Report 2000-011

Small nonparametric tolerance regions

A. Di Bucchianico, J.H.J. Einmahl,
and N.A. Mushkudiani
ISSN 1389-2355

Eindhoven, April 2000
The Netherlands

EURANDOM
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Secretariat: Laplace Building 1.09
Telephone: +31 40 247 8100
Telefax: +31 40 247 8190
Email: all@eurandom.tue.nl
WWW: http://www.eurandom.nl

ISSN 1389-2355

Small nonparametric tolerance regions*

A. Di Bucchianico[†] J.H.J. Einmahl[‡] N.A. Mushkudiani[§]

Abstract

We present a new, natural way to construct nonparametric multivariate tolerance regions. Unlike the classical nonparametric tolerance intervals, where the endpoints are determined by beforehand chosen order statistics, we take the shortest interval, that contains a certain number of observations. We extend this idea to higher dimensions by replacing the class of intervals by a general class of indexing sets, which specializes to the classes of ellipsoids, hyperrectangles or convex sets. The asymptotic behaviour of our tolerance regions is derived using empirical process theory, in particular the concept of generalized quantiles. Finite sample properties of our tolerance regions are investigated through a simulation study. A real data example is also presented.

Key words and phrases. Nonparametric tolerance region, prediction region, empirical process, asymptotic normality, minimum volume set.

AMS 1991 subject classification. 62G15, 62G20, 62G30, 60F05.

---

*This report is a revision of COSOR memorandum 98-16, Eindhoven University of Technology. Research partially supported by European Union HCM grant ERB CHRX-CT 940693.

[†]EURANDOM and Eindhoven University of Technology, Department of Technology Management, Section Quality of Products and Processes, both P.O. Box 513, 5600 MB Eindhoven, the Netherlands, email: `A.d.Bucchianico@tm.tue.nl`

[‡]Corresponding author. EURANDOM and Eindhoven University of Technology, Department of Mathematics and Computing Science, both P.O. Box 513, 5600 MB Eindhoven, The Netherlands, email: `j.h.j.einmahl@tue.nl`

[§]Eindhoven University of Technology, Department of Mathematics and Computing Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, email: `anino@win.tue.nl`

1 Introduction

Several practical statistical problems require information on the distribution itself rather than on functionals of the distribution like, e.g., mean and variance. For example, in life testing of new products it is required that a certain percentage of sold products will not fail before the end of the warranty period. There are many other examples of this kind in various fields, such as reliability theory, medical statistics, chemistry, quality control, etc. (see e.g., Aitchison and Dunsmore (1975)). The statistical literature provides tolerance intervals and regions as a solution to these problems. Starting with Wilks (1941), many papers on this topic have appeared. The monographs Aitchison and Dunsmore (1975) and Guttman (1970) provide thorough overviews of the literature, while extensive bibliographies can be found in Jílek (1981) and Jílek and Ackermann (1989). Although there is a vast literature on the two types of tolerance regions (guaranteed coverage and mean coverage in the terminology of Aitchison and Dunsmore (1975) or β-content and β-expectation in the terminology of Guttman (1970)), statistics text books, both the mathematically and the engineering oriented ones, hardly deal with this topic explicitly. This is surprising since prediction regions are in fact β-expectation/mean coverage tolerance regions. We refer to the introduction of Carroll and Ruppert (1991) for useful remarks on this issue, in particular on when to use which type of tolerance region. In case tolerance regions are mentioned in text books, the treatment is often confined to tolerance *intervals* for the normal distribution. In practice, however, one often encounters situations where the data are not normally distributed or univariate. In order to deal with the first problem, nonparametric tolerance intervals are used. The idea, which first appeared in the seminal paper Wilks (1941), is to consider intervals with two order statistics as endpoints. It is important to note that it is decided *beforehand* which order statistics to take.

In the spirit of the shorth (see e.g. Rousseeuw and Leroy (1988), Grübel (1988)), we propose a new approach to nonparametric tolerance intervals by taking the shortest interval that contains a certain number of order statistics. Surprisingly, the asymptotic theory concerning content (or coverage) is the same as for the classical procedure, although obviously by definition our intervals are not longer, and often much shorter. A problem with nonparametric techniques in higher dimensions is that there is no canonical ordering. In order to overcome this problem, essentially one-dimensional procedures such as statistically equivalent blocks were developed to construct multivariate tolerance regions (see Wald (1943), Tukey (1947, 1948), Fraser (1953) and more recently Ackermann (1983)). From a statistical point of view, there is much arbitrariness in these procedures, since they depend on auxiliary ordering functions. Moreover, they are not necessarily asymptotically minimal (see Chatterjee and Patra (1980)). Instead, one would like to have a genuine multivariate procedure, that is not based on ordering the data. In Chatterjee and Patra (1980) a procedure is presented based on nonparametric density estimation, which yields asymptotically minimal tolerance regions. Our procedure is inspired by empirical process theory and extends to higher dimensions in a natural way. It avoids the choices that have to be made when estimating densities. On the other hand, we have to choose an indexing class to parametrize our empirical process, which however has the advantage that we can choose the shape of the tolerance region. We will show that our procedures are asymptotically *correct*, in contrast to those in Chatterjee and Patra (1980) where only asymptotic conservatism is shown. Our tolerance regions are asymptotically minimal with respect to the indexing class and have desirable invariance properties. In medical statistics, multivariate tolerance regions based on data from, e.g., blood counts, can be used for screening of patients. In this paper, we will illustrate our approach by computing tolerance regions for bi- and trivariate observations in such a situation. Multivariate tolerance regions can be applied in several other fields. E.g., in statistical process control a multivariate approach to capability studies (which, if properly conducted, should be based on tolerance regions) is highly desirable, when various quality characteristics are taken into account.

This paper is organized as follows. Section 2 reviews some background material, while Section 3 contains the main results. In Section 4 we study the finite sample properties of our tolerance regions through simulations and apply the methods to a real data example. Section 5 contains the proofs of the results in Section 3.

