ELSEVIER Computational Statistics & Data Analysis 27 (1998) 47–60

COMPUTATIONAL STATISTICS & DATA ANALYSIS

A comparison between Rosenblatt's estimator and parametric density estimators for determining test limits

W. Albers, W.C.M. Kallenberg*, G.D. Otten

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received 1 March 1996; received in revised form 1 July 1997; accepted 1 August 1997

Abstract

Because of measurement errors, test limits instead of specification limits are used for inspection to realize a prescribed bound on the consumer loss. Test limits based on the assumption of normality lead to severe violation of the prescribed bound when normality fails. While relaxing the assumption of normality, it is important to estimate the density of the inspected characteristic at the specification limit correctly. It is investigated whether larger parametric models provide a useful improvement. Simulations are carried out for several such models. It turns out that for estimating a density at a fixed point, the parametric estimators give improvements compared to application of the normal density. However, for small or moderate sample sizes Rosenblatt's estimator is, in general, more accurate than the parametric density estimators.

Keywords: Specification limit; Normal test limit; Robust test limit; Box–Cox model; Johnson system; Pearson system; Exponential power distribution; Monte Carlo experiments

1. Introduction

A well-known adage in statistics reads as follows: "Give me four parameters and I shall describe an elephant; with five, it will wave its trunk". It is the aim of this paper to show that for estimating a density at a fixed point, parametric models and the corresponding densities are not as powerful as suggested by the preceding dictum.

We encountered the aforementioned problem of estimating the density at a fixed point when studying test limits. In order to ensure that a customer is not saddled

* Corresponding author.

0167-9473/98/$19.00
PII S0167-9473(97)00049-2

with a lot of bad products an inspection is carried out on several characteristics of the produced items. For each characteristic a specification limit is given, but because of measurement errors usually a slightly more stringent test limit is set to decide whether an item should be approved. If the test limit is too restrictive, too many items will be rejected which is at the cost of the yield, whereas a test limit that is set too liberally will possibly lead to complaining customers.

Consider the situation where a product is called conforming if the value of the characteristic is below a given specification limit s. Let X denote the value of the characteristic and U the measurement error. The consumer loss is the probability that the product is non-conforming and still accepted. It is given by

$$CL = P(X > s,\ X + U < t). \tag{1}$$

On the one hand, to avoid complaining customers, we require $CL \le \gamma$ with γ quite small. On the other hand, within this restriction, t is as large as possible to protect the yield. Therefore, we are looking for a test limit t such that $CL = \gamma$. Let F be the distribution function of X and f its density. Assume that X and U are independent and that U is normally $N(0,\sigma^2)$-distributed. As a rule the standard deviation σ of the measurement error is small compared to the standard deviation of X. Therefore we expand CL as a function of σ. Let $a = (s-t)/\sigma$ and $Y = -U/\sigma$, implying $Y \sim N(0,1)$. Denote by Φ the standard normal distribution function and by ϕ its density. In view of (1) we write

$$\begin{aligned} CL &= P(X > s,\ X - \sigma Y < t) = P(Y > a,\ s < X < s + \sigma(Y-a)) \\ &= \int_a^\infty \{F(s + \sigma(y-a)) - F(s)\}\phi(y)\,\mathrm{d}y = \sigma f(s) g_1(a) + \cdots \end{aligned} \tag{2}$$

with

$$g_1(a) = \int_a^\infty (y-a)\phi(y)\,\mathrm{d}y = \phi(a) - a\{1 - \Phi(a)\}.$$

Hence as a first order approximation we get

$$a = g_1^{-1}\left(\frac{\gamma}{\sigma f(s)}\right). \tag{3}$$

In most of the literature about statistical tolerancing, screening, inspection etc., X is assumed to be normal too (cf. Mee, 1990; Easterling et al., 1991, and further references in these papers). In that case $f(s) = \phi(s)$ is inserted in (3) (assuming for convenience that $EX = 0$ and $\operatorname{var} X = 1$), leading to the test limit

$$t_N = s - g_1^{-1}\left(\frac{\gamma}{\sigma\phi(s)}\right)\sigma. \tag{4}$$

It is inferred from (2) that the corresponding consumer loss satisfies

$$CL_N = \gamma\,\frac{f(s)}{\phi(s)} + \cdots .$$

Simulation results and supporting theory show that under nonnormality of X test limits as (4), based on normality, violate the prescribed bound on the consumer loss drastically (cf. Albers et al., 1997).

Therefore a new test limit is proposed in Albers et al. (1997), essentially replacing the unknown $f(s)$ in (3) by an estimator $\hat{f}(s)$, using Rosenblatt's kernel estimator. The test limit now reads as

$$\hat{t} = s - g_1^{-1}\left(\frac{\gamma}{\sigma\hat{f}(s)}\right)\sigma$$

and, in view of (2), we get for the corresponding consumer loss

$$\widehat{CL} = \gamma\frac{f(s)}{\hat{f}(s)} + \cdots .$$

The nonparametric approach, using Rosenblatt's estimator, has the desired robustness property: a small loss under normality and a large gain in case of nonnormality in comparison to t_N.

