PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 5, Fasc. 1 (1985), p. 31-44

ON INVARIANT CURVE ESTIMATORS

BY

WOLFGANG WERTZ (WIEN)

Abstract. Invariantly optimal estimators of functions, with respect to mean integrated squared error as a risk, are considered. Their Bayes properties are investigated and, for several examples, explicit solutions are computed.

1. INTRODUCTION

The purpose of this paper is to construct nonparametric estimators, satisfying optimality conditions with respect to a nonparametric risk, namely the mean integrated squared error (MISE). Actually, it is possible to carry through the computations only for parametric models. On the other hand, this enables us to compare estimators corresponding to a nonparametric criterion with estimators which are obtained by estimating the unkown parameters.

Without further restrictions on the estimators under consideration it is impossible to find estimators which are optimal in the sense that they minimize uniformly the risk.

The assumption of expectation-unbiasedness is too strong for nonparametric situations, in larger models even void. Hence we restrict ourselves to the study of invariant estimators.

In [10] we have derived invariantly optimal estimators (IOE's) of functions, based on mean integrated p-th power loss ($p > 1$); the case $p = 1$ with several examples is treated in [11]. For $p = 2$, in [10] further examples are included, concerning density estimation on the real line, on the circle and on the sphere; moreover, estimation of the density, of the density of order statistics, etc. is considered there. The IOE of the reliability function in the exponential case is investigated in [3], including its asymptotic deficiency compared with other common types of estimators, in a local manner, however.

Restriction to invariant estimators can be justified for several reasons in our situation:

(i) For many applications, invariance is a desirable property of estimators.

(ii) Under reasonable conditions, estimators constructed by substituting estimators of the parameter turn out to be invariant.

To be more specific, we give the following example: let us estimate the density, which is assumed to be specified up to location and scale, say

$$x \mapsto \frac{1}{\sigma} f\left(\frac{x-\xi}{\sigma}\right).$$

Estimating ξ by the empirical mean $\bar{x}$ and σ by the empirical standard deviation s yields

$$x \mapsto \frac{1}{s} f\left(\frac{x-\bar{x}}{s}\right)$$

as a density estimate which is invariant in the sense below.

(iii) Although IOE's are not strict in general (a situation which is usually met in nonparametric curve estimation), there are examples in which the IOE has uniformly smaller risk than the usually strict "parametric estimator" (see e.g. [8], [2]). It is not yet known whether this is the typical situation. In some cases an expectation-unbiased estimator exists, but it is inferior to the IOE. All of this holds, e.g., if a normal density with unknown location is to be estimated.

(iv) If the model has certain invariance properties, for every estimator $\hat{f}$ there exists an invariant estimator $\hat{f}_0$ such that the maximum risk of $\hat{f}_0$ is not greater than the maximum risk of $\hat{f}$. In particular, the existence of an invariant minimax estimator can be proved (see [4] and [9]).

In the following we give some more examples for the case $p = 2$. For many problems (in particular estimating distribution functions, reliability functions, hazard rate functions and so on), an essential integrability assumption of [10] (the square integrability of the function Df_g below, in the case treated here) fails to hold. A natural way to avoid such difficulties in case of sample space $\mathscr{R}$ is to truncate the function to be estimated at two quantiles; this is carried out in [4]. From a practical point of view, the truncation at quantiles is reasonable under many circumstances, say, in a purely nonparametric framework, when nothing is assumed about the analytic form of the density (e.g. exponentially or algebraically decreasing tails), no reliable information is available about the behaviour of the distribution outside an interval between two appropriate quantiles.

In chapter 3 the IOE is shown to be a generalized Bayes estimator (GBE) in the sense of Sacks [5].

2. PRELIMINARIES

Let G be a locally compact group, H a closed subgroup, $X := G/H$ the corresponding homogeneous space, $\mathscr{X}$ its σ-algebra of Borel sets, ν a right Haar measure on G and μ a nontrivial, relatively invariant, σ-finite measure on $\mathscr{X}$, i.e.

$$\bigwedge_{g\in G} \bigwedge_{A\in\mathscr{X}} \mu(gA) = \chi(g)\,\mu(A)$$

with a function $\chi\colon G \to (0, \infty)$ independent of A.

Let f be a probability density, $P := f\mu$ the corresponding probability measure, f_g the probability density $x \mapsto \chi(g^{-1}) f(g^{-1}x)$ and $P_g := f_g\,\mu$. For every $g \in G$ let $Df_g\colon X \to \mathscr{R}$ be a μ-square-integrable function with

$$Df_g(x) = \varphi(g^{-1})\, Df(g^{-1}x) \tag{1}$$

for every $g \in G$ and $x \in \mathscr{X}$ with a function $\varphi\colon G \to \mathscr{R}\setminus\{0\}$.

