Piotr Kulczycki
Associate Professor,
Systems Research Institute,
Polish Academy of Sciences,
ul. Newelska 6,
PL-01-447 Warsaw, Poland
e-mail: kulczycki@ibspan.waw.pl

An Algorithm for Bayes Parameter Identification

This paper deals with the task of parameter identification using the Bayes estimation method, which makes it possible to take into account the differing consequences of positive and negative estimation errors. The calculation procedures are based on the kernel estimators technique. The final result constitutes a complete algorithm usable for obtaining the value of the Bayes estimator on the basis of an experimentally obtained random sample. An elaborated method is provided for numerical computations.
[DOI: 10.1115/1.1409552]

1 Introduction

One of the elementary issues of contemporary engineering is parameter identification, i.e., the specification of the value of a parameter. In the case typical for engineering applications, its realizations are directly measurable (observable). In that situation, one has-not knowing the "true" value of the parameter x-its m measurements $x_1, x_2, \ldots, x_m$, obtained by using independent experiments, and in practice burdened with errors of varying origin. On the basis of these measurements, that number $\hat{x}$, which would most nearly approximate the true (but unknown) value of the parameter x must be determined. If such measurements can be treated as the sum of the true value and the random disturbances, then the task from the mathematical point of view becomes a typical problem for point estimation, while $\hat{x}$ is called the estimator of the parameter x [1].

The procedures generally used for specifying the estimator, such as the least squares or maximum likelihood methods, are noted for their great simplicity and general availability in the literature. However, they do not make it possible to take into account the differing consequences of positive and negative estimation errors. Yet, in engineering practice, it often turns out that one of the two has only a minor impact on the quality of work of the device, while the other has a far more profound influence, not excluding system failure.

The Bayes estimation method [1; Section 4.1] used in this paper to solve the problem of parameter identification has no such shortcomings. The calculational procedures worked out below will be based on the kernel estimators technique [2]. The final result is a complete usable algorithm for specifying the value of the parameter estimator, which in a natural way makes it possible to take into account the consequences of estimation errors differing in size and sign.

2 Bayes Estimation

Assume the probability space (Ω,Σ,P), where Ω denotes the set of elementary events, Σ means its σ-algebra, and P is a probability measure [3,4]. Suppose that the real random variable $X:\Omega\rightarrow\mathrm{R}$ represents the measurement process, and its realizations are interpreted as the particular independent measurements of the value of the estimated parameter x. Consider also the loss function $l:\mathrm{R}\times\mathrm{R}\rightarrow\mathrm{R}\cup\{\pm\infty\}$; its values $l(\hat{x},x)$ denote the losses which may be incurred by assuming $\hat{x}$ as the estimator, whereas the true (but unknown) value of the estimated parameter is x. Let $l_b:\mathrm{R}\rightarrow\mathrm{R}\cup\{\pm\infty\}$ be a function of the so-called Bayes losses

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

where $\int_{\Omega}\cdot dP(\omega)$ denotes the integral with respect to the probability measure P. Therefore, $l_b(\hat{x})$ constitutes the expectation value of losses if the value $\hat{x}$ is assumed. Every element $\hat{x}_b\in\mathrm{R}$ such that

$$l_b(\hat{x}_b)=\inf_{\hat{x}\in\mathrm{R}} l_b(\hat{x}) \tag{2}$$

is known as a Bayes estimator. For details see Section 4.1 of [1].

In the present paper, consideration will be given to an asymmetrical form of the loss function

$$l(\hat{x},x)=\begin{cases}-p_1(\hat{x}-x) & \text{if } \hat{x}-x\leqslant 0\\ p_2(\hat{x}-x) & \text{if } \hat{x}-x\geqslant 0\end{cases}, \tag{3}$$

where $p_1,p_2>0$. The constants p_1 and p_2 constitute the coefficients of proportionality of losses suffered after obtaining a value of the estimator that is either smaller or greater than the true value of the estimated parameter, i.e., for negative and positive estimation errors, respectively. With the values p_1 and p_2 given, it is possible to calculate easily the quantity r such that

$$r=\frac{p_1}{p_1+p_2}=\frac{\frac{p_1}{p_2}}{\frac{p_1}{p_2}+1}. \tag{4}$$

It is not difficult to show (see e.g., [5]) that if the quantile of order r is uniquely defined, then it constitutes the Bayes estimator for the loss function given by formula (3). (The notion of the quantile can be found e.g., in [4].)