It is clear that a nonparametric estimator of $f(s)$ is only based on observations close to s, since there is no relation assumed between the density at s and the density elsewhere. Therefore, if $f(s)$ is small and the number n of observations not too large, estimation is based on only a few observations and hence not very accurate. (Note that the specification limit s in general lies rather in the tail of the distribution.)

To involve also the bulk of observations in such a case, it is needed to relate $f(s)$ to the density elsewhere. This leads to a parametric model, which should be sufficiently rich to describe densities as they arise in practice rather well.

So, writing $\{f_\vartheta, \vartheta \in \Theta\}$ for the parametric family of densities, there should be $f_\theta(s)$ for some $\theta \in \Theta$ close to the unknown $f(s)$ and $f_{\hat{\theta}}(s)$ should in turn be close to $f_\theta(s)$. Since parametric estimators as a rule converge faster than nonparametric estimators, it is hoped that $f_{\hat{\theta}}(s)$ is closer to $f(s)$ than $\hat{f}(s)$.

Another way of looking at this approach is that the parametric model is a compromise between assuming perfect knowledge of the form of the density (normality) and no knowledge at all (nonparametric estimator). Instead of making a big step at once, an intermediate approach assuming moderate knowledge of the form of the density may give an improvement in particular in situations, where the nonparametric approach is less reliable, i.e. when $f(s)$ is small and n not too large.

As is seen before, the consumer loss is determined by the ratio of the true density at s and its estimator. Therefore from now on we consider the problem of estimating the density of X at a fixed point s by Rosenblatt's estimator and by a parametric estimator.

Starting from the normal family we discuss well-known parametric families, with a great variety of densities, as the exponential power distribution, the Pearson- and Johnson system and Box–Cox's transformation model. Two things are of interest. How well is the true density estimated if the observations are from the model itself and how compare the estimators to each other for reasonably smooth distributions, not from the model.

It turns out that larger parametric models give improvements for a great variety of distributions compared to application of the normal density, but that they hardly can compete with Rosenblatt's estimator. Surprisingly, even with observations from the parametric family itself, for sample sizes up to 400 there is no guarantee for improvement compared to the nonparametric Rosenblatt estimator. Therefore in the test limit problem we recommend to apply the one based on Rosenblatt's estimator, as presented in Albers et al. (1997), when normality fails.

2. Definitions and some theoretical results

Let $X_1,\dots,X_n$ be i.i.d. r.v.'s with unknown density f. Let s be a given point. It is aimed to estimate $f(s)$. Rosenblatt's estimator is defined by

$$\hat{f}_R(s)=(2nh)^{-1}\sum_{i=1}^{n} I_{[s-h,s+h]}(X_i),$$

where

$$I_{[a,b]}(x)=\begin{cases}1\\0\end{cases} \text{if } x \begin{matrix}\in\\ \notin\end{matrix} [a,b].$$

Since

$$p_0=P(\hat{f}_R=0)>0,$$

the "mean squared error" $E\{(f/\hat{f}_R)-1\}^2=\infty$. However, as is shown in Albers et al. (1997) there exists a set B with $P(B^c)$ exponentially small (as $n\to\infty$), such that

$$E\left(\frac{f}{\hat{f}_R}-1\right)^2 I_B\approx c_1h^4+c_2(nh)^{-1},$$

leading to the recommendation (cf. Albers et al., 1997)

$$h=\left\{\frac{1}{\sigma_X}\phi\left(\frac{s-\mu_X}{\sigma_X}\right)n\right\}^{-1/5}, \tag{5}$$

where, if unknown, the expectation μ_X and variance σ_X^2 should be estimated. The convergence rate now equals $n^{-4/5}$.

In the numerical examples of the following sections the exact mean and mean squared error of $(f/\hat{f}_R)I_{\{\hat{f}_R>0\}}$ as well as p_0 are calculated numerically by using the binomial distribution. As we deal in the numerical examples with standardized densities only, in (5) $\mu_X=0$ and $\sigma_X^2=1$ is taken, yielding

$$h=\{n\phi(s)\}^{-1/5}.$$

Also in the parametric case there exists under weak regularity conditions a set B^* with $P((B^*)^c)$ exponentially small (as $n\to\infty$), such that

$$E\left(\frac{f}{f_{\hat{\theta}}}-1\right)^2 I_{B^*}=\left(\frac{f}{f_\theta}-1\right)^2+\mathcal{O}(n^{-1})$$

as $n \to \infty$. Note that if $f = f_\theta$, then the convergence rate equals n^{-1} and hence from a theoretical point of view faster convergence is obtained using the parametric estimator than when using the nonparametric one. For more details we refer to Albers et al. (1994).

In the next sections the parametric estimators are compared with Rosenblatt's estimator. Note that all simulated densities have expectation 0 and variance 1. The fixed point s corresponding to the specification limit is chosen as the 0.99 quantile of the distribution involved.