The set of all *φ-invariant estimators* $\hat{f}$ of Df_g, i.e.

$$\hat{f} \in \bigcap_{g\in G} L_2(X^{n+1}, \mathscr{X}^{n+1}, P_g^n \otimes \mu),$$

with

$$\bigwedge_{x\in X} \bigwedge_{x^n\in X^n} \bigwedge_{g\in G} \hat{f}(x^n, x) = \varphi(g)\, \hat{f}(gx^n, gx),$$

will be denoted by $S_2(\varphi)$.

Let

$$R(\hat{f}, g) = R(\hat{f}, P_g) := \int_{X^n} \int_X [\hat{f}(x^n, x) - Df_g(x)]^2\, d\mu(x)\, dP_g^n(x^n)$$

serve as the risk of $\hat{f}$ at g.

The formula

$$R(\hat{f}, g) = \chi(g)\, \varphi^2(g^{-1})\, R(\hat{f}, e)$$

for every $\hat{f} \in S_2(\varphi)$, where e is the unit of the group, is proved in [10], and further the following is stated:

THEOREM. *Let*

$$g \mapsto \varphi^2(g)\, \chi^{n+1}(g^{-1}) \prod_{i=1}^{n} f(g^{-1}x_i)$$

be ν-integrable for P^n-almost all $x^n = (x_1, \ldots, x_n)$. Then

$$\text{(2)} \qquad \hat{f}_0(x^n, x) = \frac{\int\limits_G Df(g^{-1}x)\,\varphi(g)\,\chi^{n+1}(g^{-1}) \prod\limits_{i=1}^{n} f(g^{-1}x_i)\,d\nu(g)}{\int\limits_G \varphi^2(g)\,\chi^{n+1}(g^{-1}) \prod\limits_{i=1}^{n} f(g^{-1}x_i)\,d\nu(g)}$$

defines the φ-invariantly optimal estimator of Df_g.

3. THE IOE AS A GENERALIZED BAYES ESTIMATOR

The estimation problem treated here can also be regarded as a purely parametric problem, namely estimating $Df_g(x)$ for fixed $x \in X$. Just as it is interesting to consider parametric methods from a nonparametric point of view, it is informative to regard (2) as a pointwise estimator of $Df_g(x)$ and to apply parametric criteria.

Let β be a σ-finite measure on G with Radon-Nikodym density p with respect to ν and $g \mapsto d(g)$ a given function from G to $\mathscr{R}$ which is to be estimated. The (generalized) Bayes risk $B(\hat{d})$ of an estimator

$$\hat{d} \in \bigcap_{g \in G} L_2(X^n, P_g^n)$$

of $d(g)$ equals

$$B(\hat{d}) = \int\limits_G \int\limits_{X^n} [\hat{d}(x^n) - d(g)]^2 f(x^n|g)\,d\mu^n(x^n)\,d\nu(g),$$

where

$$f(x^n|g) := \chi^n(g^{-1}) \prod_{i=1}^{n} f(g^{-1}x_i).$$

We make the following assumptions:

(3a) $f(x^n) := \int\limits_G f(x^n|g)\,p(g)\,d\nu(g)$ fulfils $0 < f(x^n) < \infty$ for P^n-almost all x^n;

(3b) $g \mapsto d(g) f(x^n|g)\,p(g)$ is ν-integrable for P^n-almost all x^n;

(3c) $d \in L_2(G, \beta)$.

Then $q(g|x^n) := f(x^n|g)\,p(g)/f(x^n)$ is defined for P^n-almost every x^n, and $g \mapsto q(g|x^n)$ is a probability density.

The mapping

$$h \mapsto T(h) = \Big(x^n \mapsto \int\limits_G h(x^n, g) \cdot q(g|x^n)\,d\nu(g)\Big),$$

defined on $L_2\big(X^n \times G, f(x^n|g)\,p(g)\,d\mu^n \otimes \nu(x^n, g)\big)$, is easily seen to be a projection to the subspace of functions depending only on the variable x^n. Hence $T(d)$ minimizes B; $T(d)$ has the form

$$T(d)(x^n) = \int d(g)\, q(g|x^n)\, d\nu(g).$$

Since β is σ-finite, there is an ascending sequence (C_k) of sets with $0 < \beta(C_k) < \infty$, $C_k \uparrow G$. A sequence of probability measures β_k can be defined e.g. by

$$d\beta_k(g) := \mathbb{1}_{C_k}(g)\, p(g)\, d\nu(g)/\beta(C_k).$$

Let $\hat{d}_k$ be the Bayes estimator of d with a priori distribution β_k; then, by the Lebesgue theorem,

$$\hat{d}_k(x^n) = \frac{\dfrac{1}{\beta(C_k)} \displaystyle\int_G d(g)\, f(x^n|g)\, \mathbb{1}_{C_k}(g)\, p(g)\, d\nu(g)}{\dfrac{1}{\beta(C_k)} \displaystyle\int_G f(x^n|g)\, \mathbb{1}_{C_k}(g)\, p(g)\, d\nu(g)}$$

converges to $T(d)(x^n)$ for almost all x^n. Hence $T(d)$ is a generalized Bayes estimator (see [5], and cf. also [6]).

