

Parameter Identification Using Bayes and Kernel Approaches

PIOTR KULCZYCKI

Faculty of Electrical Engineering
Cracow University of Technology
Cracow, Poland

(Received July 4, 1997; Accepted July 15, 1998)

ABSTRACT

This paper deals with the task of parameter identification (point estimation) using the Bayes approach. The calculation procedures are based on the kernel estimators technique. As a result of presented considerations, a complete algorithm usable for obtaining the value of the estimated parameter is worked out. An elaborated method is provided for numerical computations, including computer systems working in the real-time regime.

Key Words: parameter identification, point estimation, Bayes approach, kernel estimator, numerical algorithm

I. Introduction

One of the elementary issues in contemporary engineering is parameter identification, i.e., the specification of the values of the parameters occurring in the model being used. If the measurements obtained are treated as the sum of the "true" value of the parameter and the random disturbances, then the task from the mathematical point of view becomes a typical problem for point estimation – a fundamental discipline of mathematical statistics. The primary task here consists in calculating, on the basis of the obtained measurements of the given parameter, an appraisal of its value, i.e., an estimator. In practice, simple procedures are typically used, e.g., the least squares or the highest reliability methods. However, the so-called Bayes estimation offers a number of important advantages (Lehmann, 1983). This approach allows for a more universal use of accessible information about the subject of estimation, also outside the statistical sample; in particular, it affords possibilities for taking into account the consequences of estimation errors, which has considerable significance in modern engineering. Some of these errors often have only a minor impact on the quality of work of the device while others have far more profound influence, not excluding system failure.

The final result of the considerations presented in this paper will be a complete usable algorithm for specifying the value of the Bayes estimator on the basis

of independent measurements obtained experimentally. For calculation procedures, the kernel estimators technique is used. Since any sort of judgmental statistical research has been eliminated here through the application of optimizing criteria, the proposed method may be successfully adapted to numerical calculational procedures. The speed of the algorithms makes it also suitable for systems working in a real-time regime.

II. Bayes Estimation

Assume the probability space (Ω, Σ, P) , where Ω means the set of elementary events, Σ is its σ -algebra, and P denotes a probability. Let the random variable $X: \Omega \rightarrow \mathbb{R}$ represent the measurement process while its realizations will be interpreted as the particular results of the independent measurements obtained for the estimated parameter. Consider also the loss function $l: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \cup \{\pm\infty\}$; its values $l(\hat{x}, x)$ denote the losses which may be incurred by assuming \hat{x} as the estimator whereas, in reality, the estimated parameter is x . If for every $\hat{x} \in \mathbb{R}$ the integral $\int_{\Omega} l(\hat{x}, X(\omega)) dP(\omega)$ exists, let $l_b: \mathbb{R} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ be a function of the Bayes losses

$$l_b(\hat{x}) = \int_{\Omega} l(\hat{x}, X(\omega)) dP(\omega). \quad (1)$$

Then, every element $\hat{x}_b \in \mathbb{R}$ such that

$$l_b(\hat{x}_b) = \inf_{\hat{x} \in \mathbb{R}} l_b(\hat{x}) \quad (2)$$

is known as a Bayes estimator. Note that with fixed \hat{x} , the value of the function of the Bayes losses constitutes the expected value of losses after the value \hat{x} is assumed.

In the present paper, consideration will be given to a nearly intuitive non-symmetrical form of the loss function:

$$l(\hat{x}, x) = \begin{cases} -p_1(\hat{x} - x) & \text{if } \hat{x} - x \leq 0 \\ p_2(\hat{x} - x) & \text{if } \hat{x} - x \geq 0. \end{cases} \quad (3)$$

where $p_1, p_2 > 0$. The constants p_1 and p_2 constitute, therefore, the coefficients of proportionality of losses suffered after obtaining the value of the estimator, either smaller or greater than the "true" value of the estimated parameter, respectively.

The quantile of order r ($0 < r < 1$) of the random variable X is defined as every number $q \in \mathbb{R}$ fulfilling the conditions

$$P(\{\omega \in \Omega : X(\omega) \leq q\}) \geq r \quad (4)$$

$$P(\{\omega \in \Omega : X(\omega) \geq q\}) \geq 1 - r. \quad (5)$$

If the distribution function F of this random variable is continuous, the quantile is given by the equation

$$F(q) = r \quad (6)$$

while the strong monotonicity of this function in the set $F^{-1}((0, 1))$ indicates its uniqueness.

In Appendix 1 can be found the proof that if the quantile of order r with

$$r = \frac{p_1}{p_1 + p_2} = \frac{p_2}{p_2 + 1} \quad (7)$$

is uniquely defined, then it constitutes the Bayes estimator for the loss function given by Eq. (3). The second part of the above equality shows that the order r does not depend on the parameters p_1 and p_2 themselves, but rather on their ratio.

The next section will present the procedure for calculating the value of the quantile of order r using the kernel estimators technique, which in accordance with the above, will complete the solution of the Bayes approach of the point estimation task considered in this paper.