2 Preliminaries

Below we specify our setup and notation. We also state some preliminary results for convenient reference later on. Let $X_1, \ldots, X_n$, $n \geq 1$, be i.i.d. $\mathbb{R}^k$-valued random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a common probability distribution P, absolutely continuous with respect to Lebesgue measure, and corresponding distribution function F. Let $\mathcal{B}$ be the σ-algebra of Borel sets on $\mathbb{R}^k$ and define the pseudo-metric d_0 on $\mathcal{B}$ by

$$d_0(B_1, B_2) = P(B_1 \triangle B_2), \quad \text{for } B_1, B_2 \in \mathcal{B}.$$

Denote by P_n the empirical distribution:

$$P_n(B) = \frac{1}{n} \sum_{i=1}^{n} I_B(X_i), \quad B \in \mathcal{B},$$

where I_B is the indicator function of the set B.

Let $\mathcal{E}$ be the class of all closed ellipsoids A in $\mathbb{R}^k$. Fix $t_0 \in (0,1)$ and $C \in \mathbb{R}$. Set $p_n = t_0 + \frac{C}{\sqrt{n}}$. For n large enough, we need existence and uniqueness of an ellipsoid $A_{n,t_0,C} \in \mathcal{E}$ of minimum volume such that $P_n(A_{n,t_0,C}) \geq p_n$, almost surely. In other words, $A_{n,t_0,C}$ should contain at least $\lceil np_n \rceil$ observations. The sets $A_{n,t_0,C}$ are our candidate tolerance regions. The existence and a.s. uniqueness of such an ellipsoid $A_{n,t_0,C}$ was proved in Davies (1992). There are between $k+1$ and $k(k+3)/2$ points on the boundary of $A_{n,t_0,C}$ in dimension k (see Silverman and Titterington (1980)) and hence,

$$t_0 + \frac{C}{\sqrt{n}} \leq P_n(A_{n,t_0,C}) < t_0 + \frac{C}{\sqrt{n}} + \frac{k(k+3)}{2n} \quad \text{a.s. .}$$

However with some more effort it can be shown that a minimum volume ellipsoid that contains *at least* m out of n points, contains *exactly* m points, a.s. (see Lemma 3 at the end of Section 5). This result seems not to be present in the literature. It yields that

$$P_n(A_{n,t_0,C}) = \frac{1}{n} \left\lceil n \left(t_0 + \frac{C}{\sqrt{n}} \right) \right\rceil \quad \text{a.s. .} \tag{1}$$

Let $\mathcal{R}$ be the class of all closed hyperrectangles with faces parallel to the coordinate axes. It is easy to adapt the proof of Davies (1992) to $\mathcal{R}$. Hence, there exists an a.s. unique smallest volume hyperrectangle $A_{n,t_0,C} \in \mathcal{R}$, with $P_n(A_{n,t_0,C}) \geq p_n$. Since with probability one, all hyperplanes parallel to the coordinate axes contain at most one observation, the equality in (1) holds here too.

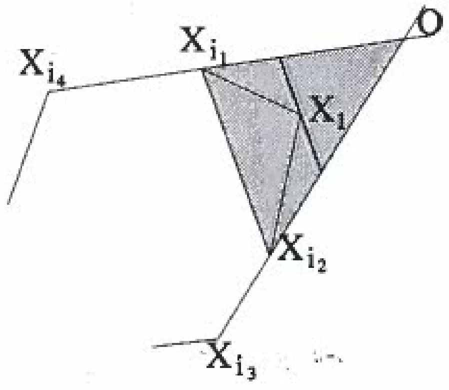

Figure 1: Uniqueness of minimum area convex set.

Consider now the existence and a.s. uniqueness problem of $A_{n,t_0,\mathcal{C}}$ for $\mathcal{C}$, the class of all closed convex sets in $I\!R^2$. It is a well-known fact that the convex hull of $\mathcal{X} = \{X_1, \ldots, X_n\}$ is a bounded polyhedral set in $I\!R^2$ (i.e., a bounded set which is the intersection of finitely many half-planes, see e.g., Webster (1994), Theorem 3.2.5), and thus a polygon. Since the convex hull of $\mathcal{X}$ is the smallest (with respect to set inclusion) convex set containing $\mathcal{X}$, it follows that the closed convex hull of $\mathcal{X}$ is the a.s. unique smallest area closed convex set containing $\mathcal{X}$. As the number of subsets of $\mathcal{X}$ is finite, the existence of a smallest area convex subset containing $\lceil np_n \rceil$ points from $\mathcal{X}$ is assured. Hence, it is left to show that with probability 1, any two different convex hulls of subsets of the sample will have different areas. Suppose we have two sets of vertices $\{X_{i_1}, \ldots, X_{i_\ell}\}$ and $\{X_{j_1}, \ldots, X_{j_k}\}$, $3 \leq \ell$, $k \leq n$ with convex hulls A_1 and A_2, respectively. Without loss of generality we assume that X_1 is a vertex of A_1, but not of A_2. If we condition on $\{X_2, \ldots, X_n\}$, then we have to show that for any positive v

$$I\!P\{X_1 : V(A_1) = v \mid X_2, \ldots, X_n\} = 0, \tag{2}$$

where $V(A_1)$ denotes the area of A_1. Since A_1 is convex, X_1 lies in the interior of the triangle $X_{i_1}OX_{i_2}$ (see Figure 1), for any neighbouring vertices X_{i_1} and X_{i_2}. As the area of A_1 is fixed, X_1 can be only on some interval parallel to $X_{i_1}X_{i_2}$. (Actually, we assumed $5 \leq \ell \leq n$, but a similar argument works for $\ell = 3$ or $\ell = 4$.) Hence, we see that (2) holds. Finally, it is obvious that (1) holds for $\mathcal{C}$.

Remark 1 Observe that unlike for the classes above, the minimum volume problem has no unique solution for the case of *all* hyperrectangles in $I\!R^k$. Consider a random sample of size n in, e.g., $I\!R^2$. Then with positive probability, there are 3 sample points that form an acute triangle such that the remaining $n-3$ sample points are in the interior of that triangle. In this case, there are 3 minimal area rectangles that contain the sample.

Here are some more definitions and results. By the *Blaschke Selection Principle* (see e.g. Webster (1994), Theorem 2.7.10), every sequence of non-empty compact convex sets contained in a compact subset of $I\!R^k$ has a subsequence that converges in the Hausdorff metric to some non-empty compact convex set in $I\!R^k$. By Shephard and Webster (1965), the Hausdorff and the symmetric difference metric $d(A, B) := V(A \triangle B)$, where V denotes volume (Lebesgue measure), are equivalent on the class of all compact convex subsets of $I\!R^k$ with non-empty interior. Hence, we have convergence in the Hausdorff metric if and only if we have convergence in the symmetric difference metric d.