Application of these considerations to the case $d(g) = \varphi(g^{-1})\, Df(g^{-1}x)$, $p(g) = \varphi^2(g)\, \chi(g^{-1})$ proves (2) to be a generalized Bayes estimator.

Using a result by Strasser [7], generalizing a theorem of De Groot and Rao [1], formula (2) can also be proved to be a generalized Bayes estimator in the nonparametric sense, under mild restrictions.

Let λ be a probability measure on G, and let $g \mapsto \int (Df_g)^2\, d\mu$ be λ-integrable, further $Df_g \geq 0$. In an obvious manner, λ induces an a priori distribution on $\{P_g\colon g \in G\}$. We assume that $L_2(X, \mu)$ has a countable base and

(4) $$(x, g) \mapsto \varphi(g^{-1})\, Df(g^{-1}x)$$

is $\mu \otimes \lambda$-integrable.

We take $\Delta = L_2(X, \mu)$ as a decision space and $L(t, P_g) = \int (t - Df_g)^2\, d\mu$ as a loss function. Then

$$B(\hat{f}) = \int_G \int_{X^n} \int_X [\hat{f}(x^n, x) - Df_g(x)]^2\, d\mu(x) \prod_{i=1}^{n} f_g(x_i)\, d\mu^n(x^n)\, d\lambda(g)$$

is the Bayes risk of the estimator $\hat{f}$.

Denote by $L'_s(t, P_g)$ the directional derivative of $t \mapsto L(t, P_g)$ at $t \in \Delta$ in direction $s \in \Delta$. Writing $\bar{q}(\cdot|\cdot)$ for the a posteriori density, it follows from theorem 4.5 of [7] that

(5) $$\int L'_s(\hat{f}(x^n, \cdot), P_g)\, \bar{q}(g|x^n)\, d\nu(g) \geq 0 \quad \text{a.e. for every } s \in \Delta$$

is a sufficient condition for $\hat{f}$ to be a Bayes estimator. In our case,

$$L'_s(t, P_g) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[L(t+\varepsilon s, P_g) - L(t, P_g)] = 2\int (t - Df_g)\, s\, d\mu,$$

and (5) is equivalent to

$$\int_G \int_X [\hat{f}(x^n, x) - Df_g(x)]\, s(x)\, d\mu(x)\, \bar{q}(g \mid x^n)\, dv(g) \geqslant 0 \quad \text{a.e.} \tag{6}$$

Take the σ-finite measure β and the probability measures β_k as above. Then

$$\hat{f}_k(x^n, x) = \frac{\int_G Df_g(x)\, \chi^n(g^{-1}) \prod_{i=1}^n f(g^{-1} x_i)\, d\beta_k(g)}{\int_G \chi^n(g^{-1}) \prod_{i=1}^n f(g^{-1} x_i)\, d\beta_k(g)}$$

obviously fulfils (5) and hence is a Bayes estimator with a priori distribution $\lambda = \beta_k$; further,

$$\hat{f}_0 = \lim_{k \to \infty} \hat{f}_k \quad \text{a.e.}$$

Let, for $k = 0, 1, 2, \ldots$,

$$A_k(x^n, x) := \int_{C_k} Df_g(x)\, \chi^n(g^{-1}) \prod_{i=1}^n f(g^{-1} x_i)\, d\beta(g)$$

and

$$B_k(x^n) := \int_{C_k} \chi^n(g^{-1}) \prod_{i=1}^n f(g^{-1} x_i)\, d\beta(g),$$

where $C_0 := G$. Obviously, $A_k \uparrow A_0$, $B_k \uparrow B_0$ and $\hat{f}_k = A_k/B_k$.

Now, since $Df_g \geqslant 0$,

$$|\hat{f}_0 - \hat{f}_k| = \left|\frac{A_0}{B_0} - \frac{A_k}{B_k}\right| = \left|\frac{A_0}{B_0}\right| \cdot \left|1 - \frac{A_k B_0}{B_k A_0}\right|$$

$$\leqslant |\hat{f}_0| \cdot \left(1 + \left|\frac{B_0}{B_k}\right| \cdot \left|\frac{A_k}{A_0}\right|\right) \leqslant |\hat{f}_0| \cdot \left(1 + \left|\frac{B_0}{B_1}\right|\right).$$

Because $\hat{f}_0(x^n, \cdot) \in L_2(X, \mu)$, and B_0/B_1 is independent of x, Lebesgue's theorem can be applied, yielding

$$\|\hat{f}_0(x^n, \cdot) - \hat{f}_k(x^n, \cdot)\|_{2,\mu} \to 0$$

for almost every x^n.

