is reduced to 1 dB with an increase in the signal-to-noise ratio \( q \), while it remains that the traditional estimate asymptotically tends to the unique effective one when \( q \to \infty \).

In addition, numerical modeling demonstrated the stability of the algorithm. The algorithm effectively estimates the pulse repetition period using the CSS methodology with respect to variations in the threshold level \( H \) (red line in Fig. 1) over a wide range (up to 10 dB). With the same variation, the estimate synthesized on the basis of the MLE approach is not stable.

Let us compare the computational costs required to find the pulse repetition period using the MLE and CSS methods. Both methods are linear with respect to the sample size of signal pulses (they belong to the complexity class \( O(n) \)). In the case of the traditional estimation of the pulse repetition period, obtained by the MLE method according to formula (12), the number of addition/subtraction \( CMLE \), multiplication \( C^{MLE} \) and division \( C^{MLE} \) operations after optimization depends on the number of pulses in the sample as follows

\[
C^{MLE}_+(n) = 3n + 1, \quad C^{MLE}_-(n) = 0, \quad C^{MLE}_/(n) = 1 + n
\]  

(43)

In the case of the pulse repetition period estimation obtained by the CSS method according to formula (27), the number of operations of addition/subtraction \( C^{CSS}_+, \) multiplication \( C^{CSS}_* \) and division \( C^{CSS}_/ \) after optimization is

\[
C^{CSS}_+(n) = 3(n + 1) + 5, \quad C^{CSS}_*(n) = 2(n + 1) + 6, \quad C^{CSS}_/(n) = 1
\]  

(44)

Comparing (43) and (44), we conclude that the CSS method requires more operations, however, additional optimization of expression (29) for a specific processor architecture or logical cell array (LCA) is possible, examples of which are described in \(^{13,15} \). This optimization is a separate engineering problem and is beyond the scope of this work.

Conclusion

The methodology of complete sufficient statistics is applied to solve the problem of finding an effective estimate of the pulse repetition period. Taking into account the probability distribution functions, a mathematical apparatus is constructed that allows using the Lehmann–Scheffe theorem and its corollary. An analytical expression for the effective estimation of the pulse repetition period is obtained.

Expressions for the variance and relative error of the pulse repetition period estimation are obtained using the CSS methodology depending on the duty-off-factor, squareness coefficient, signal-to-noise ratio and sample size. In the framework of the estimation theory, the lower bound of the accuracy of the estimation of the pulse repetition period is calculated. The relative errors of estimates are compared using traditional methods of MLE and the CSS methodology.

It is shown that for a finite and small value of the sample size \( (n = 16) \) in the range of signal-to-noise ratios of 5–12 dB, the relative root-mean-square error of estimates for the pulse repetition period by CSS method is significantly less than the error of the traditional estimate by MLE method.

The results obtained have a wide range of applicability and can be used in the design of control and measuring equipment in the oil and gas industry, in the development of medical equipment, in the field of telecommunications, in the design of pulse-Doppler radars, etc.

Data availability

The datasets generated during and/or analyzed during the current study are available from the first author on reasonable request.

Received: 23 February 2022; Accepted: 15 November 2022
Published online: 19 November 2022