Define for any class $\mathcal{A} \subset \mathcal{B}$:

$$\begin{aligned} \widetilde{F}_n(y) &= \sup_{A \in \mathcal{A}} \{P_n(A) : V(A) \leq y\}, \\ \widetilde{F}(y) &= \sup_{A \in \mathcal{A}} \{P(A) : V(A) \leq y\}, \quad y > 0, \end{aligned}$$

and introduce as in Einmahl and Mason (1992) the generalized empirical quantile and quantile functions, based on P, V and $\mathcal{A}$ by

$$\begin{aligned} U_n(t) &= \inf_{A \in \mathcal{A}} \{V(A) : P_n(A) \geq t\}, \\ U(t) &= \inf_{A \in \mathcal{A}} \{V(A) : P(A) \geq t\}, \quad t \in (0,1); \end{aligned}$$

set $U(t) = 0$ for $t \leq 0$, and $U(t) = \lim_{s \uparrow 1} U(s)$ for $t \geq 1$.

3 Main results

In this section we present the asymptotic results on our tolerance regions. Recall the notation of Section 2. Let $\mathcal{A}$ be a class of Borel-measurable subsets of $\mathbb{R}^k$. (We assume that $\mathcal{A}$ is such that no measurability problems occur.)

Theorem 1 *Fix $t_0 \in (0,1)$ and let $C \in \mathbb{R}$. Assume the following conditions are fulfilled:*

C1) $\mathcal{A}$ is P-Donsker: $\sqrt{n}(P_n - P)$ converges weakly on $\mathcal{A}$ (in the sense of Dudley (1978)) to a bounded, mean zero Gaussian process B_P; B_P is uniformly continuous on $(\mathcal{A}, d_0)$ and has covariance function $P(A_1 \cap A_2) - P(A_1)P(A_2)$, $A_1, A_2 \in \mathcal{A}$,
C2) With probability 1, there exist a unique set $A_{n,t_0,C} \in \mathcal{A}$ with minimum volume and

$$P_n(A_{n,t_0,C}) \geq t_0 + \frac{C}{\sqrt{n}},$$

C3) As $n \to \infty$,

$$P_n(A_{n,t_0,C}) = t_0 + \frac{C}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad a.s.,$$

C4) A_{t_0}, the set in $\mathcal{A}$ with minimum volume and $P(A_{t_0}) = t_0$, exists, is unique, and

$$d(A_{n,t_0,C}, A_{t_0}) \to 0 \quad a.s. \quad (n \to \infty).$$

Then we have

$$\text{(3)} \qquad \sqrt{n}(t_0 - P(A_{n,t_0,C})) + C \xrightarrow{d} Z\sqrt{t_0(1-t_0)} \quad (n \to \infty),$$

where Z is a standard normal random variable.

The following theorems, which are corollaries to Theorem 1, are actually our main general results about tolerance regions. In fact, we will show that the sets $A_{n,t_0,C}$, for suitable C, are asymptotic tolerance regions. Theorem 2 gives the result for guaranteed coverage tolerance regions, whereas Theorem 3 deals with mean coverage tolerance (or prediction) regions. We show that the guaranteed coverage tolerance regions have indeed asymptotically the correct confidence level, whereas the mean coverage tolerance regions have the correct mean coverage with error rate $o(1/\sqrt{n})$. These results are new and of interest in any finite dimension, *including* dimension one. The numbers t_0 and $1-\alpha$ denote the (desired) coverage and confidence level, respectively.

Theorem 2 *Fix $\alpha \in (0,1)$ and let $C = C(\alpha)$ be the $(1-\alpha)$-th quantile of the distribution of $Z\sqrt{t_0(1-t_0)}$. Under the conditions of Theorem 1 we have*

$$\lim_{n\to\infty} \mathbb{P}\{P(A_{n,t_0,C}) \geq t_0\} = 1-\alpha. \tag{4}$$

Theorem 3 *If the conditions of Theorem 1 hold and $\sqrt{n}(t_0 - P(A_{n,t_0,0}))$ is uniformly integrable, then*

$$\mathbb{E}P(A_{n,t_0,0}) = t_0 + o\left(\frac{1}{\sqrt{n}}\right), \quad n \to \infty. \tag{5}$$

Note that $\mathbb{E}P(A_{n,t_0,C}) \to t_0$, $n \to \infty$, for *every* $C \in \mathbb{R}$.

In the final theorem, we will specialize our general results to three natural and relevant indexing classes, which satisfy the conditions of the above theorems. From the point of view of applications, this is the main result of the paper. In the sequel, $\mathcal{A}$ will be one of the following classes: all closed

(a) ellipsoids,

(b) hyperrectangles with faces parallel to the coordinate axes,

(c) convex sets (for $k = 2$)

that have probability strictly between 0 and 1.

These classes of sets are very natural for constructing nonparametric tolerance regions. The class of ellipsoids in (a) is a good choice, since elliptically contoured distributions are considered to be natural and important in probability and statistics. The multivariate normal distribution is of course a prominent example. One should choose the parallel hyperrectangles of (b) as indexing class, if it is desirable, like in many applications, to have a multivariate tolerance region that can be decomposed into (easily interpretable) tolerance intervals for the individual components of the random vectors. The convex sets of (c), which reduce to tolerance regions that are convex polygons, are very natural, since when taking the convex hull of a finite set of data points, one hardly feels the restriction due to the underlying indexing class.

Theorem 4 *Fix $t_0 \in (0,1)$. If the density f of the distribution function F is positive on some connected, open set $S \subset \mathbb{R}^k$ and $f \equiv 0$ on $\mathbb{R}^k \backslash S$, and if A_{t_0}, the set in $\mathcal{A}$ with minimum volume and $P(A_{t_0}) = t_0$, exists and is unique, then we have for the cases (a) and (b) that (3),(4) and (5) hold.*
If $k = 2$ and, in addition, f is bounded, then (3),(4) and (5) also hold for case (c).