This proves $\hat{f}_0$ to be a generalized Bayes estimator in the global sense.

4. EXAMPLES

The first examples refer to the case of distributions defined on the real line: $X \subseteq \mathscr{R}$ and μ is the Lebesgue measure.

(a) *Location parameter b unknown.* We take $G = \mathscr{R}$, the additive group, hence $\chi \equiv 1$. The solution is

$$\hat{f}_0(x^n, x) = \frac{\int_{\mathscr{R}} Df(x-b)\,\varphi(b) \prod_{i=1}^{n} f(x_i - b)\,db}{\int_{\mathscr{R}} \varphi^2(b) \prod_{i=1}^{n} f(x_i - b)\,db}. \tag{7a}$$

(b) *Scale parameter $a > 0$ unknown.* We take the multiplicative group $\mathscr{R}_+$, $\chi(a) \equiv a$. We have

$$\hat{f}_0(x^n, x) = \frac{\int_{(0,\infty)} Df(x/a)\,\varphi(a)\,a^{-(n+2)} \prod_{i=1}^{n} f(x_i/a)\,da}{\int_{(0,\infty)} \varphi^2(a)\,a^{-(n+2)} \prod_{i=1}^{n} f(x_i/a)\,da}. \tag{7b}$$

An examination of the proofs in [10] shows that (7b) is also the IOE in the case $X = \mathscr{R}$, $G = \mathscr{R}_+$.

(c) *Both location and scale parameters are unknown.* We take G as the group of all mappings

$$x \mapsto ax + b \qquad (a > 0,\ b \in \mathscr{R})$$

and H as the subgroup of mappings $x \mapsto ax$; then $\chi(a, b) = a$. We have

$$\hat{f}_0(x^n, x) = \frac{\int_{\mathscr{R}} \int_{(0,\infty)} Df\left(\frac{x-b}{a}\right) \varphi(a, b)\,a^{-(n+2)} \prod_{i=1}^{n} f\left(\frac{x_i - b}{a}\right) da\,db}{\int_{\mathscr{R}} \int_{(0,\infty)} \varphi^2(a, b)\,a^{-(n+2)} \prod_{i=1}^{n} f\left(\frac{x_i - b}{a}\right) da\,db}. \tag{7c}$$

As remarked in chapter 1, frequently the square integrability of Df fails to hold. To overcome this difficulty, one can truncate Df_g at given quantiles $x_\alpha^{(g)}$ and $x_\beta^{(g)}$ of the distribution P_g $(0 \leqslant \alpha < \beta \leqslant 1)$.

Let us denote by $\bar{D}$ the truncation of D, namely

$$\bar{D}f_g = (Df_g)\,1\,[x_\alpha^{(g)}, x_\beta^{(g)}].$$

In each of the cases (a)-(c), a problem invariant under G remains invariant under truncation, with the same function φ.

We consider case (c). Writing $x_\alpha = x_\alpha^{(1,0)}$ and $x_\beta = x_\beta^{(1,0)}$, $g = (a, b)$, we have $x_\alpha^{(a,b)} = ax_\alpha + b$, and

$$Df_{(a,b)}(x) = \varphi\big((a, b)^{-1}\big) Df\left(\frac{x-a}{b}\right)$$

follows. Hence formulae (7a) to (7c) can be applied directly.

The calculations leading to the following solutions are omitted here (for sake of completeness, results of [10] are also included).

A. Exponential distribution. Density: $f(x) = e^{-x}\mathbb{1}_{(0,\infty)}(x)$.

Estimation of density: $Df = f$.

(i) $a = 1$, b unknown.

$$\begin{cases} \dfrac{n}{n+1} e^{n(x-x_{(1)})} & \text{if } x < x_{(1)}, \\[2ex] \dfrac{n}{n+1} e^{-(x-x_{(1)})} & \text{if } x \geqslant x_{(1)}. \end{cases}$$

(ii) $b = 0$, $a > 0$ unknown.

$$\frac{n-1}{n}\frac{1}{\bar{x}}\left[1+\frac{x}{n\bar{x}}\right]^{-n}.$$

(iii) $a > 0$, b unknown.