References

1. Van Trees, H. Detection, Estimation, and Modulation Theory, Part I (Wiley, 1968).
2. Abratkiewicz, K., Samczyński, P. & Czarnecki, K. Radar signal parameters estimation using phase accelerogram in the time-frequency domain. IEEE Sens. J. 19(13), 5078–5085. https://doi.org/10.1109/JSEN.2019.2903352 (2019).
3. Wu, N., Li, B., Wang, H., Hou, L. & Kuang, J. Parametric message passing-based relative navigation in joint tactical information distribution system. J. Syst. Eng. Electron. 27, 81–89. https://doi.org/10.1109/JSEE.2016.00009 (2016).
4. Ravi Shankar Reddy, G. & Rao, R. Oscillatory-plus-transient signal decomposition using TQWT and MCA. J. Electron. Sci. Technol. 17(2), 135–151. https://doi.org/10.1109/JEST.1674-862X.607/1911.19 (2019).
5. Matuszewski, J. The analysis of modern radar signals parameters in Electronic Intelligence system. in Proceedings of the 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science, 298–302 (TCSET, 2016). https://doi.org/10.1109/TTCSET.2016.7452040.
6. Abe, H. et al. Improving the quality of XAFS data. J. Synchrotron Rad. 25, 972–980. https://doi.org/10.1107/S1600577518006602 (2018).
7. Zhuchkov, K. N., Shuvaeva, V. A., Yagi, K. & Terauchi, H. Deglitching procedure for XAFS. J. Synchrotron Rad. 8, 302–304. https://doi.org/10.1107/S0909049500020951 (2001).
8. Terauchi, H. et al. Observation and numerical analysis of plasma parameters in a capillary discharge-produced plasma channel waveguide. J. Phys. D Appl. Phys. 109, 053304. https://doi.org/10.1063/1.3560529 (2011).
9. Morel, B. et al. A solver based on pseudo-spectral analytical time-domain method for the two-fluid plasma model. Sci. Rep. 11, 3151. https://doi.org/10.1038/s41598-021-82173-9 (2021).
10. Vasilenko, M., Zavyalov, A. & Zhuchkov, K. Increasing the stability of a spatially distributed information system using a robust algorithm for filtering anomalous measurements. IT Ind. 3(8), 1–7 (2020).
11. Zhuchkov, K. N., Zasyyalov, A. P. & Lukyanov, V. A. On the necessity to develop approaches to the assessment of reliability parameters of equipment and pipelines of gas transport systems. Gas Ind. 11(745), 54–60 (2016).
12. Oparin, V., Vostretsov, A., Krivetsky, A., Buzyaev, A. & Yakovitskaya, G. Integrated measurement and recording of loads, displacements and electromagnetic emission in rocks under uniaxial compression. J. Min. Sci. 47, 547–557. https://doi.org/10.1134/S1062739114050017 (2011).
13. Lehmann, E. L. & Scheffé, H. Completeness, similar regions, and unbiased estimation: Part I Sankhyā. Indian J. Stat. 10(4), 305–340 (1950).
14. Parkhomenko, N. G., Khoruzhiy, S. G., Zhuchkov, K. N., Aleksandrov, V. P. & Stepantov, I. M. Perspective approaches in signal parameter estimation problems. Mar. Radioelectron. 16(2), 46–48 (2006).
15. Kay, S. M. Fundamentals of Statistical Signal Processing: Estimation Theory (Prentice Hall PTR, 1985).
16. Levin, B. R. Theoretical Foundations of Statistical Radio Engineering Vol. 2 (Sov Radio, 1975) (in Russian).
17. Borovkov, A. A. Mathematical Statistics (Nauka, 1984) (in Russian).
18. Xu, L. & Ding, F. Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. Circuits Syst. Signal Process 36(4), 1735–1753. https://doi.org/10.1007/s00034-016-0378-4 (2017).
19. Wu, M.-W., Jin, Y., Li, Y., Song, T. & Kam, P.-Y. Maximum-likelihood, magnitude-based, amplitude and noise variance estimation. IEEE Signal Process. Lett. 28, 414–418. https://doi.org/10.1109/LSP.2021.3055464 (2021).
20. Li, Y., Zhao, H., Fan, W. & Shen, C. Extended noise resistant correlation method for period estimation of pseudoperiodic signals. IEEE Trans. Instrum. Meas. 70, 1–11. https://doi.org/10.1109/TIM.2021.3052027 (2021).
21. White, A., Hong, J.-W., Hong, S. & Choi, J. Parameter estimation for wavelet transformed ultrasonic signals. NDT E Int. 44(1), 32–40. https://doi.org/10.1016/j.ndteint.2010.09.003 (2011).
22. Atamanskiy, D. V. & Sadovyi, K. V. Multithreshold procedure for evaluating parameters of meteorological objects against the background of local objects clutter in pulsed doppler weather radars. Radioelektron. Commun. Syst. 61, 317–324. https://doi.org/10.3103/S073527721807004X (2018).
23. Weight, D., Horry, R. & Skagerberg, E. Functions for traditional and multilevel approaches to signal detection theory. Behav. Res. Methods 41, 257–267. https://doi.org/10.3758/BRM.41.2.257 (2009).
24. Lo, J. & Bassu, D. Adaptive multilayer perceptrons with long- and short-term memories. IEEE Trans. Neural Netw. 13(1), 22–33. https://doi.org/10.1109/72.977262 (2002).
25. Krasnov, A. Yu., Zhuchkov, K. N. & Khoruzhiy, S. G. Determination of the initial phase of the clock oscillation for pulse signals with MSK in Proceedings of the 7th International conference on digital signal processing and its application, 192–195 (DSPA, 2005).
26. Leonov, G. A., Kuznetsov, N. V., Yuldashev, M. V. & Yuldashev, R. V. Hold-in, pull-in, and lock-in ranges of PLL circuits: Rigorous mathematical definitions and limitations of classical theory. Circ. Syst. I IEEE Trans. 62(10), 2454–2464. https://doi.org/10.1109/TCSI.2015.2476295 (2015).
27. Trifonov, A. P. & Bespalova, M. B. Quasi-plausible estimation of the arrival time and period of video pulses. Radiotechnika 3, 78–80 (1998).
28. Bogdanovich, V. A. & Vostretsov, A. G. The Theory of Robust Detection, Discrimination and Estimation of Signals (Physmathlit, 2004) (in Russian).
29. Lyne, A. & Rickett, B. Measurements of the pulse shape and spectra of the pulsating radio sources. Nature 218, 326–330. https://doi.org/10.1038/218326a0 (1968).
30. Fisher–Neyman Factorization Theorem [Electronic Resource]. Wikipedia. Free encyclopedia. https://en.wikipedia.org/wiki/Sufficient_statistic#Fisher%E2%80%93Neyman_factorization_theorem.
31. Lehmann, E. L. & Casella, G. Theory of Point Estimation (Springer, 2006).
32. Cramer, G. Introduction à l’Analyse des lignes Courbes algébriques. (Europeana, 1750).
33. Govorukhina, A., Zhuchkov, K. & Khoruzhiy, S. Code optimization methods for the TMS320C6000 digital signal processor. Dig. Signal Process. 4, 47–56 (2004).
34. Wallace, R., Marvuglia, A., Benetto, E. & Tiruta-Barna, L. A new approach to optimization with life cycle assessment: combining optimization with detailed process simulation. in Computational Science and Its Applications—ICCSA 2014. Lecture Notes in Computer Science, vol. 8581. (Springer, 2022). https://doi.org/10.1007/978-3-319-09150-1_52.