The following will present the practical procedure for calculating the value of the quantile of order r using the kernel estimators technique, which in accordance with the above result will complete the solution of the Bayes method for the point estimation task considered here.

3 Kernel Estimators Technique

The kernel estimator of the density function of the real random variable X, calculated on the basis of m realizations $x_1, x_2, \ldots, x_m$, is defined by the dependence

$$\hat{f}(x)=\frac{1}{mh}\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right), \tag{5}$$

where the measurable and symmetrical function $K:\mathrm{R}\rightarrow[0,\infty)$ with a unique integral and a maximum in point zero is called the kernel, while the positive constant h is known as the smoothing parameter. Detailed information concerning the rules for choosing the function K and fixing the value of the parameter h is included

Contributed by the Dynamic Systems and Control Division for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received by the Dynamic Systems and Control Division December 6, 1999. Associate Editor: S. Fassois.

in [2]. Specifically, the approximate value of the optimal (in the mean squared sense) smoothing parameter can be calculated by assuming the normal distribution; one then obtains

$$h=\left(V_K\frac{8}{3}\sqrt{\pi}\frac{1}{m}\right)^{1/5}\hat{\sigma}, \tag{6}$$

while

$$V_K=\int_{-\infty}^{\infty}K(x)^2dx\cdot\left(\int_{-\infty}^{\infty}x^2K(x)dx\right)^{-2} \tag{7}$$

$$\hat{\sigma}^2=\frac{1}{m}\sum_{i=1}^{m}\left(x_i-\frac{1}{m}\sum_{i=1}^{m}x_i\right)^2=\frac{1}{m}\sum_{i=1}^{m}x_i^2-\left(\frac{1}{m}\sum_{i=1}^{m}x_i\right)^2. \tag{8}$$

On the other hand, the choice of the type of the kernel K does not have a major impact on the statistical quality of estimation. In practice, it becomes possible to take into account primarily the desired properties of the estimator obtained, e.g., the simplicity of calculation or the finiteness of the support, etc.

In many applications, it proves to be particularly advantageous to introduce the concept of modification of the smoothing parameter. The estimator can then be constructed in the following manner:

(A) the kernel estimator $\hat{f}$ is calculated in accordance with basic dependence (5);

(B) the modifying parameters $s_i>0$ $(i=1,2,\ldots,m)$ are stated as

$$s_i=\left(\frac{\hat{f}(x_i)}{b}\right)^{-1/2}, \tag{9}$$

where b denotes the geometric mean of the numbers $\hat{f}(x_1),\hat{f}(x_2),\ldots,\hat{f}(x_m)$, given in the form of the logarithmic equation

$$\log(b)=\frac{1}{m}\sum_{i=1}^{m}\log(\hat{f}(x_i)); \tag{10}$$

(C) the kernel estimator with the modified smoothing parameter is defined by the formula

$$\hat{f}(x)=\frac{1}{mh}\sum_{i=1}^{m}\frac{1}{s_i}K\left(\frac{x-x_i}{h\ s_i}\right). \tag{11}$$

One of the essential features of such an estimator consists in its slight sensitivity to the exactness of the choice of the parameter h. In practice, when the modification procedure is applied, it quite often proves sufficient to accept the approximate value given by dependence (6).

A detailed description of the above technique can be found in [2].