$$\begin{cases} \dfrac{n-2}{n+1}(\bar{x}-x_{(1)})^{-1}\left[1-\dfrac{x-x_{(1)}}{\bar{x}-x_{(1)}}\right]^{-(n-1)} & \text{if } x < x_{(1)}, \\[2ex] \dfrac{n-2}{n+1}(\bar{x}-x_{(1)})^{-1}\left[1+\dfrac{x-x_{(1)}}{n(\bar{x}-x_{(1)})}\right]^{-(n-1)} & \text{if } x \geqslant x_{(1)}. \end{cases}$$

Estimation of reliability function truncated at quantiles:

$$Df(x) = e^{-x}\mathbb{1}_{(x_\alpha, x_\beta)}(x);$$

$$x_\alpha = -\log(1-\alpha), \quad x_\beta = -\log(1-\beta) \quad (0 \leqslant \alpha < \beta \leqslant 1).$$

(i) $a = 1$, b unknown.

$$\begin{cases} \dfrac{n}{n+1}[(1-\alpha)^{n+1}-(1-\beta)^{n+1}]\, e^{n(x-x_{(1)})} & \text{if } x < x_{(1)}+x_\alpha, \\[2ex] \dfrac{n}{n+1}[e^{x_{(1)}-x}-(1-\beta)^{n+1}\, e^{n(x-x_{(1)})}] & \text{if } x_{(1)}+x_\alpha \leqslant x < x_{(1)}+x_\beta, \\[2ex] 0 & \text{if } x \geqslant x_{(1)}+x_\beta. \end{cases}$$

(ii) $b = 0$, $a > 0$ unknown.

$$\frac{\Gamma\left(n+1;\left(1+\frac{n\bar{x}}{x}\right)x_\beta\right)-\Gamma\left(n+1;\left(1+\frac{n\bar{x}}{x}\right)x_\alpha\right)}{\Gamma(n+1)}\left(1+\frac{x}{n\bar{x}}\right)^{-(n+1)}\mathbf{1}_{(0,\infty)}(x)$$

($\Gamma(k, z) = \int_0^z e^{-t}t^{k-1}\,dt$ denotes the incomplete gamma-function).

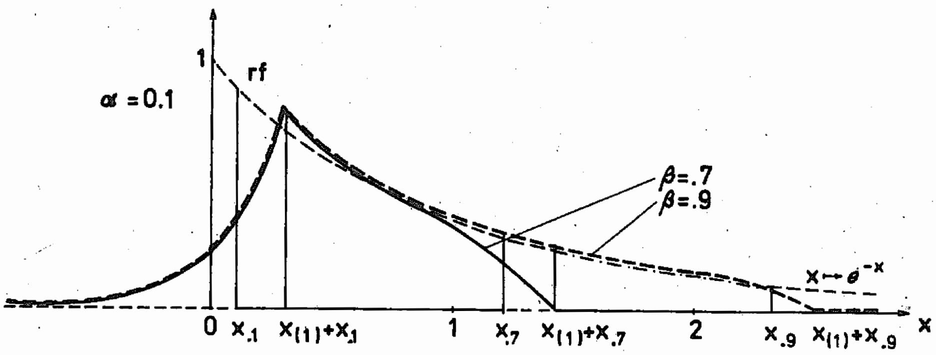

Fig. 1. Estimate of reliability function truncated at x_α, x_β, $n = 5$, $x_{(1)} = 0.2$.
rf – reliability function

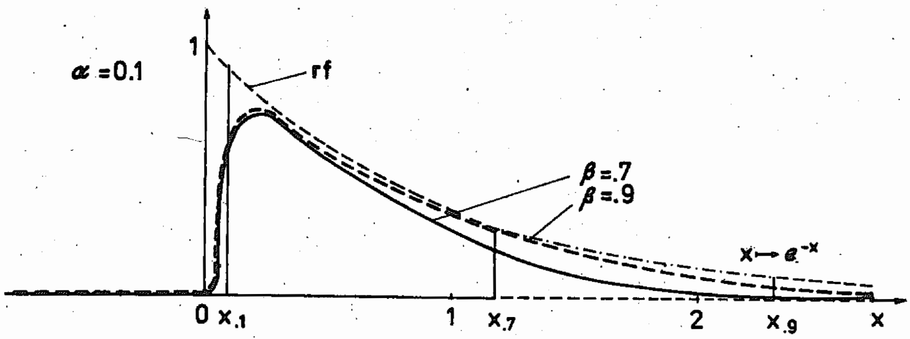

Fig. 2. Estimate of reliability function truncated at x_α, x_β. $n = 5$, $\bar{x} = 1$.
rf – reliability function

Estimation of hazard rate truncated at quantiles: $Df = \mathbf{1}_{(x_\alpha, x_\beta)}$.
$a = 1$, b unknown.