3. Normal distribution

Consider the family of normal distributions with parameter $\theta = (\mu, \sigma)$, μ being the mean and σ the standard deviation. The unknown parameter θ may be estimated by the maximum likelihood estimator

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}) \quad \text{with } \hat{\mu} = \bar{X} = n^{-1}\sum_{i=1}^{n} X_i \text{ and } \hat{\sigma} = \left\{ n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \right\}^{1/2}.$$

We are interested in estimating the unknown density at s, which lies in the tail of the distribution. By introducing a parametric family we want to use apart from information in the tail also information in the middle of the distribution. Therefore another way of estimating θ is to use the sample median and sample 0.975-quantile. Then $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ is given by (1.96 is the 0.975-quantile of the $N(0,1)$-distribution)

$$\hat{\mu} = \text{sample median}, \quad \hat{\mu} + 1.96\,\hat{\sigma} = \text{sample 0.975-quantile}.$$

So, we compare three estimators of $f(s)$: Rosenblatt's estimator, the normal density with the maximum likelihood estimators of μ and σ, the normal density with the quantile estimators of μ and σ.

Table 1 shows the difference between the three estimators, when sampling from the normal family itself, i.e. when $f(s) = \phi(s)$. (Note that in the simulations we always take "standardized" densities with $EX = 0$ and $\operatorname{var} X = 1$.)

As is seen the parametric estimator based on the maximum likelihood estimators is the best one. In terms of MSE, Rosenblatt's estimator is for $n \leq 400$ almost as good. It performs for $n \leq 400$ better than the parametric estimator based on quantiles. For very large n all three estimators behave very well, the convergence rate of Rosenblatt's estimator being slightly slower.

Next we consider some densities not from the normal family, in particular standardized β- and Γ-densities

$$\beta(p,q): \quad (a_2 - a_1)^{-p-q+1} B(p,q)^{-1} (x - a_1)^{p-1} (a_2 - x)^{q-1}, \quad a_1 < x < a_2,$$

$$\Gamma(p): \quad \frac{1}{b\Gamma(p)} \left(\frac{x-a}{b}\right)^{p-1} \exp\left\{ -\left(\frac{x-a}{b}\right) \right\}, \quad x > a,$$

where B denotes the beta-function and Γ the gamma-function and where a_1, a_2 and a, b, respectively, are chosen in such a way that $EX = 0$ and $\operatorname{var} X = 1$. Analogous to

Table 1
$\phi(s)$ estimated in three ways with s the 0.99-quantile; mean-squared error (MSE) denotes $E(\phi/\hat{f} - 1)^2$ with $\hat{f} = \hat{f}_R$ or $\phi_{\hat{\theta}}$, respectively; mean and MSE of $\phi/\hat{f}_R$ calculated numerically; mean and MSE of $\phi/\phi_{\hat{\theta}}$ estimated by simulation with 10 000 replications

	Rosenblatt			Maximum likelihood estimators of μ, σ		Quantile estimators of μ, σ	
n	p_0	$E\phi/\hat{f}_R$	$\sqrt{MSE}$	$E\phi/\phi_{\hat{\theta}}$	$\sqrt{MSE}$	$E\phi/\phi_{\hat{\theta}}$	$\sqrt{MSE}$
100	0.001	0.809	0.522	1.160	0.599	1.155	1.074
400	0.000	0.828	0.285	1.033	0.214	1.047	0.354
1600	0.000	0.878	0.180	1.007	0.098	1.010	0.157
6400	0.000	0.921	0.113	1.002	0.049	1.002	0.076

the value of p_0 for Rosenblatt's estimator, with the simulation results the number of replications is provided for which the parametric estimate of the density is less than a certain quantity (as a rule 10^{-5}) or cannot be determined by numerical problems. Under f the maximum likelihood estimators (derived in the normal family) converge to $(0,1)$ as $n \to \infty$, since $EX = 0$ and $\sqrt{\operatorname{var} X} = 1$ under f. Therefore the limiting value $f(s)/\phi(s)$ is also presented in Table 2.

Under f the sample median and sample 0.975-quantile converge to the median and 0.975-quantile of X with density f as $n \to \infty$. Let $\theta = (\mu, \sigma)$ be the parameter value for which the normal distribution with mean μ and standard deviation σ has the same median and 0.975-quantile as the distribution of X with density f. Now the limiting value of $f(s)/\phi_{\hat{\theta}}(s)$ equals $f(s)/\phi_\theta(s)$ and therefore this quantity is also presented in Table 2. For some theoretical results on the behaviour of $f(s)/\phi_\theta(s)$ we refer to Albers et al. (1994).

First of all it is seen in Table 2 that estimating $f(s)$ with a normal density, where μ and σ are estimated by the maximum likelihood estimators, may produce large errors. They are (mainly) due to the large ratios $f(s)/\phi(s)$. This confirms the results of Albers et al. (1997). Secondly, the ratio $f(s)/\phi_\theta(s)$ turns out to be very close to 1 and hence a substantial improvement is obtained by using quantile estimators. However, the normal density supplied with the quantile estimators can not be considered as a competitor of Rosenblatt's estimator, especially in the case we are mainly interested in, i.e., $n \le 400$.