III. Kernel Estimators

1. Kernel Estimators of the Density Function

In this subsection, the basic elements associated with kernel estimators of the density function are presented, with particular attention given to aspects that will be used in further parts of this paper. Illustrative considerations can be found in the book by Silverman (1986) while the mathematical side is the topic of Prakasa Rao (1983).

A. Basic Definitions

It will be assumed here that the distribution of the random variable X has the density function f . Its estimator $\hat{f}: \mathbb{R} \rightarrow [0, \infty)$ is calculated on the basis of the value of the m -elements simple random sample $x_1, x_2, \dots, x_m \in \mathbb{R}$. (The dependence of samples and estimators on the random factor $\omega \in \Omega$ will not be explicitly noted hereinafter.) The fundamental form of the kernel estimator can be defined by the dependence

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \quad (8)$$

where

$$h > 0 \quad (9)$$

while the measurable function $K: \mathbb{R} \rightarrow [0, \infty)$ fulfills the condition

$$\int_{\mathbb{R}} K(x) dx = 1 \quad (10)$$

and for every $x \in \mathbb{R}$:

$$K(x) = K(-x) \quad (11)$$

$$K(0) \geq K(x). \quad (12)$$

The function K is called the kernel whereas the constant h is known as the smoothing parameter. For an illustrative description of this concept, see e.g. Section 2.4 of Silverman (1986).

B. Choice of Kernel and Smoothing Parameter

The form of the kernel K and the value of the smoothing parameter h can be fixed on the basis of the minimum mean squared error criterion. (More details are found in Chapter 3 of Silverman (1986).) It is then additionally assumed that $f \in C^2$, and that the functions

f and f' are bounded. Let the following finite and non-zero quantities be given:

$$U(K) = \int_{\mathbb{R}} x^2 K(x) dx \quad (13)$$

$$V(K) = \int_{\mathbb{R}} K(x)^2 dx \quad (14)$$

$$W(f) = \int_{\mathbb{R}} f'(x)^2 dx. \quad (15)$$

The minimum mean squared error occurs for

$$h_o = \left(\frac{V(K)}{U(K)^2 W(f) m} \right)^{1/5}; \quad (16)$$

then, the value of the mean square functional J is

$$J(h_o) = \frac{5}{4} \left(\frac{U(K)^2 V(K)^4 W(f)}{m^4} \right)^{1/5}. \quad (17)$$

Equality (16) constitutes the basis for the procedure to calculate the value of the smoothing parameter h . The direct application of this formula encounters difficulties, however, since the expression $W(f)$ depends on a density function that is estimated, i.e., not known *a priori*.

The approximate value of the smoothing parameter can be calculated by assuming that for normal distribution with the standard deviation σ , one obtains

$$W(f) = \frac{3}{8} \pi^{-1/2} \sigma^{-5}. \quad (18)$$

Then, from Eq. (16), the result is that

$$h_o^* = \left(\frac{V(K)}{U(K)^2} \frac{8}{3} \pi^{1/2} \frac{1}{m} \right)^{1/5} \sigma. \quad (19)$$

In the general case, however, one calculates the value h_o^* which assumes the minimum of the function $g: \mathbb{R} \rightarrow [0, \infty)$ defined as

$$g(h) = \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \tilde{K} \left(\frac{x_j - x_i}{h} \right) + \frac{2}{mh} K(0), \quad (20)$$

where

$$\tilde{K}(x) = K^{*2}(x) - 2K(x) \quad (21)$$

and K^{*2} denotes the convolution square of the function K , i.e.,

$$K^{*2}(x) = \int_{\mathbb{R}} K(y) K(x-y) dy. \quad (22)$$

In practice, the quantity h_o^* constitutes a sufficient approximation of the optimal value of the smoothing parameter h_o .

Equality (17), on the other hand, provides indications regarding the choice of the type of kernel, i.e., the form of the function K . In accordance with this dependence, the minimized mean squared error is proportional to the expression $(U(K)^2 V(K)^4)^{1/5}$, but for the types of kernels applied in practice, the values of this expression do not much differ from each other, on which the exponent 1/5 also has an impact. As a result, it becomes possible in choosing the type of kernel to take into account primarily the properties of the estimator obtained, e.g., the class of regularity or the finiteness of the support, without significant worsening the statistical quality of estimation. In what follows, the exponential kernel

$$K_e(x) = \frac{e^{-x}}{(1+e^{-x})^2} \quad (23)$$

will be applied. Its primitive assumes a form convenient for further considerations:

$$I_e(x) = \frac{1}{1+e^{-x}}. \quad (24)$$

The convolution square, used in calculating the optimal value of the smoothing parameter, is then expressed by the formula

$$\tilde{K}_e^{*2}(x) = \begin{cases} e^x \left(\frac{x(e^x+1)}{(e^x-1)^3} - \frac{2}{(e^x-1)^2} \right) & \text{for } x \neq 0 \\ \frac{1}{6} & \text{for } x = 0 \end{cases} \quad (25)$$