Remark 2 Theorem 4 is valid under very mild conditions. In particular, there are no smoothness conditions on the density f. The uniqueness of A_{t_0}, however, is crucial for the results as stated. If it is not satisfied the results can be substantially different. On the other hand, uniqueness of A_{t_0} is a mild condition and holds for many (multimodal) distributions.

Note that it is well-known, see e.g. Dudley (1982), that for dimension 3 or higher there is no weak convergence of the empirical process indexed by closed convex sets, since this class of sets has a too large entropy. (Actually the supremum of the absolute value of this empirical process tends to infinity, in probability, as $n \to \infty$.) This means that for this case Theorem 4, if true at all, can not be proved with the methods presented in this paper.

Remark 3 Since our general tolerance regions $A_{n,t_0,C}$ converge in probability to A_{t_0}, they are asymptotically minimal with respect to the chosen indexing class. That means, e.g. for case (a), that no tolerance ellipsoids can be found the volume of which converge to a number smaller than $V(A_{t_0})$. It is well-known that under weak additional conditions (see, e.g., Chatterjee and Patra (1980)) there exists a region of the form $\{x \in I\!R^k : f(x) \geq c\}$, for some $c > 0$, that has probability t_0 and minimal Lebesgue measure. Such a minimal region is unique up to sets of Lebesgue measure 0. If the above level set belongs to the indexing class we use, then our tolerance regions are asymptotically minimal (with respect to *all* Borel-measurable sets).

At a finer scale, it seems possible to prove (under additional conditions) along the lines of Einmahl and Mason (1992) that in fact $V(A_{n,t_0,C}) = V(A_{t_0}) + O_{I\!P}(n^{-1/2})$.

Remark 4 It is rather easy to show that the tolerance regions of Theorem 4 have desirable invariance properties. For cases (a) and (c) the tolerance region $A_{n,t_0,C}$ is affine equivariant, i.e. for a nonsingular $k \times k$ matrix M and a vector v in $I\!R^k$, we have that $MA_{n,t_0,C} + v$ is the tolerance region corresponding to the $MX_i + v$. (Here $MA_{n,t_0,C} = \{Mx : x \in A_{n,t_0,C}\}$.) Since case (b) deals with parallel hyperrectangles, this property does not hold in full generality for this case, but it does hold when M is a nonsingular diagonal matrix, which means that we allow affine transformations of the coordinate axes.

Remark 5 Let $m > 1$ be an integer and let $\mathcal{A} \subset \mathcal{B}$ be the class consisting of
(a') unions of m closed ellipsoids,
(b') unions of m closed parallel hyperrectangles, or
(c') unions of m closed convex sets, contained in a fixed, large compact set (for $k = 2$),
with probability strictly between 0 and 1, respectively. Note that a minimum volume set $A_{n,t_0,C}$ consists of at most m 'components' and that some of these components may have an empty interior. Note as well that now a minimum volume set $A_{n,t_0,C}$ need not be almost surely unique, hence C2) is not satisfied, but we still have the second part of C4) of Theorem 1, which yields 'asymptotic uniqueness'. Since also C1), the 'existence part' of C2), and C3) are satisfied we see that Theorem 4 remains true when replacing the cases (a), (b) and (c) by (a'), (b') and (c'), respectively. This can be relevant for multimodal distributions. However, often the indexing classes of cases (a), (b) and (c) suffice, since in many (multimodal) situations the smallest closed set having probability t_0, is a 'nice' connected set, because t_0 is typically close to 1. Note that Remark 4, mutatis mutandis, holds true for the classes defined in (a'), (b') and (c').

4 Simulation study and a real data example

First we present results on the finite sample behaviour of our tolerance regions through simulations. Each simulation consisted of 1000 replications. Note that the asymptotic behaviour of our tolerance regions does not change if we vary the number of observations in the tolerance regions within $o(\sqrt{n})$. However, even for the classical tolerance *intervals*, the finite sample behaviour is very sensitive to the actual number of used order statistics (see Table 1).

number of order statistics	93	94	95	96	97
confidence level	67.9%	79.3%	88.3%	94.2%	97.6%

Table 1: Sensitivity of classical 90% guaranteed coverage tolerance intervals with $n = 100$.

Simulations showed a similar sensitivity for our tolerance regions. Moreover, including exactly $\lceil np_n \rceil$ observations we obtained slightly too low coverages, resulting in too low simulated confidence levels. Since the boundary of a tolerance region has probability zero, we decided to add the number of points on the boundary of our tolerance regions to $\lceil np_n \rceil$.

For the classical tolerance intervals, we of course used an exact calculation, based on the beta distribution, for the number of observations to be included. These intervals were chosen in such a way that the indices of the order statistics that serve as endpoints are (almost) symmetric around $(n + 1)/2$. We thus expect *our* tolerance intervals to be substantially shorter for skewed distributions, as they automatically scan for the interval with highest mass concentration. As mentioned above, we added 2 observations when constructing our tolerance intervals. Tables 2 and 3 contain our simulation results for guaranteed coverage and mean coverage tolerance intervals.

distribution	sample size	simulated confidence level		average length	
		classical	new	classical	new
standard normal	300	95.7%	92.2%	3.61	3.57
	1000	95.3%	90.5%	3.46	3.43
standard Cauchy	300	95.8%	94.2%	18.9	18.3
	1000	96.4%	93.2%	15.3	14.9
exponential(1)	300	95.3%	96.8%	3.31	2.73
	1000	94.6%	96.9%	3.11	2.50
Pareto(1)	300	96.2%	97.5%	28.4	14.6
	1000	95.1%	96.0%	22.8	11.2
chi-square(5)	300	95.8%	94.3%	11.0	10.0
	1000	95.2%	92.6%	10.4	9.42

Table 2: 90% guaranteed coverage tolerance intervals with confidence level 95%.