4 The Algorithm

To continue the considerations of the previous section, suppose the kernel K is positive and has the primitive given by

$$I(x)=\int_{-\infty}^{x}K(y)\ dy. \tag{12}$$

Then the estimator of the distribution function $\hat{F}$ with modified smoothing parameter can be described as

$$\hat{F}(x)=\frac{1}{m}\sum_{i=1}^{m}I\left(\frac{x-x_i}{hs_i}\right), \tag{13}$$

and, therefore, the estimator of quantile of order r, denoted hereinafter as $\hat{q}$, is uniquely defined by the equation

$$\frac{1}{m}\sum_{i=1}^{m}I\left(\frac{\hat{q}-x_i}{hs_i}\right)=r. \tag{14}$$

Moreover, the estimator may be calculated recurrently, using Newton's method [6], as the limit of the sequence $\{\hat{q}^k\}_{k=0}^{\infty}$ defined by the formulas

$$\hat{q}^0=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{15}$$

$$\hat{q}^{k+1}=\hat{q}^k+\frac{r-\hat{F}(\hat{q}^k)}{\hat{f}(\hat{q}^k)}\quad\text{for } k=0,1,\ldots, \tag{16}$$

since-according to formulas (11)–(13)-the function $\hat{f}$ constitutes the derivative of the mapping $\hat{F}$.

For the purposes of the method elaborated here, the kernel

$$K(x)=\frac{e^{-x}}{(1+e^{-x})^2} \tag{17}$$

can be proposed. It fulfills all the requirements formulated above, and in particular its primitive has a form convenient for calculations, namely,

$$I(x)=\frac{1}{1+e^{-x}}. \tag{18}$$

In this case, the quantity V_k occurring in dependence (6) amounts to

$$V_K=\frac{3}{2\pi^4}. \tag{19}$$

5 Final Suggestions and Conclusions

The substance of the present paper provides complete material defining the practical algorithm used to calculate the Bayes estimator for loss function (3). It is assumed that m independent measurements $x_1,x_2,\ldots,x_m$ of the unknown value of the estimated parameter x are available. Based on prior process knowledge, the user should also identify the ratio p_1/p_2 which characterizes the proportion of losses resulting from negative and positive estimation errors, i.e., underestimating or overestimating the parameter. It is then easy to successively calculate the following values:

(A) the order r of the quantile-from the second part of formula (4),

(B) the smoothing parameter h-on the basis of dependence (6), along with (8) and (19),

(C) quantities $f(x_i)$ for basic form (5), applying also equality (17),

(D) modifying parameters s_i thanks to procedure (9)–(10).

Since the forms of the functions K and I are given by dependencies (17) and (18), then all the quantities needed to apply algorithm (15)–(16) have already been defined. This is tantamount to specifying the value of the Bayes estimator.

The estimator obtained in this fashion is strongly consistent, i.e., with probability 1 it converges on the proper value along with the increase in the size of sample. The strict formulation of this fact, under very mild assumptions, is presented in the Appendix, whereas the proof can be found in [7]. It should be emphasized that the quite general condition of the uniqueness of the quantile of order r, fulfilled e.g., when the random variable X has a density function with a connected support, is in practice the only limitation on the possibility of applying the method proposed in this paper.

The correct functioning of the algorithm here designed has been verified using a numerical simulation. Random disturbances of various distributions, including asymmetrical, long-tailed, and multimodal, were subjected to testing. The results obtained for

Table 1 Results obtained for parameter zero and disturbances with standard normal distribution: (*a*) value of the Bayes estimator proposed in this work; (*b*) value of the classical sample mean (only for $p_1=1$, $p_2=1$); (*c*) precision (in relation to theoretical) of the quantile estimator proposed in this work; (*d*) precision (in relation to theoretical) of the estimator Y_8 recommended in survey paper [8].

$p_1=1$, $p_2=5$ $\left(\frac{p_1}{p_2}=0.2;\ r=0.167\right)$

m	(a)	(c)	(d)
10	-1.031	0.328	0.340
20	-1.021	0.231	0.241
50	-1.007	0.152	0.162
100	-0.993	0.108	0.115
200	-0.980	0.075	0.081
500	-0.972	0.049	0.053
1000	-0.967	0.034	0.037

$p_1=1$, $p_2=3$ $\left(\frac{p_1}{p_2}=0.333;\ r=0.25\right)$

m	(a)	(c)	(d)
10	-0.717	0.296	0.328
20	-0.707	0.213	0.229
50	-0.698	0.138	0.154
100	-0.690	0.099	0.107
200	-0.682	0.069	0.075
500	-0.677	0.045	0.048
1000	-0.674	0.031	0.034