whereas the value of the quantity occurring in Eq. (19) is

$$\frac{V(K_e)}{U(K_e)^2} = \frac{3}{2\pi^4} \quad (26)$$

C. Strong Consistency Property

If the function K is Borelian and fulfills the condition

$$\lim_{x \rightarrow \infty} xK(x) = 0 \quad (27)$$

while the value of the smoothing parameter h is selected

in such a way that

$$\lim_{m \rightarrow \infty} h = 0 \tag{28}$$

$$\lim_{m \rightarrow \infty} mh = \infty, \tag{29}$$

then at every point of continuity x of the density function f , the kernel estimator is strongly consistent, i.e.,

$$P(\lim_{m \rightarrow \infty} \hat{f}(x) = f(x)) = 1, \tag{30}$$

and, therefore, also consistent:

$$\lim_{m \rightarrow \infty} P(|\hat{f}(x) - f(x)| \geq \varepsilon) = 0 \text{ for every } \varepsilon > 0. \tag{31}$$

The proof can be found in the fundamental work by Parzen (1962). Note also that if the value of the parameter h is defined on the basis of Eqs. (16), (19) or (20), then Eqs. (28) and (29) are fulfilled.

Since kernel estimators are predominantly applied when the sample size is rather large, the property of consistency is of fundamental significance in considering estimators of this type.

D. Modification of the Smoothing Parameter

In many applications, it proves to be particularly advantageous to introduce the concept of modification of the smoothing parameter (see Section 5.3 of Silverman (1986)). Construction of the estimator can then be done in the following manner:

- (1) the kernel estimator \hat{f} is calculated in accordance with the procedure previously presented;
- (2) the modifying parameters $s_i > 0$ ($i=1, 2, \dots, m$) are stated in the form

$$s_i = \left(\frac{\hat{f}(x_i)}{b} \right)^{-a}, \tag{32}$$

where $a \in [0,1]$ while b is the geometric mean of the numbers $\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_m)$, given in the form of the logarithmic equation

$$\log(b) = m^{-1} \sum_{i=1}^m \log(\hat{f}(x_i)); \tag{33}$$

- (3) the kernel estimator with the modified smoothing parameter is defined using the formula

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^m \frac{1}{s_i} K\left(\frac{x-x_i}{hs_i}\right). \tag{34}$$

The considerations resulting from the criterion of minimum mean squared error point to the value

$$a = \frac{1}{2}. \tag{35}$$

Definition (8) is a particular case of Eq. (34) when $a=0$. The concept of modifying the smoothing parameter improves the quality of the estimation by respectively differentiating this parameter in areas of greater and lesser density of the random sample values. From the practical point of view, another essential feature of the estimator with the modified smoothing parameter consists in its slight sensitivity to the exactness of the choice of the constant h . In practice, this property is exceptionally advantageous, and when such an estimator is applied, it often proves sufficient to accept the approximate value h_o^* given by Eq. (19).

2. Kernel Estimators of the Distribution Function

To carry forward the concept presented in Subsection III.1, the natural estimator of the distribution function is the mapping $\hat{F} : \mathbb{R} \rightarrow [0,1]$ defined by the formula

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}(y) dy. \tag{36}$$

Equality (10) guarantees the existence of the primitive $I : \mathbb{R} \rightarrow [0,1]$ of the kernel K , i.e.,

$$I(x) = \int_{-\infty}^x K(y) dy. \tag{37}$$

(For the purposes stated in the next subsection, note that the function I is continuous while inequality (12) implies that it also fulfills the Lipschitz condition with the constant

$$L_f = K(0).) \tag{38}$$

Thus, due to the basic properties of integrals, the kernel estimator of the distribution function can be finally expressed as

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^m I\left(\frac{x-x_i}{h}\right). \tag{39}$$

If the condition

$$\lim_{m \rightarrow \infty} h = 0 \tag{40}$$

is fulfilled, then kernel estimator (39) is strongly consistent at the points of its continuity. The proof

of the above fact, under very mild assumptions, is found in Appendix 2.

In the case where the smoothing parameter is modified, definition (39) takes on the form

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^m I\left(\frac{x-x_i}{hs_i}\right). \quad (41)$$

3. Kernel Estimators of the Quantile

To continue with the considerations given in the previous subsection: if

$$K(x) > 0 \text{ for every } x \in \mathbb{R}, \quad (42)$$

then the estimator of the distribution function \hat{F} given by Eq. (39) is a strictly increasing mapping, which – together with its continuity – indicates that the kernel estimator of the quantile of order r , denoted hereinafter as \hat{q} , may be uniquely defined by the equation

$$\hat{F}(\hat{q}) = r; \quad (43)$$

therefore, finally,

$$\sum_{i=1}^m I\left(\frac{\hat{q}-x_i}{h}\right) = mr. \quad (44)$$

If the quantile of order r is uniquely defined and the condition

$$\lim_{m \rightarrow \infty} h = 0 \quad (45)$$

is fulfilled, then the above designed kernel estimator of the quantile is strongly consistent. The proof of this fact, under very mild assumptions, is presented in Appendix 3.