These tables show very good behaviour of our tolerance intervals. In particular, for the highly skewed distributions they perform much better with respect to length; e.g., for the Pareto distribution the length is reduced with 50%. In general, we see that the asymptotic theory works well.

distribution	sample size	simulated coverage		average length	
		classical	new	classical	new
standard normal	300	90.0%	89.0%	3.31	3.24
	1000	90.0%	89.5%	3.29	3.26
standard Cauchy	300	90.1%	89.5%	13.6	12.7
	1000	90.0%	89.6%	12.9	12.4
exponential(1)	300	90.0%	90.0%	2.98	2.35
	1000	90.0%	90.0%	2.96	2.32
Pareto(1)	300	90.1%	90.1%	20.5	9.71
	1000	90.1%	90.0%	19.5	9.22
chi-square(5)	300	90.0%	89.4%	10.0	8.91
	1000	90.0%	89.7%	9.97	8.92

Table 3: 90% mean coverage tolerance intervals.

Table 4 gives simulation results for mean coverage rectangles with sides parallel to the coordinate axes. We included 4 extra observations in all cases, i.e. we used 274 observations for $n = 300$ and 904 for $n = 1000$. We simulated from the following distributions:

- bivariate standard normal with mean $\binom{0}{0}$ and covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- bivariate half-normal with density $f(x, y) = \frac{2}{\pi} e^{-\frac{1}{2}(x^2+y^2)}$, $x, y \geq 0$
- bivariate Cauchy distribution with density $f(x, y) = \frac{1}{2\pi} \left(1 + x^2 + y^2\right)^{-3/2}$
- bivariate exponential (1,1) distribution with density $f(x, y) = e^{-(x+y)}$, $x, y \geq 0$
- bivariate pyramid distribution with density $f(x, y) = \frac{1}{8(|x| \vee |y|)} e^{-(|x| \vee |y|)}$; see Figure 2 below.

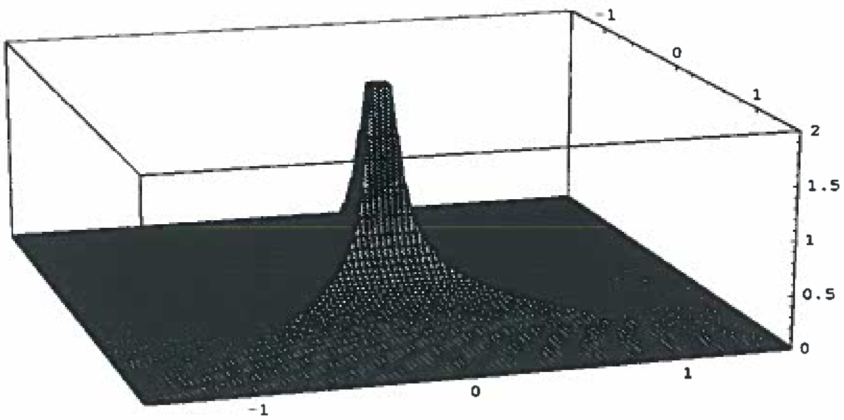

Figure 2: Bivariate pyramid density.

From this table, we again see that our tolerance regions perform well: the coverages are close to 90%, but slightly too low. This effect is caused by the minimum area property of our tolerance

distribution	sample size	
	300	1000
bivariate normal	87.7%	88.7%
bivariate half-normal	88.3%	88.9%
bivariate Cauchy	86.2%	86.3%
bivariate exponential	88.5%	89.0%
bivariate pyramid	86.4%	87.1%

Table 4: Simulated coverages of 90% mean coverage tolerance rectangles.

regions, and has a drastic impact on the confidence level of guaranteed coverage tolerance rectangles. Therefore, we do not present simulation results for those rectangles. However, a better performance of the mean coverage tolerance rectangles is possible by including more observations.

We have also performed simulations for tolerance hyperrectangles in $I\!R^3$, from the following trivariate distributions:

- trivariate standard normal with mean $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
- trivariate half-normal with density $f(x,y,z) = \left(\frac{2}{\pi}\right)^{3/2} e^{-\frac{1}{2}(x^2+y^2+z^2)}$, $x, y, z \geq 0$
- trivariate Cauchy distribution with density $f(x,y,z) = \frac{1}{\pi^2}\left(1 + x^2 + y^2 + z^2\right)^{-2}$
- trivariate exponential distribution with density $f(x,y,z) = e^{-(x+y+z)}$, $x, y, z \geq 0$

In Table 5 simulation results for the mean coverage hyperrectangles for $n = 300$ are presented. Here we included 6 extra points. Hence for the 95% mean coverage tolerance regions 291 data points were included. As is clear from this table the results are again very good. Replacing 90% (Table 4) by 95% seems to improve the asymptotics, as could be expected. We chose 95% here, not to improve on the coverage, but to speed up the computations; now the number of points that have to be excluded is substantially less (9 against 24).

distribution	simulated coverage
trivariate normal	93.6%
trivariate half-normal	94.1%
trivariate Cauchy	94.8%
trivariate exponential	94.2%

Table 5: Simulated coverages of 95% mean coverage tolerance hyperrectangles.

We give a brief description of the algorithm that was used for computing the minimum area parallel rectangles, which led to Table 4. This algorithm can be easily extended to one for minimum volume hyperrectangles; this was used for Table 5. The basic idea is that since tolerance regions typically have a coverage of 90% or 95%, it is the outermost points that determine the minimum area rectangle. As we have to find the smallest rectangle over $\lceil np_n \rceil + 4$ observations from $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, we 'peel' our data $r + 1$ times, where

$r := n - \lceil np_n \rceil - 4$. Each peeling consists in removing the boundary observations of the smallest rectangle over the remaining observations. Save the peeled observations in the set $\mathcal{P}$ and let ℓ be the cardinality of $\mathcal{P}$. Denote the order statistics of the y-coordinates of the elements of $\mathcal{P}$ by $Y_{j:\ell}$, $j = 1, \dots, \ell$. The horizontal sides of the minimum area rectangle can lie only on the lines $y = Y_{j:\ell}$ and $y = Y_{\ell-r+i-1:\ell}$, where $j = 1, \dots, r+1$ and $i = j, \dots, r+1$. For each fixed i and j, there are $n - r - j + i$ sample points (including two observations on the horizontal sides) between $y = Y_{j:\ell}$ and $y = Y_{\ell-r+i-1:\ell}$. Hence, for each i and j we can construct $i - j + 1$ rectangles containing exactly $\lceil np_n \rceil + 4$ points such that horizontal sides lie on $y = Y_{j:\ell}$ and $y = Y_{\ell-r+i-1:\ell}$, and the vertical sides each contain an observation. Find the minimum area rectangle R_{ij} in this set. The final minimum area rectangle is the minimum area rectangle among the R_{ij}'s.