4. Exponential power distribution

To get a more rich parametric family, the normal family is extended to the exponential power distribution (EPD) with an extra parameter affecting the tail of the distribution. Thus, apart from location- and scale parameters, a kurtosis parameter comes in. Its density is given by

$$f(x) = \frac{1}{2\sigma\Gamma(\beta+1)} \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|^{1/\beta}\right\}, \quad -\infty < x < \infty.$$

Table 2
Standardized β- and Γ-densities at s (0.99-quantile), estimated in three ways; bias denotes $E(f/\phi_{\hat{\theta}} - f/\phi_{\theta})$; mean squared error ($MSE$) denotes $E(f/\hat{f} - 1)^2$ with $\hat{f} = \hat{f}_R$ or $\phi_{\hat{\theta}}$, respectively; mean and MSE of $f/\hat{f}_R$ calculated numerically; mean and MSE of $f/\phi_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 10^{-5} or the parameters could not be determined

		Rosenblatt			Maximum likelihood estimators of μ, σ			Quantile estimators of μ, σ		
	n	p_0	$Ef/\hat{f}_R$	$\sqrt{MSE}$	f/ϕ	bias	$\sqrt{MSE}$	f/ϕ_θ	bias	$\sqrt{MSE}$
$\beta(2,2)$ ($s = 1.973$)										
	100	0.000	0.948	0.328	1.305	0.102	0.579	1.671	0.031	0.882
	400	0.000	0.972	0.173	1.305	0.022	0.369	1.671	0.012	0.726
	1600	0.000	1.008	0.102	1.305	0.005	0.321	1.671	0.002	0.683
	6400	0.000	1.026	0.065	1.305	0.001	0.309	1.671	0.001	0.674
$\beta(8,32)$ ($s = 2.633$)										
	100	0.003	0.851	0.567	1.664	0.681	2.554	0.980	0.259	1.710
	400	0.000	0.853	0.296	1.664	0.133	0.982	0.980	0.062	0.416
	1600	0.000	0.893	0.179	1.664	0.030	0.737	0.980	0.016	0.178
	6400	0.000	0.930	0.111	1.664	0.006	0.681	0.980	0.004	0.087
$\beta(2,8)$ ($s = 2.853$)										
	100	0.002	0.869	0.567	2.837	2.294	10.60	1.003	0.314	2.360
	400	0.000	0.865	0.287	2.837	0.343	2.570	1.003	0.064	0.423
	1600	0.000	0.901	0.173	2.837	0.090	2.009	1.003	0.013	0.178
	6400	0.000	0.935	0.106	2.837	0.021	1.879	1.003	0.003	0.087
$\Gamma(2)$ ($s = 3.280$)										
	100	0.005	0.834	0.579	6.676	25.44	95.27(5)	0.977	1.022	9.999
	400	0.000	0.848	0.321	6.676	2.741	12.27	0.977	0.146	0.753
	1600	0.000	0.889	0.190	6.676	0.560	6.725	0.977	0.028	0.264
	6400	0.000	0.927	0.117	6.676	0.148	5.935	0.977	0.009	0.125
$\Gamma(6)$ ($s = 2.902$)										
	100	0.005	0.859	0.588	2.707	2.762	13.31	0.962	0.484	3.586
	400	0.000	0.860	0.315	2.707	0.406	2.554	0.962	0.096	0.526
	1600	0.000	0.896	0.185	2.707	0.085	1.888	0.962	0.022	0.218
	6400	0.000	0.931	0.114	2.707	0.022	1.752	0.962	0.005	0.106
$\Gamma(32)$ ($s = 2.582$)										
	100	0.003	0.845	0.568	1.464	0.493	1.856	0.972	0.258	1.755
	400	0.000	0.850	0.300	1.464	0.102	0.727	0.972	0.061	0.413
	1600	0.000	0.890	0.182	1.464	0.022	0.528	0.972	0.017	0.179
	6400	0.000	0.928	0.112	1.464	0.006	0.480	0.972	0.002	0.090

For $\beta = \frac{1}{2}$ we get the family of normal distributions. The larger β, the more heavy the tails. Note that the distribution is symmetric around μ, but that the scale parameter σ is not equal to the standard deviation. In the simulations, for a given β the parameters μ and σ are chosen such that $EX = 0$, $\mathrm{var} X = 1$ (implying $\mu = 0$, $\sigma^2 = \Gamma(\beta)/\Gamma(3\beta)$). This density is denoted by $EPD(\beta)$.

Table 3
Estimation of the *EPD* density at s (0.99 quantile) in four ways; mean squared error (MSE) denotes $E(f_\theta/\hat{f}-1)^2$ with $\hat{f}=\hat{f}_R$ or $f_{\hat{\theta}}$; mean and *MSE* of $f_\theta/\hat{f}_R$ calculated numerically; mean and *MSE* of $f_\theta/f_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 10^{-5} or the parameters could not be determined