In practice, the value of the quantile estimator given by Eq. (44) can be calculated recurrently as the limit of the sequence $\{\hat{q}^k\}_{k=1}^{\infty}$ defined by the formulas

$$\hat{q}^1 = \frac{1}{m} \sum_{i=1}^m x_i \quad (46)$$

$$\hat{q}^{k+1} = \hat{q}^k + c \left[mr - \sum_{i=1}^m I\left(\frac{\hat{q}^k - x_i}{h}\right) \right] \text{ for } k=1, 2, \dots, \quad (47)$$

where $c \in \mathbb{R}$; however, global convergence is guaranteed by the condition

$$0 < c < \frac{2h}{mL_I} \quad (48)$$

while the value L_I can be obtained using Eq. (38). Moreover, if the function $g: \mathbb{R} \rightarrow \mathbb{R}$ defined as

$$g(x) = \sum_{i=1}^m I\left(\frac{x-x_i}{h}\right) \quad (49)$$

is not linear with the slope $\frac{mL_I}{h}$ in the neighborhood of the quantile, then the above algorithm is also convergent when

$$c = \frac{2h}{mL_I}. \quad (50)$$

For the majority of cases occurring in practice, this value yields the best results for algorithms (46) and (47). In particular, if the exponential kernel (23) is applied, then, in view of Eqs. (38) and (23), one obtains

$$c_e = \frac{8h}{m}. \quad (51)$$

In the case when the smoothing parameter is modified, which can often be advantageous in estimating the quantile, Eq. (44) takes on the form

$$\sum_{i=1}^m I\left(\frac{\hat{q}-x_i}{hs_i}\right) = mr. \quad (52)$$

Then, formulas (47)-(51) need to be changed, such that the constant h is replaced by hs_i in Eqs. (47) and (49), but by $h \sum_{i=1}^m s_i$ in dependences (48), (50), and (51).

IV. Final Comments and Conclusions

Formulas (7), (46), (47), and (24), (51), in conjunction with the methods for fixing the smoothing parameter described in the second part of Subsection III.1 (especially Eqs. (19) and (26) when the procedure of modification (32)-(35) is used), provide a complete set of rules defining the algorithm used to calculate the Bayes estimator for the loss function given by Eq. (3). The achieved form renders this algorithm exceptionally convenient for application using computer systems, and also for working in the real-time regime. According to the principles of the Bayes approach, the elaborated procedure possibilities for taking into account the consequences of estimation errors differing in sign and size.

The previously proposed form of kernel (23) satisfies all the requirements assumed in the course of defining the method of quantile estimation: its primitive has a form that is convenient for the calculational procedures, strictly increasing in its entire domain, and not linear in any restriction of the domain. It should be emphasized, however, that there is considerable latitude in using other forms of the kernel if this should prove to be advantageous for a given application.

Correctness of the algorithm presented here has

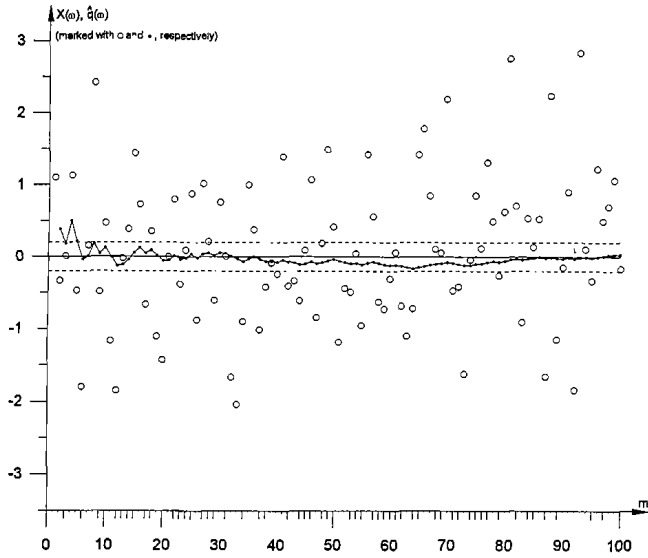


Fig. 1. Results of numerical simulation for $p_1=1$ and $p_2=1$.