$$\begin{cases}[(1-\alpha)^n-(1-\beta)^n]\,e^{n(x-x_{(1)})} & \text{if } x < x_{(1)}+x_\alpha,\\ 1-(1-\beta)^n e^{n(x-x_{(1)})} & \text{if } x_{(1)}+x_\alpha \leqslant x < x_{(1)}+x_\beta,\\ 0 & \text{if } x \geqslant x_{(1)}+x_\beta.\end{cases}$$

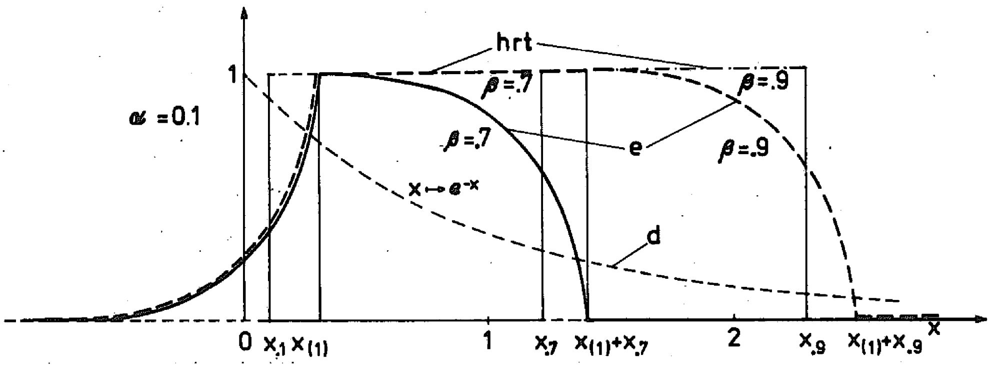

Fig. 3. $n = 5$, $x_{(1)} = 0.2$.
hrt – hazard rate truncated at x_α, x_β; e – estimates; d – density

B. Laplace distribution. Density: $f(x) = \frac{1}{2} e^{-|x|}$.

Estimation of density:
$b = 0$, $a > 0$ unknown.

$$\frac{n-1}{2\sum_{i=1}^{n} |x_i|}\left(1+\frac{|x|}{\sum_{i=1}^{n} |x_i|}\right)^{-n}.$$

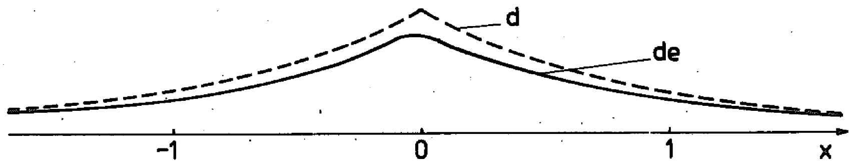

Fig. 4. $n = 5$, $\sum_{i=1}^{n} |x_i| = 5$.
d – density; de – density estimate

C. Uniform distribution. Density: $f = \mathbf{1}_{(0,1)}$.

Estimation of density:
(i) $a = 1$, b unknown (See Fig. 5).

$$\begin{cases} \dfrac{x - x_{(n)} + 1}{x_{(1)} - x_{(n)} + 1} & \text{if } x_{(n)} - 1 \leqslant x < x_{(1)}, \\ 1 & \text{if } x_{(1)} \leqslant x < x_{(n)}, \\ \dfrac{x_{(1)} - x + 1}{x_{(1)} - x_{(n)} + 1} & \text{if } x_{(n)} \leqslant x < x_{(1)+1}, \\ 0 & \text{otherwise.} \end{cases}$$

(ii) $b = 0$, $a > 0$ unknown (See Fig. 6).

$$\begin{cases} 0 & \text{if } x < 0, \\ \dfrac{n-1}{nx_{(n)}} & \text{if } 0 \leqslant x < x_{(n)}, \\ \dfrac{n-1}{n}\dfrac{x_{(n)}^{n-1}}{x^n} & \text{if } x \geqslant x_{(n)}. \end{cases}$$

(iii) $a > 0$, b unknown.

$$\begin{cases} \dfrac{n-2}{n}\dfrac{[x_{(n)}-x_{(1)}]^{n-2}}{[x_{(n)}-x]^{n-1}} & \text{if } x < x_{(1)}, \\ \dfrac{n-2}{n}\dfrac{1}{x_{(n)}-x_{(1)}} & \text{if } x_{(1)} \leqslant x < x_{(n)}, \\ \dfrac{n-2}{n}\dfrac{[x_{(n)}-x_{(1)}]^{n-2}}{[x-x_{(1)}]^{n-1}} & \text{if } x \geqslant x_{(n)}. \end{cases}$$

Estimation of distribution function.

$$F(x) = x\mathbb{1}_{(0,1)}(x)+\mathbb{1}_{[1,\infty)}(x).$$

(i) $a = 1$, b unknown. F truncated at $b+1$: $Df = F\mathbb{1}_{(-\infty,1)}$.