		Rosenblatt			Parametric estimators based on:					
					1st and 3rd absolute central moment		Variance and 3rd absolute central moment		Variance and kurtosis	
	n	p_0	$Ef_\theta/\hat{f}_R$	$\sqrt{MSE}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{MSE}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{MSE}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{MSE}$
EPD($\frac{1}{4}$)										
($s=1.913$)	100	0.000	0.709	0.396	2.501	13.89 (63)	1.383	3.460 (8)	4.253	50.05 (130)
	400	0.000	0.764	0.279	1.064	0.279	1.057	0.252	1.551	14.99 (12)
	1600	0.000	0.831	0.193	1.012	0.114	1.012	0.112	1.011	0.098
	6400	0.000	0.889	0.126	1.004	0.055	1.005	0.054	1.004	0.049
EPD($\frac{1}{2}$)										
($s=2.326$)	100	0.001	0.809	0.522	1.603	4.724 (2)	1.594	8.462	3.817	42.04 (5)
	400	0.000	0.828	0.285	1.066	0.250	1.065	0.242	1.060	0.215
	1600	0.000	0.878	0.180	1.013	0.107	1.014	0.106	1.013	0.094
	6400	0.000	0.921	0.113	1.005	0.050	1.005	0.050	1.004	0.045
EPD(1)										
($s=2.766$)	100	0.016	0.925	0.625	1.562	2.807	1.545	2.563	1.864	9.312
	400	0.000	0.918	0.385	1.055	0.220	1.059	0.221	1.065	0.226
	1600	0.000	0.924	0.196	1.014	0.098	1.015	0.095	1.017	0.097
	6400	0.000	0.947	0.116	1.003	0.047	1.003	0.046	1.004	0.046
EPD(2)										
($s=3.107$)	100	0.068	0.894	0.549	1.587	2.895	1.638	3.768	1.788	5.241
	400	0.001	1.023	0.600	1.094	0.275	1.057	0.263	1.044	0.286
	1600	0.000	0.964	0.243	1.027	0.110	1.012	0.103	0.997	0.109
	6400	0.000	0.966	0.134	1.006	0.048	0.999	0.049	0.990	0.059

The parameter μ is estimated by the sample mean. As absolute central moments are easily expressed in the parameters σ and β, estimators of σ and β can be based on it, using the corresponding sample versions. Applying three different combinations of absolute central moments, we obtain as many parametric estimators, which are compared to Rosenblatt's estimator. Table 3 shows the results, when f belongs itself to the *EDP* family.

The main conclusion of this table is that improvement of Rosenblatt's estimator only occurs for large n. But as a rule in that case Rosenblatt's estimator is already sufficiently accurate. Since *EPD*($\frac{1}{2}$) gives the normal density $\phi(s)$, this case may be compared with Table 1. It is seen that for large n the behaviour of the parametric estimators of the *EDP* family is similar to that of the parametric estimators in the normal family. Finally we conclude that the bias and *MSE* do not vary much with the parameter β for the parametric estimators (again for large n).

Table 4
Standardized β- and Γ-densities at s (0.99-quantile), estimated in two ways; bias denotes $E(f/f_{\hat{\theta}} - f/f_{\theta})$; mean squared error ($MSE$) denotes $E(f/\hat{f}-1)^2$ with $\hat{f}=\hat{f}_R$ or $f_{\hat{\theta}}$, respectively; mean and MSE of $f/\hat{f}_R$ calculated numerically; mean and MSE of $f/f_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 10^{-5} or the parameters could not be determined

		Rosenblatt			*EPD*		
	n	p_0	$Ef/\hat{f}_R$	$\sqrt{MSE}$	f/f_θ	Bias	$\sqrt{MSE}$
$\beta(2,2)$ ($s=1.973$)	100	0.000	0.948	0.328	1.341	3.775	93.59 (11)
	400	0.000	0.972	0.173	1.341	0.087	0.580
	1600	0.000	1.008	0.102	1.341	0.019	0.397
	6400	0.000	1.026	0.065	1.341	0.005	0.356
$\beta(8,32)$ ($s=2.633$)	100	0.003	0.851	0.567	1.527	7.046	60.94 (3)
	400	0.000	0.853	0.296	1.527	0.305	1.231
	1600	0.000	0.893	0.179	1.527	0.063	0.651
	6400	0.000	0.930	0.111	1.527	0.016	0.557
$\beta(2,8)$ ($s=2.853$)	100	0.002	0.869	0.567	2.068	13.427	88.21 (11)
	400	0.000	0.865	0.287	2.068	0.484	2.241
	1600	0.000	0.901	0.173	2.068	0.090	1.226
	6400	0.000	0.935	0.106	2.068	0.022	1.105
$\Gamma(2)$ ($s=3.280$)	100	0.005	0.834	0.579	1.797	11.861	65.49 (16)
	400	0.000	0.848	0.321	1.797	0.537	2.106
	1600	0.000	0.889	0.190	1.797	0.108	0.968
	6400	0.000	0.927	0.117	1.797	0.029	0.841
$\Gamma(6)$ ($s=2.902$)	100	0.005	0.859	0.588	1.625	12.148	77.94 (11)
	400	0.000	0.860	0.315	1.625	0.493	1.781
	1600	0.000	0.896	0.185	1.625	0.094	0.795
	6400	0.000	0.931	0.114	1.625	0.024	0.665
$\Gamma(32)$ ($s=2.582$)	100	0.003	0.845	0.568	1.349	4.605	55.84 (2)
	400	0.000	0.850	0.300	1.349	0.213	0.845
	1600	0.000	0.890	0.182	1.349	0.044	0.448
	6400	0.000	0.928	0.112	1.349	0.013	0.375

To investigate whether the unsatisfactory behaviour of the parametric estimators is due to the type of estimators, we have also considered maximum likelihood estimators. Although there was some improvement, the results were still not satisfying. For more details we refer to Albers et al. (1994).