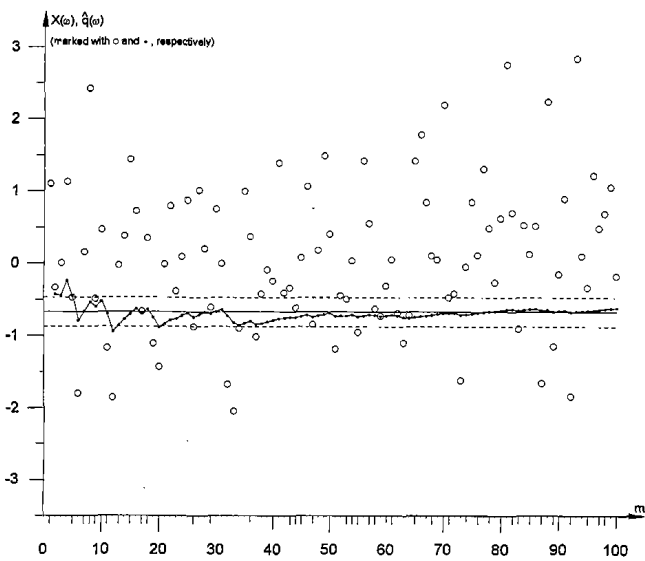


Fig. 2. Results of numerical simulation for $p_1=1$ and $p_2=3$.

been verified using a numerical simulation. The value of the Bayes estimator was calculated with a precision to 0.2 of the standard deviation for a minimal sample size of 10-50, or – in favorable cases – even beginning with 4. The precision increased significantly when the value of the estimator was located in the areas of greater density of obtained sample values.

Suppose, for example, that the estimated parameter has a normal standard distribution. For $p_1=1$ and $p_2=1$ ($r=0.5$), the value of the Bayes estimator is 0. The results obtained for this case are shown in Fig. 1. A precision of 0.2 is now obtained for the sample size $m=6$. If, in turn, $p_1=1$ and $p_2=3$ ($r=0.25$), then the value of the estimator comes to -0.67 whereas the above

precision is achieved for $m=21$ (Fig. 2). Finally, Fig. 3 illustrates the results for $p_1=1$ and $p_2=6$ ($r=0.14$), where the value of the estimator equals -1.07 while the minimal sample size has increased to $m=40$. (It is obvious that, in order to assure proper precision of estimation, a sufficient number of sample values located on either side of the quantile – i.e., greater and lesser – is necessary. As order r approaches 0 or 1, there is an increase in the random sample size essential from the point of view of precision requirements.)

The presented algorithm has also been successfully applied to the positional time-optimal control system described in Kulczycki (1996a, 1996b). This task consists in bringing the object state to the target set in a minimal and finite time. In the event that the estimator of the values of resistances to motion is understated, sliding trajectories appear in the controlled system, increasing the time needed to reach the target proportionally to the magnitude of the underestimation. If, however, this estimator is overstated, over-regulations occur in the system, with a much greater impact on the increase in the time to reach the target (likewise proportionally to the value of the overestimation), in the extreme case threatening failure of the device. The estimator of the values of resistances to motion was calculated by using the above elaborated procedure for $p_1=1$ and $p_2=6$; in such case, sliding trajectories clearly dominated in the controlled system. The speed of the method presented in this paper allowed for frequent adaptation of the control algorithm, consisting in successively updating the value of the calculated Bayes estimator of resistances to motion, following the changing work conditions of the controlled positional system.

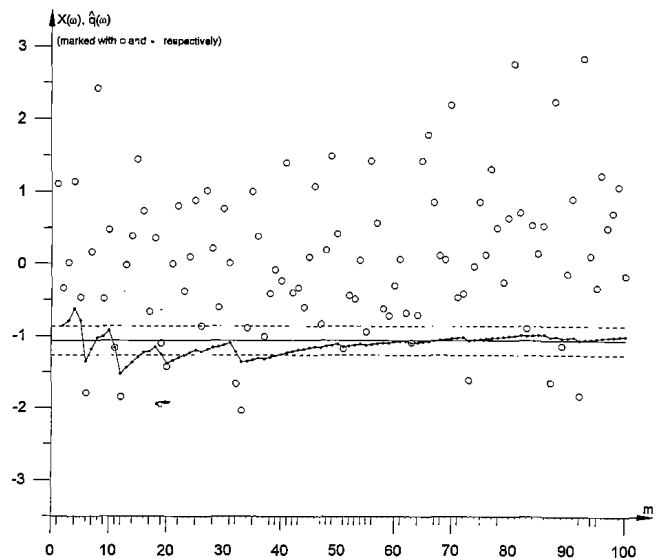


Fig. 3. Results of numerical simulation for $p_1=1$ and $p_2=6$.

Appendix 1

In the following, it will be shown that for loss function (3), the quantile of order r given by Eq. (7), if unique, constitutes the Bayes estimator.