Given the discrete nature of the empirical measure and the aforementioned sensitivity of tolerance regions it can be, in particular when the density f is smooth, that a smoothed version of the empirical measure yields somewhat better tolerance regions than the ones presented in Section 3. We will briefly consider this here and will restrict ourselves to the one dimensional situation and guaranteed coverage tolerance intervals. It can be shown, see e.g. Azzalini (1981), Shorack and Wellner (1986), Section 23.2, and van der Vaart (1994), that an integrated kernel density estimator ($\hat{P}_n$, say) as an estimator for the probability measure yields the same limiting behaviour as in Section 3, when the bandwidth is chosen to be $K/n^{1/3}$, $K \in (0, \infty)$. So asymptotically, in first order, there is no difference between the two procedures, i.e. Theorem 2 holds true, when $A_{n,t_0,C}$ is based on $\hat{P}_n$ instead of on P_n. However, for finite n it may be that a 'smoothed procedure' works better. We investigated this through a simulation. Table 6 gives the results. We chose the Epanechnikov kernel (with support $[-1, 1]$) and $K = \frac{1}{2}\sqrt{5}S$, with S the sample standard deviation, as suggested in Azzalini (1981). Since $\hat{P}_n$ is absolutely continuous we did not add the 2 observations as indicated above.

distribution	sample size	simulated conf. level	average length
standard normal	300	92.6%	3.58
	1000	92.7%	3.44
chi-square(5)	300	96.4%	9.98
	1000	96.8%	9.50
beta(5,10)	300	94.5%	.415
	1000	94.3%	.400
logistic	300	93.4%	6.51
	1000	93.1%	6.23
Student-t(5)	300	93.6%	4.52
	1000	92.7%	4.28

Table 6: 'Smoothed' 90% guaranteed coverage tolerance intervals with confidence level 95%.

This table shows excellent behaviour of the 'smoothed' tolerance intervals. We see indeed that there is some evidence that, when the underlying density is smooth, our procedures can be somewhat improved by properly smoothing the empirical.

All simulations were performed on a SunSparc5 and SunUltra10. Simulations in dimensions one and three were performed using the statistical packages of the computer algebra system Mathematica. The (two-dimensional) rectangles algorithm was implemented in C++, which was linked with a Mathematica notebook where data were generated and coverages were computed. The computation for one replication (including the coverage computation) with $n = 1000$ took at most 6 seconds. Our simulations procedures for parallel hyperrectangles can easily be extended to dimensions 4 and higher.

As mentioned before medical statistics is one of the fields where tolerance regions are used. Here we illustrate our theory with an application to Leukemia diagnosis. Leukemia is a cancer of blood-forming tissue such as bone marrow. The diagnosis of Leukemia is based on the results of both blood and bone marrow tests. There are only three major types of blood cells: red blood cells, white blood cells and platelets. These cells are produced in the bone marrow and circulate through the blood stream in a liquid called plasma. When the bone marrow is functioning normally the count of blood cells remains stable. In the case of this disease the number of blood cells changes drastically and is therefore easy to detect with tolerance regions. We now construct a 95% mean coverage tolerance ellipse and two 95% mean coverage tolerance (hyper)rectangles (for dimension $k = 2$ and $k = 3$) for blood count data kindly provided by Blood bank de Meierij, Eindhoven. Blood samples were taken from 1000 adult, supposedly healthy potential blood donors. Among the measured variables were the total number of white blood cells (WBC), red blood cells (RBC), and platelets (PLT) in one nanoliter, picoliter, and nanoliter, respectively, of whole blood. We computed tolerance regions (ellipse, rectangle, hyperrectangle) for the following combinations of variables: (WBC, PLT), (WBC, RBC) and (WBC, RBC, PLT), for 500, 1000 and 500 observations, respectively (see Figures 3, 4 and 5 below).

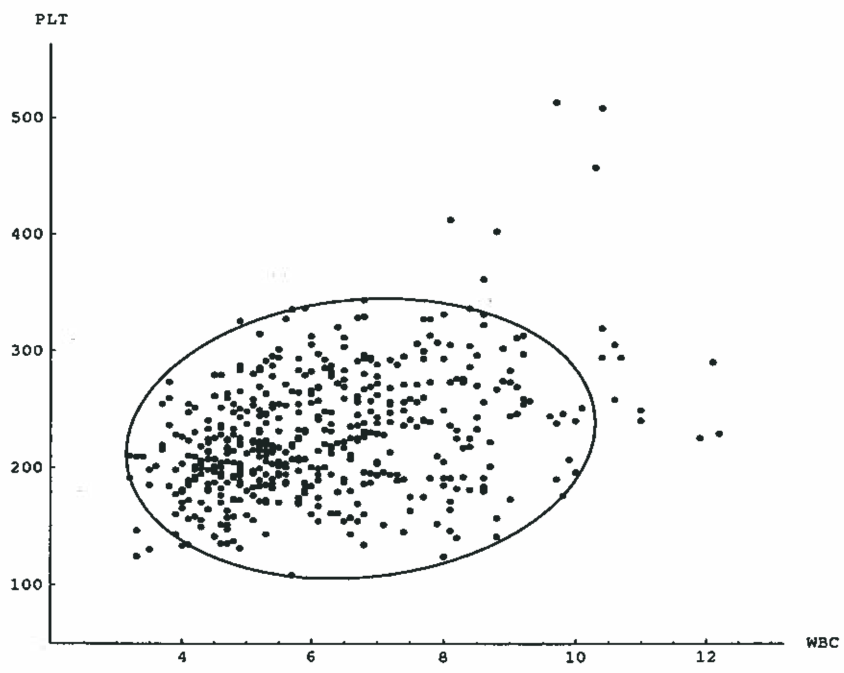

Figure 3: 95% mean coverage tolerance ellipse.