$$\begin{cases} \dfrac{(x-x_{(1)}+1)^2}{2(x_{(1)}-x_{(n)}+1)} & \text{if } x_{(n)}-1 \leqslant x < x_{(1)}, \\ x-\dfrac{x_{(1)}+x_{(n)}-1}{2} & \text{if } x_{(1)} \leqslant x < x_{(n)}, \\ \dfrac{1-(x-x_{(1)})^2}{2(x_{(1)}-x_{(n)}+1)} & \text{if } x_{(n)} \leqslant x < x_{(1)}+1, \\ 0 & \text{otherwise.} \end{cases}$$

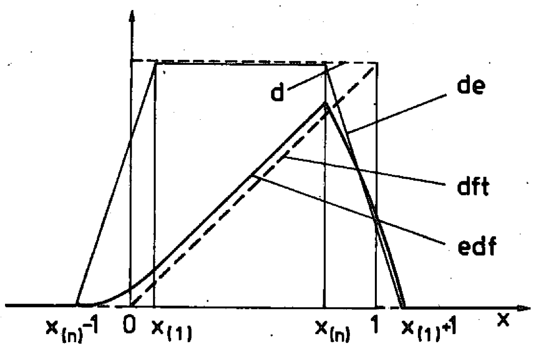

Fig. 5. $n = 5$, $x_{(1)} = 0.1$, $x(n) = 0.8$.
d – density; de – density estimate; dft – distribution function truncated at $b+1 = 1$; edf – estimate of distribution function

(ii) $b = 0$, $a > 0$ unknown. F truncated at ac $(c > 1)$: $Df = F\mathbb{1}_{(-\infty,c)}$.

$$\begin{cases} 0 & \text{if } x < 0, \\ \dfrac{n+1}{n+2}\dfrac{x}{x_{(n)}} & \text{if } 0 \leqslant x < x_{(n)}, \\ 1-\dfrac{1}{n+2}\dfrac{x_{(n)}^{n+1}}{x^{n+1}} & \text{if } x_{(n)} \leqslant x < cx_{(n)}, \\ \left(c^{n+1}-\dfrac{1}{n+2}\right)\dfrac{x_{(n)}^{n+1}}{x^{n+1}} & \text{if } x \geqslant cx_{(n)}. \end{cases}$$

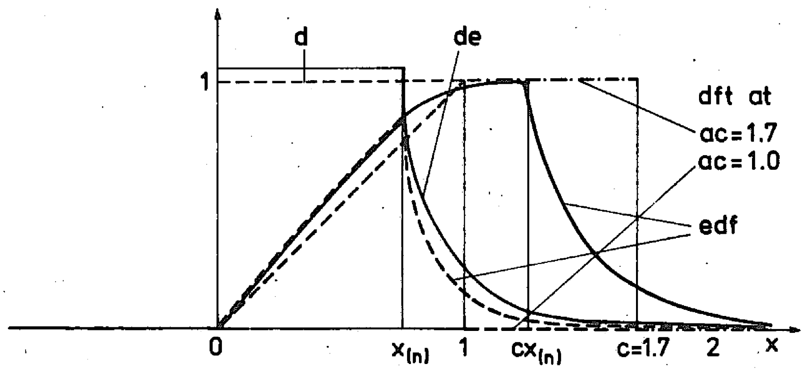

Fig. 6. $n = 5$, $x_{(n)} = 0.75$.
d – density; de – density estimate; dft – distribution function truncated; edf – estimates of distribution function

The following examples are direct consequences of formula (2).

D. k-dimensional normal distribution. $X = G = \mathscr{R}^k$, density $f(\chi) = (2\pi)^{-k/2}e^{-\chi'\chi/2}$ $(\chi\in\mathscr{R}^k)$, $\mu = \nu = \lambda_k$ (k-dimensional Lebesgue measure).

The class of considered distribution consists of the normal ones with unknown vector of expectations and unit covariance matrix, with density $f_b(\chi) = (2\pi)^{-k/2}e^{-(\chi-b)'(\chi-b)/2}$.