Next the analogue of Table 2 is presented. Since the three parametric estimators do not vary much, we restrict attention to the estimator based on variance and kurtosis. Hence the present results extend those of Table 2 in the sense that now in addition the kurtosis is involved in the estimation. Let $\theta=(\mu,\sigma,\beta)$ be the parameter value for which the *EPD* has the same mean, variance and kurtosis as the distribution of X with density f. The limiting value of $f(s)/f_{\hat{\theta}}(s)$ equals $f(s)/f_\theta(s)$ and therefore this quantity is also presented in Table 4. For some theoretical results on the behaviour of $f(s)/f_\theta(s)$ we refer to Albers et al. (1994).

It is obvious from this table that the parametric estimator cannot beat Rosenblatt's estimator. Compared to the maximum likelihood estimators in the normal family (see Table 2) there is substantial improvement for large n, due to a better ratio f/f_θ when the kurtosis is also involved.

5. Pearson system

To approximate unknown densities a well-known parametric family is the Pearson system. For a detailed description of the system we refer to Johnson and Kotz (1970). Each combination of skewness and kurtosis corresponds to a distribution in the system, implying that the Pearson system contains a large variety of densities.

As parametric estimators we use moment estimators, which are explicitly given in Johnson and Kotz (1970). A quantile approach is also possible, but more complicated. In the Johnson system, which is much alike, the quantile approach is more suitable (see Section 6).

Table 5 shows the comparison between Rosenblatt's estimator and the parametric estimator when f itself belongs to the Pearson system. By β_1 the squared skewness is denoted, while β_2 equals the kurtosis.

While we hoped for an improvement of Rosenblatt's estimator, especially for not too large n, the bad behaviour of the parametric estimator, when sampling from the family itself, is striking. Although from Table 5 clearly the parametric estimator of the Pearson system is not at all a competitor of Rosenblatt's estimator, we shortly discuss how well the Pearson system approximates the unknown density. Therefore we consider *EPD*-densities and the contamined normal distribution (*CND*), given by

$$f(x)=(1-\tau)\frac{1}{\sigma_1}\,\phi\left(\frac{x-\mu_1}{\sigma_1}\right)+\tau\frac{1}{\sigma_2}\,\phi\left(\frac{x-\mu_2}{\sigma_2}\right)$$

with $0<\tau<1$. (Note that β-densities are contained in the Pearson system. Therefore we take other densities than in Tables 2, 4.)

Instead of giving analogues of Tables 2 and 4 we restrict attention to the limiting values $f(s)/f_\theta(s)$. The densities are standardized. In case of *CND*-densities we take $\tau=0.9$ and varying μ_2,σ_2, while μ_1 and σ_1 are such that the mean equals 0 and the variance equals 1.

Table 6 clearly supports the claim in the first paragraph of the introduction.

6. Johnson system

Another well-known parametric family to approximate unknown densities is the Johnson system (cf. Johnson and Kotz, 1970). Quantile estimators are derived by Slifker and Shapiro (1980). They use $\Phi(-3z)$–, $\Phi(-z)$–, $\Phi(z)$–, $\Phi(3z)$-quantiles. More information on these estimators is given in Albers et al. (1994).

Table 5
Estimation of densities from the Pearson system at s (0.99-quantile) in two ways; mean squared error (MSE) denotes $E(f_\theta/\hat{f} - 1)^2$ with $\hat{f} = \hat{f}_R$ or $f_{\hat{\theta}}$, respectively; mean and MSE of $f_\theta/\hat{f}_R$ calculated numerically; mean and MSE of $f_\theta/f_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 10^{-5} or the parameters could not be determined

		Rosenblatt			Pearson	
	n	p_0	$Ef_\theta/\hat{f}_R$	$\sqrt{MSE}$	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{MSE}$
$(\beta_1,\beta_2)=(0,2.5)$	100	0.000	0.814	0.434	6.700	88.321 (124)
	400	0.000	0.842	0.246	1.116	0.567 (1)
	1600	0.000	0.890	0.157	1.016	0.120
$(\beta_1,\beta_2)=(0,3)$	100	0.001	0.808	0.523	5.143	43.697 (123)
	400	0.000	0.828	0.285	1.116	0.915 (89)
	1600	0.000	0.878	0.180	1.019	0.130 (101)
$(\beta_1,\beta_2)=(0,6)$	100	0.012	0.862	0.605	5.441	43.698 (61)
	400	0.000	0.876	0.370	1.431	8.937 (2)
	1600	0.000	0.900	0.201	1.031	0.176
$(\beta_1,\beta_2)=(0,15)$	100	0.023	0.872	0.609	9.638	73.251 (59)
	400	0.000	0.905	0.425	1.241	0.919 (6)
	1600	0.000	0.913	0.211	1.298	8.421 (5)
$(\beta_1,\beta_2)=(0.5,3)$	100	0.001	0.883	0.538	3.424	30.423 (107)
	400	0.000	0.877	0.263	1.083	0.327 (1)
	1600	0.000	0.910	0.159	1.014	0.118
$(\beta_1,\beta_2)=(0.5,4)$	100	0.006	0.859	0.593	5.704	42.318 (139)
	400	0.000	0.861	0.323	1.079	0.411 (196)
	1600	0.000	0.896	0.188	1.014	0.123 (250)
$(\beta_1,\beta_2)=(0.5,6)$	100	0.013	0.866	0.607	2.913	10.446 (80)
	400	0.000	0.881	0.375	1.103	0.415 (13)
	1600	0.000	0.903	0.201	1.010	0.146