$p_1=1$, $p_2=2$ $\left(\frac{p_1}{p_2}=0.5;\ r=0.333\right)$

m	(a)	(c)	(d)
10	-0.462	0.279	0.310
20	-0.451	0.206	0.229
50	-0.447	0.130	0.145
100	-0.442	0.093	0.105
200	-0.436	0.065	0.072
500	-0.434	0.043	0.046
1000	-0.431	0.030	0.033

$p_1=1$, $p_2=1$ $\left(\frac{p_1}{p_2}=1;\ r=0.5\right)$

m	(a)	(b)	(c)	(d)
10	-0.008	-0.005	0.268	0.293
20	-0.004	-0.003	0.200	0.222
50	-0.005	-0.007	0.125	0.138
100	-0.003	-0.003	0.089	0.099
200	-0.001	-0.001	0.064	0.070
500	-0.002	-0.002	0.041	0.044
1000	-0.001	-0.001	0.029	0.031

$p_1=2$, $p_2=1$ $\left(\frac{p_1}{p_2}=2;\ r=0.667\right)$

m	(a)	(c)	(d)
10	0.440	0.279	0.309
20	0.442	0.202	0.221
50	0.434	0.130	0.143
100	0.433	0.093	0.102
200	0.433	0.067	0.073
500	0.431	0.042	0.046
1000	0.431	0.030	0.032

$p_1=3$, $p_2=1$ $\left(\frac{p_1}{p_2}=3;\ r=0.75\right)$

m	(a)	(c)	(d)
10	0.701	0.295	0.326
20	0.697	0.212	0.228
50	0.683	0.136	0.148
100	0.681	0.097	0.106
200	0.678	0.070	0.078
500	0.674	0.045	0.048
1000	0.674	0.032	0.034

$p_1=5$, $p_2=1$ $\left(\frac{p_1}{p_2}=5;\ r=0.833\right)$

m	(a)	(c)	(d)
10	1.025	0.324	0.346
20	1.016	0.235	0.257
50	0.989	0.148	0.164
100	0.984	0.108	0.117
200	0.976	0.077	0.084
500	0.969	0.049	0.053
1000	0.968	0.035	0.037

standard normal distribution are shown in columns (a) of Table 1. For simplicity, the parameter being estimated had the value zero.

When $p_1=p_2$, i.e., given the assumption that negative and positive estimation errors entail the same losses, the Bayes estimator and the classical sample mean are conditioned analogously, which renders it possible to compare the results that are obtained by using them. Columns (*a*) and (*b*) for $p_1=1$, $p_2=1$ in Table 1 indicate that their precision was comparable. In the case of the Bayes estimator, however, it is more important that, if $p_1 \neq p_2$, then its value was properly shifted in the direction of those errors for which the parameter p_1 or p_2 was less [see columns (*a*) in Table 1, keeping in mind that the standard deviation of the random disturbances was 1].

The results obtained by using the quantile estimator presented in Section 4 were more precise in comparison with those generated by other quantile estimators available in the literature, especially with small sample sizes [e.g., compare columns (*c*) and (*d*) in Table 1 created for the estimators proposed in this work and for Y_8 recommended in survey paper [8], respectively].

In sum, for $p_1=p_2$ the proposed algorithm yields results that are comparable to those obtained using the sample mean, while assuming different p_1 and p_2, opens up possibilities that are unavailable for this classical method: to properly shift the value of the estimator in the direction associated with smaller losses. The Bayes estimation method proposed here is natural, easy to interpret and used in practice.

The algorithm presented here has also been successfully applied to the positional time-optimal control system described in papers [9,10]. This task consists in bringing the object state to the target set in a minimal time. In the event that the estimator of the values of resistances to motion is underestimated, sliding trajectories appear in the controlled system, increasing the time to reach the target proportionally to the magnitude of the underestimation. If, however, this estimator is overestimated, 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), threatening in the extreme case failure of the device. The estimator of the values of resistances to motion was calculated by using the above-elaborated procedure for $p_1/p_2=0.2$; thanks to this, more desirable sliding trajectories clearly dominated in the controlled system.