The IOE is a normal density with vector of expectations $\bar{\chi} = \frac{1}{n}\sum_{i=1}^{n}\chi_i$ and covariance matrix $\frac{n+1}{n}I$:

$$\hat{f}_0(\chi^n, \chi) = \left(2\pi\frac{n+1}{n}\right)^{-k/2}\exp\left[-\frac{n}{2(n+1)}(\chi-\bar{\chi})'(\chi-\bar{\chi})\right].$$

E. Discrete uniform distribution. We take $X = G = \mathscr{Z}$ (additive group of integers), $\mu = \nu =$ counting measure on $\mathscr{Z}$:

$$\mu(A) = \sum_{i \in \mathscr{Z}} \mathbf{1}_A(i).$$

Let k be fixed and $f(x) = 1/k$ if $x \in \{1, \ldots, k\}$. We estimate $Df = f$, i.e. the probabilities of the discrete distribution.

$$\hat{f}_0(x^n, x) = \begin{cases} \dfrac{1}{k}\,\dfrac{x - x_{(n)} + k - 1}{x_{(1)} - x_{(n)} + k - 1} & \text{if } x_{(n)} - k - 1 \leqslant x < x_{(1)}, \\ 1 & \text{if } x_{(1)} \leqslant x < x_{(n)}, \\ \dfrac{1}{k}\,\dfrac{x_{(1)} - x + k - 1}{x_{(1)} - x_{(n)} + k - 1} & \text{if } x_{(n)} \leqslant x < x_{(1)} + k - 1, \\ 0 & \text{otherwise.} \end{cases}$$

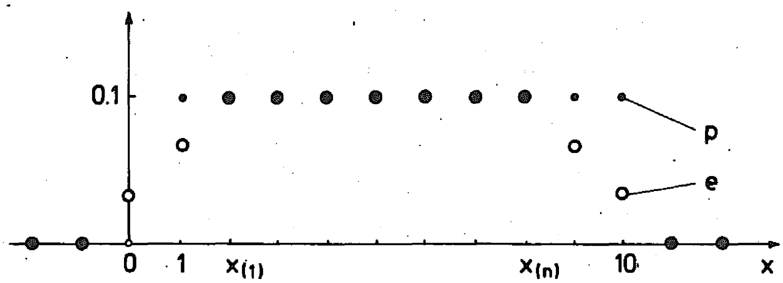

Fig. 7. $k = 10$, $n = 5$; $x_{(1)} = 2$, $x_{(n)} = 8$.
p — probabilities; e — estimates

At first sight, some of the solutions look rather strange, and they obviously differ from the usual parametric methods. The fact stems from the use of the nonparametric type of loss. But in fact, our risk $R(\hat{f}, g)$, the mean integrated square error (MISE), is used in purely nonparametric situations without any objection. Our results can be used to define a notion of efficiency of nonparametric curve estimators (e.g. defined by kernel methods, modified histograms, nearest neighbour rules) by comparing their risk with the risk of the IOE in particular parametric models.

Generally speaking, use of a loss function defined by an integral leads to smooth estimates (cf. the examples above). Uncertainty relative to location smears the estimate over the real line, even if the support of Df is a proper subset of $\mathscr{R}$ (see e.g. the estimates of the exponential density with known scale parameter). If the scale parameter is unknown, usually only finitely many moments of the density estimate exist, even in case as uniform, exponential, normal (see [10]) and Laplace distributions (see Figs. 4 and 6). This phenomenon is closely connected with the remarks in chapter 1.

REFERENCES

[1] M. H. de Groot and M. M. Rao, *Bayes estimation with convex loss*, Ann. Math. Statist. 34 (1963), p. 839-845.

[2] H. Guttman and W. Wertz, *Note on estimating normal densities*, Sankhyā Ser. B 38 (1976), p. 231-237.

[3] J. Hurt and W. Wertz, *Asymptotic properties of the invariantly optimal estimator of reliability in the exponential case*, Statistics and Decisions 1 (1983), p. 197-204.

[4] A. Kozek and W. Wertz, *Estimators with values in Banach spaces under convex loss and applications to curve estimation problems*, Coll. Math. Soc. János Bolyai 32 (1980), p. 567-586.

[5] J. Sacks, *Generalized Bayes solutions in estimation problems*, Ann. Math. Statist. 34 (1963), p. 751-768.

[6] H. Stone, *Right Haar measure for convergence in probability to quasi posterior distributions*, Ann. Math. Statist. 36 (1965), p. 440-453.

[7] H. Strasser, *On Bayes estimates*, J. Multivariate Anal. 3 (1973), p. 311-327.

[8] W. Wertz, *On unbiased density estimation*, An. Acad. Brasil. Ci. 47 (1975), p. 65-72.

[9] — *Invariant density estimation*, Monatsh. Math. 81 (1976), p. 315-324.

[10] — *Über ein nichtparametrisches Schätzproblem*, Metrika 26 (1979), p. 157-167.

[11] — *Invariantly optimal nonparametric estimators with respect to L_1-loss*. "Probability and Statistical Inference" — Proceedings of the 2nd Pannonian Symposium on Mathematical Statistics (W. Grossmann, G. Pflug, W. Wertz, eds.). D. Reidel, Dordrecht, 1982, p. 373-384.

Institut für Statistik
Karlsplatz 13
A-1040 Wien

Received on 25. 04. 1981