Table 6
Contaminated normal and exponential power densities f at s (0.99-quantile) fitted by f_θ from the Pearson system such that skewness and kurtosis correspond to f

(a) $CND(\mu_2, \sigma_2, 0.9)$

μ_2	σ_2	s	f/f_θ
0.0	0.8	2.682	1.469
0.0	0.6	3.332	3.656
−0.2	0.8	2.828	1.160
−0.2	0.6	4.078	3.021

(b) $EPD(\beta)$

β	s	f/f_θ
0.25	1.291	2.016
1	1.138	2.766
2	0.819	3.107

Table 7 compares Rosenblatt's estimator with the parametric estimator when f itself belongs to the Johnson system. Again β_1 is the squared skewness and β_2 is the kurtosis.

Table 7
Estimation of densities from the Johnson system at s (0.99-quantile) in two ways; mean squared error (MSE) denotes $E(f_\theta/\hat{f}-1)^2$ with $\hat{f}=\hat{f}_R$ or $f_{\hat{\theta}}$, respectively; mean and MSE of $f_\theta/\hat{f}_R$ calculated numerically; mean and MSE of $f_\theta/f_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 1/10th of the true density

		Rosenblatt			Johnson		
(β_1,β_2)	n	p_0	$Ef_\theta/\hat{f}_R$	$\sqrt{MSE}$	z	$Ef_\theta/f_{\hat{\theta}}$	$\sqrt{MSE}$
(0, 2.5), B	100	0.0001	0.682	0.453	0.8	1.312	1.210 (61)
	400	0.0000	0.792	0.273	0.9	1.035	0.397
	1600	0.0000	0.869	0.171	0.9	1.013	0.146
(0, 3), N	100	0.0013	0.808	0.523	0.7	1.126	0.872 (34)
	400	0.0000	0.828	0.285	0.8	1.059	0.370
	1600	0.0000	0.878	0.180	0.8	1.019	0.151
(0, 6), U	100	0.0156	0.575	0.575	0.7	1.093	0.886 (12)
	400	0.0000	0.790	0.388	0.8	1.065	0.426
	1600	0.0000	0.875	0.214	0.8	1.016	0.175
(0, 10), U	100	0.0296	0.535	0.587	0.7	1.096	0.947 (10)
	400	0.0001	0.808	0.434	0.8	1.041	0.402
	1600	0.0000	0.885	0.224	0.8	1.012	0.179
(0.49, 3), B	100	0.0009	0.735	0.502	0.8	1.241	0.985 (48)
	400	0.0000	0.834	0.272	0.9	1.011	0.279
	1600	0.0000	0.900	0.162	0.9	1.013	0.139
(0.49, 4), U	100	0.0056	0.619	0.558	0.7	1.105	0.860 (28)
	400	0.0000	0.782	0.343	0.8	1.043	0.368 (1)
	1600	0.0000	0.868	0.201	0.9	1.000	0.147
(0.49, 6), U	100	0.0142	0.577	0.574	0.7	1.146	0.905 (20)
	400	0.0000	0.788	0.384	0.8	1.042	0.412
	1600	0.0000	0.873	0.213	0.8	1.005	0.171

Table 8
Contaminated normal and exponential power densities f at s (0.99-quantile) fitted by f_θ from the Johnson system such that the $\Phi(-3z)$-, $\Phi(-z)$-, $\Phi(z)$-, $\Phi(3z)$-quantiles of X are the same when X has density f_θ as when X has density f

(a) $CND(\mu_2,\sigma_2,0.9)$

z	μ_2	σ_2	s	f/f_θ
0.7	0.0	0.8	2.682	1.529
	0.0	0.6	3.332	2.788
	−0.2	0.8	2.828	1.095
	−0.2	0.6	4.078	2.569
0.8	0.0	0.8	2.682	1.268
	0.0	0.6	3.332	2.399
	−0.2	0.8	2.828	1.250
	−0.2	0.6	4.078	3.075

(b) $EPD(\beta)$

z	β	s	f/f_θ
0.7	0.25	2.016	0.809
	0.50	2.326	1.000
	1.00	2.766	1.168
	2.00	3.107	1.204
0.8	0.25	2.016	0.761
	0.50	2.326	1.000
	1.00	2.766	1.249
	2.00	3.107	1.317

Table 9
Standardized β- and Γ-densities at s (0.99-quantile), estimated in two ways; bias denotes $E(f/f_{\hat{\theta}} - f/f_{\theta})$; mean squared error ($MSE$) denotes $E(f/\hat{f} - 1)^2$ with $\hat{f} = \hat{f}_R$ or $f_{\hat{\theta}}$, respectively; mean and MSE of $f/\hat{f}_R$ calculated numerically; mean and MSE of $f/f_{\hat{\theta}}$ estimated by simulation with 1000 replications; between brackets the number of times that the estimated density was less than 10^{-5} or the parameters could not be determined