The foregoing applicational example points up an engineering interpretation of the issue, somewhat exceeding the strict mathematical point estimation formulation presented in the Introduction. The parameter under consideration may in fact be the reflection of an entire array of phenomena, reduced to a single constant due to the necessity to simplify the model. Then the issue consists not so much in approaching the "true" value (since no such thing exists), but rather to specify the best possible characterization of these phenomena using a single number. From the mathematical point of view, the formalism of statistical decision theory [11] is then appropriate, although the results obtained using the Bayes decision rule are in such case identical with those presented in this paper for the issue of point estimation.

Appendix

In the following, the strong consistency of the kernel estimator of the quantile defined by Eq. (14) will be commented. For this purpose, consider the sequence of random variables $\{X_i\}_{i=1}^{\infty}$ defined on the common probability space $(\Omega, \Sigma, \mathcal{P})$, as well as the corresponding sequence of its realizations $\{x_i\}_{i=1}^{\infty}$. For an arbitrarily fixed $m \in \mathbb{N}\backslash\{0\}$, the mapping $\mathcal{P}_m:\mathcal{B}(\mathbb{R})\to[0,1]$ given by the formula

$$\mathcal{P}_m(B)=\frac{1}{m}\#\{i\in\{1,2,\ldots,m\}:x_i\in B\}, \qquad (20)$$

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 $\{X_i\}_{i=1}^{\infty}$. Let also $\mathcal{P}_{\sim}:\mathcal{B}(\mathbb{R})\to[0,1]$ 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 $\mathcal{P}_{\sim}$, if the condition

$$\lim_{m\to\infty}\mathcal{P}_m(E)=\mathcal{P}_{\sim}(E) \qquad (21)$$

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

As results from the Glivenko-Cantelli Theorem [3], 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 that such an assumption is accepted, the measure $\mathcal{P}_{\sim}$ is nothing other than the distribution of the variables X_i, i.e.,

$$\mathcal{P}_{\sim}(B)=P(x_i\in B) \qquad (22)$$

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

Theorem. Let the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$, defined on the common probability space $(\Omega, \Sigma, \mathcal{P})$, be empirically ergodic with the limit $\mathcal{P}_{\sim}$. If the quantile of order r (denoted below as q) is defined uniquely with respect to the measure $\mathcal{P}_{\sim}$, its estimator $\hat{q}$ is given as the solution of equation (14), the kernel K is positive, and dependence (12) as well as the condition

$$\lim_{m\to\infty} h=0 \qquad (23)$$

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

$$\lim_{m \to \infty} \hat{q} = q \tag{24}$$

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

The proof of this theorem is found in reference [7]. ■

Note that if one establishes the smoothing parameter on the basis of formula (6), then criterion (23) is obviously fulfilled.

References

[1] Lehmann, E. L., 1983, *Theory of Point Estimation*, Wiley, New York.

[2] Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[3] Billingsley, P., 1979, *Probability and Measure*, Wiley, New York.

[4] Fisz, M., 1963, *Probability Theory and Mathematical Statistics*, Wiley, New York.

[5] Kulczycki, P., 2000, "Fuzzy Controller for Mechanical Systems," IEEE Trans. Fuzzy Syst., **8**, pp. 645–652.

[6] Dertouzos, M. L., Athans, M., Spann, R. N., and Mason, S. J., 1972, *Systems, Networks, and Computation: Basic Concepts*, McGraw-Hill, New York.

[7] Kulczycki, P., and Dawidowicz, A. L., 1998, "Kernel Estimator of Quantile," Universitatis Jagiellonicae Acta Mathematica, **37**, pp. 101–112.

[8] Parrish, R. S., 1990, "Comparison of Quantile Estimators in Normal Sampling," Biometrics, **46**, pp. 247–257.

[9] Kulczycki, P., 1999a, "A Random Approach to Time-Optimal Control," ASME J. Dyn. Syst., Meas., Control, **121**, pp. 542–543.

[10] Kulczycki, P., 1999b, "Random time-optimal control for mechanical systems," European Journal of Automation, **33**, pp. 115–140.

[11] Berger, J. O., 1980, *Statistical Decision Theory*, Springer-Verlag, New York.