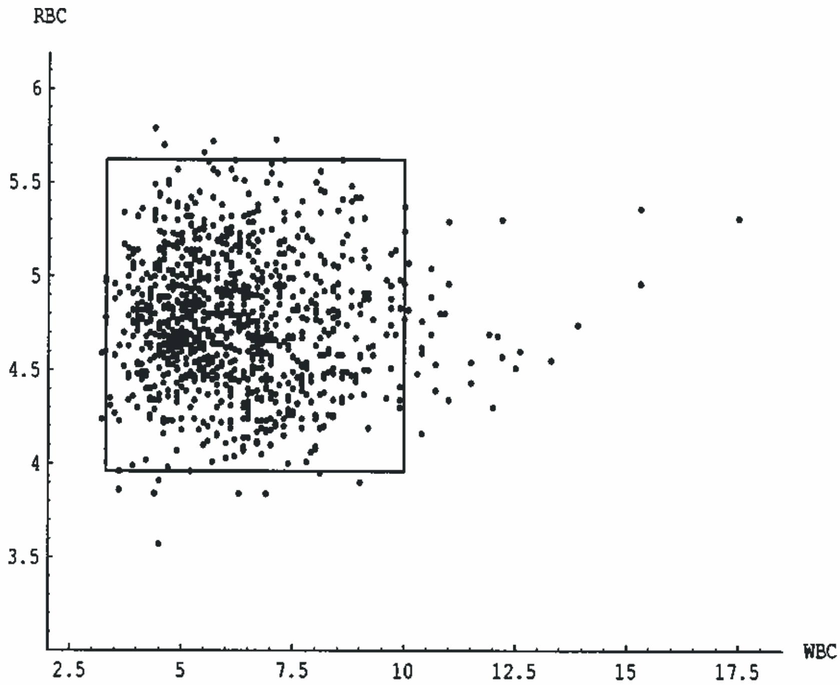

Figure 4: 95% mean coverage tolerance rectangle.

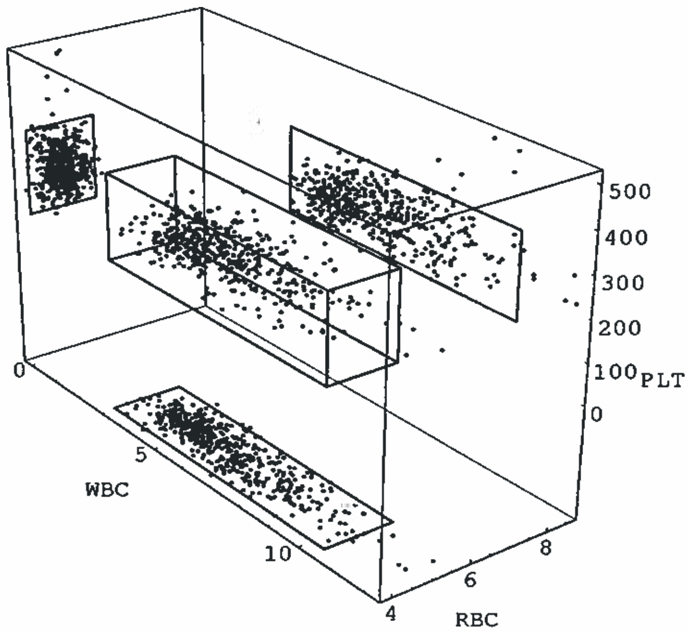

Figure 5: 95% mean coverage tolerance hyperrectangle.

Comparing the tolerance regions in Figures 3, 4 and 5 with the in practice used one-dimensional 'reference' or 'normal' values for WBC, RBC, and PLT (which we do not record here), it can be seen that our procedures work nicely. Due to the fact that the one-dimensional distributions of WBC and PLT are somewhat skewed to the right our procedures tend to give smaller regions (when these variables are involved), than those constructed (in one way or another) from the one-dimensional reference values. This is the same effect as seen in Tables 2 and 3 for the skewed distributions there. Moreover, our tolerance regions are somewhat shifted to the 'left' because of this skewness of the distributions of these variables. It is obvious, but it can be important, that in Figure 3, the tolerance ellipse does not include certain bivariate values, which would be included when forming two intervals by projecting the ellipse on the horizontal and vertical axes. For Acute Leukemia, newly diagnosed, adult patients very often have WBC values considerably over 10 (in many cases even above 100(!)) or RBC values around 3 or PLT values below 100. Clearly these values can be easily detected by the depicted tolerance regions.

Finally we give some references on computing minimum volume ellipsoids and minimum area planar convex sets (which we did not compute in this section). An algorithm for computing the minimum volume ellipsoid containing *all* data points is presented in Silverman and Titterington (1980). Algorithms for computing *approximate* minimum area ellipsoids containing m ($< n$) points are given in Nolan (1991) and Rousseeuw and van Zomeren (1991) and the *exact* algorithm we used for the minimum volume ellipse containing m ($< n$) points was developed in Agulló (1996). The computer code of this algorithm was kindly placed to our disposal by the author; it also works in higher dimensions (up to 10). As we noted in Section 2, the minimum area planar convex set containing m ($< n$) sample points is a polygon. *Exact* algorithms for computing such sets can be found in Eppstein et al. (1992) and Eppstein (1992).

5 Proofs

Here we present the proofs of the theorems of Section 3.

Proof of Theorem 1 For each $n \geq 1$, define the empirical process indexed by $\mathcal{A}$ to be

$$\alpha_n(A) = \sqrt{n}(P_n(A) - P(A)), \quad A \in \mathcal{A}.$$

Because of C1) and the Skorohod-Dudley-Wichura representation theorem (see e.g., Gaenssler (1983), p. 82), there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{I\!P})$ carrying a version $\widetilde{B}_P$ of B_P and versions $\widetilde{\alpha}_n$ of α_n, for all $n \in I\!N$, such that

$$\sup_{A \in \mathcal{A}} |\widetilde{\alpha}_n(A) - \widetilde{B}_P(A)| \to 0 \text{ a.s.}, \quad n \to \infty. \tag{6}$$

Henceforth, we will drop the tildes from the notation, for notational convenience. By C2) we obtain

$$\sqrt{n}(P_n(A_{n,t_0,C}) - P(A_{n,t_0,C})) - B_P(A_{n,t_0,C}) \to 0 \text{ a.s.}, \quad n \to \infty. \tag{7}$$

Combining this with C3) yields

$$\sqrt{n}(t_0 - P(A_{n,t_0,C})) + C - B_P(A_{n,t_0,C}) \to 0 \text{ a.s.}, \quad n \to \infty. \tag{8}$$