By inserting Eq. (3) into Eq. (1), one obtains

$$\begin{aligned} I_b(\hat{x}) &= (p_1 + p_2) \int_{\Omega} (X(\omega) - \hat{x}) [r \chi_{(\hat{x}, \infty)}(X(\omega)) \\ &\quad - (1-r) \chi_{(-\infty, \hat{x})}(X(\omega))] dP(\omega) \\ &= (p_1 + p_2) \int_{\Omega} (X(\omega) - \hat{x}) [r - \chi_{(-\infty, \hat{x})}(X(\omega))] dP(\omega), \end{aligned} \quad (53)$$

where χ_A denotes the characteristic function of the set A . Also for the unique quantile q , the following is true:

$$I_b(q) = (p_1 + p_2) \int_{\Omega} (X(\omega) - q) [r - \chi_{(-\infty, q]}(X(\omega))] dP(\omega). \quad (54)$$

The combination of the above formulas yields

$$\begin{aligned} I_b(\hat{x}) - I_b(q) &= (p_1 + p_2) \int_{\Omega} |X(\omega) - \hat{x}| [\chi_{(\hat{x}, \infty)}(X(\omega)) - \chi_{(-\infty, \hat{x})}(X(\omega))] \\ &\quad \cdot [\chi_{(-\infty, q]}(X(\omega)) - \chi_{(-\infty, \hat{x})}(X(\omega))] dP(\omega) \\ &= (p_1 + p_2) \int_{\Omega} |X(\omega) - \hat{x}| [\chi_{(\hat{x}, \infty)}(X(\omega)) \chi_{(-\infty, q]}(X(\omega)) + \chi_{(-\infty, \hat{x})}(X(\omega)) \\ &\quad \cdot (\chi_{(-\infty, \hat{x})}(X(\omega)) - \chi_{(-\infty, q]}(X(\omega)))] dP(\omega) \\ &= (p_1 + p_2) \int_{\Omega} |X(\omega) - \hat{x}| [\chi_{(\hat{x}, q]}(X(\omega)) \\ &\quad + \chi_{(-\infty, \hat{x})}(X(\omega)) (1 - \chi_{(-\infty, q]}(X(\omega)))] dP(\omega) \\ &= (p_1 + p_2) \int_{\Omega} |X(\omega) - \hat{x}| [\chi_{(\hat{x}, q]}(X(\omega)) \\ &\quad + \chi_{(-\infty, \hat{x})}(X(\omega)) \chi_{(q, \infty)}(X(\omega))] dP(\omega), \end{aligned} \quad (55)$$

finally leading to

$$\begin{aligned} I_b(\hat{x}) - I_b(q) &= (p_1 + p_2) \int_{\Omega} |X(\omega) - \hat{x}| [\chi_{(\hat{x}, q]}(X(\omega)) \\ &\quad + \chi_{(q, \hat{x})}(X(\omega))] dP(\omega) \geq 0. \end{aligned} \quad (56)$$

Inequality (56) proves that the Bayes loss function assumes the global minimum when $\hat{x} = q$.

Finally, if the quantile of order r given by Eq. (7) is unique, then it constitutes the desired Bayes estimator for the loss function defined by Eq. (3).

Appendix 2

In this appendix, the strong consistency of the kernel estimator of the distribution function defined by Eq. (39) will be shown. For this purpose, consider the sequence of random variables $\{X_i\}_{i=1}^{\infty}$

defined on the common probability space (Ω, Σ, P) , as well as the corresponding sequence of its realizations $\{x_i\}_{i=1}^{\infty}$. For an arbitrarily fixed $m \in \mathbb{N} \setminus \{0\}$, the mapping $\mathcal{P}_m: \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ given by the formula

$$\mathcal{P}_m(B) = \frac{1}{m} \# \{i \in \{1, 2, \dots, m\} : x_i \in B\}, \quad (57)$$

where $\#A$ denotes the power of the set A and $\mathcal{B}(\mathbb{R})$ represents the family of real Borelian sets, is known as the empirical distribution of the sequence $\{X_i\}_{i=1}^{\infty}$. Also let $\mathcal{P}_\infty: \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ be the distribution of a probability measure. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is called the empirically ergodic sequence with the limit \mathcal{P}_∞ , if the condition

$$\lim_{m \rightarrow \infty} \mathcal{P}_m(E) = \mathcal{P}_\infty(E) \quad (58)$$

is fulfilled with probability 1 (with respect to the measure P) for every set E of the form $(-\infty, e]$, where $\mathcal{P}_\infty(\{e\}) = 0$.

As results from the Glivenko-Cantelli Theorem (Billingsley, 1979), this condition is more general than the assumption frequently formulated in the theory of estimation concerning the identity of the distributions and the independence of the random variables X_i representing the random sample. In the case where such an assumption is accepted, the measure \mathcal{P}_∞ is nothing other than the distribution of the variables X_i , i.e.,

$$\mathcal{P}_\infty(B) = P(x_i \in B) \quad (59)$$

for any $i = 1, 2, \dots$ and $B \in \mathcal{B}(\mathbb{R})$.