		Rosenblatt			Box–Cox		
	n	p_0	$Ef/\hat{f}_R$	$\sqrt{MSE}$	f/f_θ	bias	$\sqrt{MSE}$
$\beta(2,2)\,(s=1.973)$	100	0.000	0.948	0.328	1.113	0.162	0.636 (28)
	400	0.000	0.972	0.173	1.113	0.023	0.192
	1600	0.000	1.008	0.102	1.113	0.009	0.134
$\beta(8,32)\,(s=2.633)$	100	0.003	0.851	0.567	0.990	1.064	5.300 (9)
	400	0.000	0.853	0.296	0.990	0.099	0.469
	1600	0.000	0.893	0.179	0.990	0.025	0.171
$\beta(2,8)\,(s=2.853)$	100	0.002	0.869	0.567	0.937	0.550	2.784
	400	0.000	0.865	0.287	0.937	0.046	0.296
	1600	0.000	0.901	0.173	0.937	0.013	0.136
$\Gamma(2)\,(s=3.280)$	100	0.005	0.834	0.579	1.008	0.849	3.409 (4)
	400	0.000	0.848	0.321	1.008	0.092	0.441
	1600	0.000	0.889	0.190	1.008	0.027	0.171
$\Gamma(6)\,(s=2.902)$	100	0.005	0.859	0.588	1.032	1.834	9.718 (8)
	400	0.000	0.860	0.315	1.032	0.104	0.515
	1600	0.000	0.896	0.185	1.032	0.035	0.201
$\Gamma(32)\,(s=2.582)$	100	0.003	0.845	0.568	1.020	1.850	11.19 (9)
	400	0.000	0.850	0.300	1.020	0.115	0.592
	1600	0.000	0.890	0.182	1.020	0.029	0.193

We conclude that, compared to the Pearson system, the (quantile) estimators of the Johnson system perform much better. Nevertheless, for $n = 100$ Rosenblatt's estimator is still substantially better. So again we have the surprising situation that for not too large n the parametric estimator is worse, even if the sampling is from the parametric family itself.

Table 8 shows how well unknown densities are approximated by the Johnson system, when applying quantile estimators.

Although in case of *EPD*-densities there is a substantial improvement compared to the Pearson system, approximations to *CND*-densities remain poor and hence the claim in the first paragraph of the introduction is still in action.

7. Box–Cox transformation

Box and Cox (1964) introduced a parametric model by assuming that the transformed observation has a known type of distribution, for which we take the normal family. More specifically, we say that X belongs to the Box–Cox model if Y, defined by

$$Y = \begin{cases} \frac{(X+\lambda_2)^{\lambda_1}-1}{\lambda_1}, & \lambda_1 \neq 0, \\ \log(X+\lambda_2), & \lambda_1 = 0 \end{cases}$$

is $N(\mu_Y, \sigma_Y^2)$-distributed. If $(\lambda_1\mu_Y+1)/(\lambda_1\sigma_Y)$ is sufficiently large, then the probability that $\lambda_1 Y+1$ is negative is sufficiently small. We will assume that this indeed is the case, and we ignore that X is not well defined on this set of small probability.

Estimators based on sample-quantiles are derived in Albers et al. (1994). Here we show (Table 9) how the parametric approach compares to Rosenblatt's estimator, when sampling from standardized β- and Γ-densities, not belonging to the family (cf. Tables 2,4).

Although the ratios $f(s)/f_\theta(s)$ are close to one, there is no improvement in *MSE*-terms compared to Rosenblatt's estimator. On the contrary, for $n \leq 400$ (which is of special interest to us) Rosenblatt's estimator performs (much) better.

8. Conclusions

For estimating the density f at a fixed point s parametric models are studied as a compromise between assuming perfect knowledge of the form of the density (like normality) and no knowledge at all (nonparametric approach). Special attention is focussed on the situation where $f(s)$ is small and the number of observations not too large. This estimation problem arises e.g. in determining test limits in quality control.

It turns out that larger parametric models give improvements for a great variety of distributions compared to application of the normal density, but that they hardly can compete with Rosenblatt's estimator. Surprisingly, even with observations from the parametric family itself, for sample sizes up to 400 there is no guarantee for improvement compared to the nonparametric Rosenblatt estimator. Therefore in the test limit problem we recommend to apply the one based on Rosenblatt's estimator, as presented in Albers et al. (1997), when normality fails.

References

Albers, W., Kallenberg, W.C.M., Otten, G.D., 1994. A comparison between the use of parametric density estimators and a kernel estimator to determine test limits. Memorandum 1211, University of Twente.

Albers, W., Kallenberg, W.C.M., Otten, G.D., 1997. Robust test limits. Math. Meth. Statist., to appear.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Statist. Soc. Ser. B 26, 211–252.

Easterling, R.G., Johnson, M.E., Bement, T.R., Nachtsheim, C.J., 1991. Statistical tolerancing based on consumer's risk considerations. J. Quality Technol. 23, 1–11.

Johnson, N.L., Kotz, S., 1970. Distributions in Statistics: Continuous Univariate distributions, vol. 1. Wiley, New York.

Mee, R.W., 1990. An improved procedure for screening based on a correlated, normally distributed variable. Technometrics 32, 331–337.

Slifker, J.F., Shapiro, S.S., 1980. The Johnson-system: selection and parameter estimation. Technometrics 22, 239–246.