Lemma 1. Let the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$, defined on the common probability space (Ω, Σ, P) , be empirically ergodic with the limit \mathcal{P}_∞ , which has the distribution function F . If the estimator of this function \hat{F} is given by Eq. (39), and dependencies (9), (10), and (37) are fulfilled, then for every $x^* \in \mathbb{R}$ such that

$$\mathcal{P}_\infty(\{x^*\}) = 0 \quad (60)$$

with probability 1 (with respect to the measure P) the following equality is true:

$$\lim_{h \rightarrow 0} \lim_{m \rightarrow \infty} \hat{F}(x^*) = F(x^*). \quad (61)$$

Proof. From Eq. (57), it can be directly obtained that

$$\frac{1}{m} \sum_{i=1}^m \chi_B(x_i) = \int_{\mathbb{R}} \chi_B(x) d\mathcal{P}_m(x) \text{ for any } B \in \mathcal{B}(\mathbb{R}), \quad (62)$$

where χ_B denotes the characteristic function of the set B . Since the linear and continuous operators equal in dense spaces are identical, for any measurable function $g: \mathbb{R} \rightarrow \mathbb{R}$ the above equality yields

$$\frac{1}{m} \sum_{i=1}^m g(x_i) = \int_{\mathbb{R}} g(x) d\mathcal{P}_m(x). \quad (63)$$

In particular, on the basis of Eq. (39), the formula

$$\hat{F}(x^*) = \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_m(x) \quad (64)$$

can be obtained. Therefore, Eq. (58) implies, due to the properties of weak convergence of the distribution functions (Billingsley, 1979), that with probability 1

$$\lim_{m \rightarrow \infty} \hat{F}(x^*) = \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x). \quad (65)$$

The consequences of Eqs. (10) and (37) are:

$$\lim_{x \rightarrow -\infty} I(x) = 0 \quad (66)$$

and

$$\lim_{x \rightarrow \infty} I(x) = 1, \quad (67)$$

which, thanks to condition (9), gives

$$\lim_{h \rightarrow 0} I\left(\frac{x^* - x}{h}\right) = \begin{cases} 1 & \text{for } x < x^* \\ I(0) & \text{for } x = x^* \\ 0 & \text{for } x > x^*. \end{cases} \quad (68)$$

In turn, the following equality is true:

$$\begin{aligned} & \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x) \\ &= \int_{(-\infty, x^*)} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x) + I(0) \mathcal{P}_-({x^*}) + \int_{(x^*, \infty)} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x). \end{aligned} \quad (69)$$

Therefore, from the Lebesgue Dominated Convergence Theorem, it results that

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x) = \int_{(-\infty, x^*)} d\mathcal{P}_-(x) + I(0) \mathcal{P}_-({x^*}), \quad (70)$$

i.e., taking into account assumption (60):

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x) = \int_{(-\infty, x^*)} d\mathcal{P}_-(x). \quad (71)$$

Applying Eq. (65) to the above formula, one ultimately obtains the thesis of Lemma 1. \square

Theorem 1. Let the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$, defined on the common probability space (Ω, Σ, P) , be empirically ergodic with the limit \mathcal{P}_- , which has the distribution function F . If the estimator of this function \hat{F} is given by Eq. (39), and dependencies (9), (10), and (37), as well as the condition

$$\lim_{m \rightarrow \infty} h = 0 \quad (72)$$

are fulfilled, then for every $x^* \in \mathbb{R}$ such that

$$\mathcal{P}_-({x^*}) = 0, \quad (73)$$

with probability 1 (with respect to the measure P) the equality

$$\lim_{m \rightarrow \infty} \hat{F}(x^*) = F(x^*) \quad (74)$$

is true, which means the strong consistency and, therefore, also the consistency, of the kernel estimator of the distribution function at the points of its continuity.

Proof. It suffices to demonstrate that the convergence when $m \rightarrow \infty$ occurring in Eq. (61) is uniform with respect to the variable h .

Let F_m denote the distribution function of the measure \mathcal{P}_m . For an arbitrarily fixed $m \in \mathbb{N} \setminus \{0\}$, it is obvious that

$$\lim_{x \rightarrow \infty} I\left(\frac{x^* - x}{h}\right) (F_m - F)(x) = 0 \quad (75)$$

$$\lim_{x \rightarrow -\infty} I\left(\frac{x^* - x}{h}\right) (F_m - F)(x) = 0. \quad (76)$$

Applying to the Stieltjes integral \int the integration by parts procedure, one obtains

$$\begin{aligned} & \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_m(x) - \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_-(x) \\ &= \int_{\mathbb{R}} I\left(\frac{x^* - x}{h}\right) d(F_m - F)(x) = - \int_{\mathbb{R}} (F_m - F)(x) dI\left(\frac{x^* - x}{h}\right). \end{aligned} \quad (77)$$

Since, regardless of the value of the variable h , the saltus of the function I equals 1 (is finite) whereas from the Glivenko-Cantelli Theorem (Billingsley, 1979) it results that

$$\sup_{x \in \mathbb{R}} |(F_m - F)(x)| \xrightarrow{m \rightarrow \infty} 0, \quad (78)$$

Eqs. (64), (65), and (77) finally prove Theorem 1. \square

Note that if the distribution of the measure \mathcal{P}_- has a density function, then assumptions (60) and (73) are obviously fulfilled.

Appendix 3

The strong consistency of the kernel estimator of the quantile defined by Eq. (44) will be shown below. (The notion of the empirically ergodic sequence is found in Appendix 2.)

Lemma 2. Let the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$, defined on the common probability space (Ω, Σ, P) , be empirically ergodic with the limit \mathcal{P}_- . If the quantile of order r is defined uniquely with respect to the measure \mathcal{P}_- , its estimator is defined by Eq. (44), and dependencies (9), (10), (37), and (42) are fulfilled, then with probability 1 (with respect to the measure P) the following equality is true:

$$\lim_{h \rightarrow 0} \lim_{m \rightarrow \infty} \hat{q} = q. \quad (79)$$

Proof. In order to demonstrate Eq. (79), it is sufficient to show that

$$\forall \varepsilon > 0 \exists h_* > 0 : \forall h < h_* \exists m_* \in \mathbb{N} \setminus \{0\} : \forall m > m_* |\hat{q} - q| < \varepsilon. \quad (80)$$

Let any $\varepsilon > 0$ be fixed. Since the measure \mathcal{P}_- is finite, the set of real numbers of positive measure can be at most countable. Thus, there exist $x^-, x^+ \in \mathbb{R}$ of zero measure \mathcal{P}_- and fulfilling inequalities

$$q - \varepsilon < x^- < q < x^+ < q + \varepsilon. \quad (81)$$

The distribution function is a non-decreasing mapping; therefore, due to the assumed uniqueness of the quantile, it can be inferred that there exists $\delta > 0$ such that

Bayes Parameter Identification

$$F(x^-) + \delta < F(q) < F(x^+) - \delta, \quad (82)$$

where F denotes the distribution function of the measure \mathcal{P}_n . Lemma 1 states that

$$\begin{aligned} \forall \varepsilon > 0 \exists h_* > 0 : \forall h < h_* \exists m_* \in \mathbb{N} \setminus \{0\} : \forall m > m_* \\ \hat{F}(x^-) < F(x^-) + \delta \\ \hat{F}(x^+) > F(x^+) - \delta; \end{aligned} \quad (83)$$

therefore, by combining the last two dependencies, one obtains

$$\begin{aligned} \forall \varepsilon > 0 \exists h_* > 0 : \forall h < h_* \exists m_* \in \mathbb{N} \setminus \{0\} : \forall m > m_* \\ \hat{F}(x^-) < F(q) < \hat{F}(x^+). \end{aligned} \quad (84)$$

Thus, if the quantile estimator \hat{q} is calculated in accordance with Eq. (44), then since the monotonicity of the function \hat{F} and due to dependence (81), Eq. (84) implies the truth of formula (80), which concludes Lemma 2. \square

Theorem 2. Let the sequence of real random variables $\{X_i\}_{i=1}^\infty$, defined on the common probability space (Ω, Σ, P) , be empirically ergodic with the limit \mathcal{P}_n . If the quantile of order r is defined uniquely with respect to the measure \mathcal{P}_n , its estimator is defined by Eq. (44), and dependencies (9), (10), (37), and (42), as well as the condition

$$\lim_{m \rightarrow \infty} h = 0 \quad (85)$$

are fulfilled, then with probability 1 (with respect to the measure P) the equality

$$\lim_{m \rightarrow \infty} \hat{q} = q \quad (86)$$

is true, which means the strong consistency and, therefore, also the consistency, of the kernel estimator of the quantile.

Proof. As results from the proof of Theorem 1, the value m_n introduced by formula (83) does not depend on the variable h . This means, thanks to Lemma 2, that the convergence when $m \rightarrow \infty$ is uniform, which finally proves Theorem 2. \square

Note that if the distribution of the measure \mathcal{P}_n has a density function with a connected support, then its quantile is uniquely defined.

It should also be emphasized that condition (29) required in the estimation of the density function in order to assure the consistency property, is not necessary in the cases of the distribution function and the quantile.

References

- Billingsley, P. (1979) *Probability and Measure*. Wiley, New York, NY, U.S.A.
- Kulczycki, P. (1996a) Almost certain time-optimal positional control. *IMA Journal of Mathematical Control & Information*, **13**, 63-77.
- Kulczycki, P. (1996b) Time-optimal positional feedback controller for random systems. *Proceedings of the National Science Council ROC(A)*, **20**, 79-89.
- Lehmann, E. L. (1983) *Theory of Point Estimation*. Wiley, New York, NY, U.S.A.
- Parzen, E. (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **74**, 105-131.
- Prakasa Rao, B. L. S. (1983) *Nonparametric Functional Estimation*. Academic Press, Orlando, FL, U.S.A.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, U.K.

利用貝氏法及核心法從事參數辨別

PIOTR KULCZYCKI

*Faculty of Electrical Engineering
Cracow University of Technology
Cracow, Poland*

摘 要

本論文利用貝氏法來從事參數辨別〔點估計〕的工作，其計算過程則運用所謂核心估測技巧。根據上述之考量，我們提出一個完整的參數估測演算法則。完整的數值運算方法，可運用於即時系統的數值運算工作。