From C4) we have that $d_0(A_{n,t_0,C}, A_{t_0}) \to 0$ a.s. and hence, since B_P is continuous with respect to d_0,

$$B_P(A_{n,t_0,C}) \to B_P(A_{t_0}) \text{ a.s.}, \quad n \to \infty. \tag{9}$$

From (8) and (9) we now obtain that

$$\sqrt{n}(t_0 - P(A_{n,t_0,C})) + C \to B_P(A_{t_0}) \text{ a.s.}, \quad n \to \infty.$$

Observing that

$$B_P(A_{t_0}) \stackrel{d}{=} Z\sqrt{t_0(1-t_0)},$$

completes the proof. □

Proof of Theorem 2 By Theorem 1, for all $x \in \mathbb{R}$, we have

$$\mathbb{P}\{\sqrt{n}(t_0 - P(A_{n,t_0,C})) + C \le x\} \to \mathbb{P}\{Z\sqrt{t_0(1-t_0)} \le x\}, \quad n \to \infty.$$

Hence, taking $x = C$, we obtain

$$\lim_{n\to\infty} \mathbb{P}\{P(A_{n,t_0,C}) \ge t_0\} = \mathbb{P}\{Z\sqrt{t_0(1-t_0)} \le C\} = 1 - \alpha.$$

□

Proof of Theorem 3 Theorem 1 with $C = 0$ yields

$$\sqrt{n}(t_0 - P(A_{n,t_0,0})) \stackrel{d}{\longrightarrow} Z\sqrt{t_0(1-t_0)}, \; n \to \infty. \tag{10}$$

By assumption $\sqrt{n}(t_0 - P(A_{n,t_0,0}))$ is uniformly integrable, hence

$$\mathbb{E}\sqrt{n}(t_0 - P(A_{n,t_0,0})) \longrightarrow \mathbb{E}(Z\sqrt{t_0(1-t_0)}) = 0, \quad n \to \infty,$$

which is the statement of the theorem. □

We next present two lemmas. Lemma 2 is crucial for the proof of Theorem 4, whereas Lemma 1 is needed for the proof of Lemma 2. Until further notice we shall, for case (c), tacitly *restrict* ourselves to those closed convex sets that are contained in some large circle B (which will be specified later on). In the proof of Theorem 4 we will show that this restriction can be removed. For Lemma 1, recall the functions U and $\widetilde{F}$, defined in Section 2.

Lemma 5.1 *Under the assumptions of Theorem 4 we have for the cases (a), (b) and (c), that the functions U and $\widetilde{F}$ are inverses of each other. Hence, U is continuous on $(0,1)$, $\widetilde{F}$ is continuous on $\mathbb{R}^+$, and they are both strictly increasing.*

Proof We first prove the continuity of U. Note that absolute continuity of P implies that

$$U(t) = \inf_{A\in\mathcal{A}}\{V(A) : \; P(A) > t\}, \text{ for any } t \in (0,1),$$

and

$$\widetilde{F}(y) = \sup_{A\in\mathcal{A}}\{P(A) : \; V(A) < y\}, \quad \text{for any } \; y \in \mathbb{R}^+.$$

Let us now take an arbitrary decreasing sequence $t_m \downarrow t$, where $t_m, t \in (0,1)$. Consider the sequence of sets

$$D_m = \{V(A): \; P(A) > t_m, \quad A \in \mathcal{A}\}.$$

It is easy to see that this is a nested sequence of sets, with limit set

$$\bigcup_{m=1}^{\infty} D_m = \{V(A): \; P(A) > t\}$$

and hence,

$$\lim_{m\to\infty} U(t_m) = \lim_{m\to\infty} \inf D_m = \inf_{A\in\mathcal{A}} \{V(A): \; P(A) > t\} = U(t).$$

In case $t_m \uparrow t$ the proof is analogous. Similar arguments yield continuity of $\widetilde{F}$.

Note that absolute continuity of P also implies that

$$U(t) = \inf_{A\in\mathcal{A}} \{V(A): \; P(A) = t\}, \text{ for any } t \in (0,1), \tag{11}$$

and

$$\widetilde{F}(y) = \sup_{A\in\mathcal{A}} \{P(A): \; V(A) = y\}, \quad \text{for any} \quad y \in I\!R^{+}. \tag{12}$$

It follows from (11) and (12) that U is the generalized inverse of $\widetilde{F}$, i.e.

$$U(t) = \inf\{y : \widetilde{F}(y) \geq t\} \quad \text{for any } t \in (0,1).$$

Hence, clearly both U and $\widetilde{F}$ are strictly increasing and continuous. Thus we conclude that they are inverses of each other. □

Lemma 5.2 *Under the assumptions of Theorem 4 we have for the cases (a), (b) and (c) that with probability one*

$$d(A_{n,t_0,C}, A_{t_0}) \to 0,$$

and hence $d_0(A_{n,t_0,C}, A_{t_0}) \to 0 \quad (n \to \infty)$.

Note that an in-probability-version of this lemma, with $k = 1$ and $C = 0$, can be found in Beirlant and Einmahl (1995), Corollary 1; see also Einmahl and Mason (1992).

Proof of Lemma 2 Since for cases (a) and (b) $\mathcal{A}$ is a Vapnik-Chervonenkis (VC) class we have that C1) of Theorem 1 holds. The (restricted) class of convex sets is not a VC class, but we still have C1), see Bolthausen (1978) and Dudley (1978), p. 918. Hence we have (6) for all three cases. Since B_P is bounded, this yields

$$\sup_{A\in\mathcal{A}} |P_n(A) - P(A)| \to 0 \;\text{ a.s.}, \quad n \to \infty. \tag{13}$$

It now trivially follows from (13) and the definitions of $\widetilde{F}_n$ and $\widetilde{F}$ that

$$\sup_{y>0} |\widetilde{F}_n(y) - \widetilde{F}(y)| \to 0 \;\text{ a.s.}, \quad n \to \infty. \tag{14}$$

Let $\ell < 1$ be arbitrary. Since $U(t)$ is continuous, increasing and nonnegative on $(0,1)$ by Lemma 5.1, it is uniformly continuous on $(0,\ell]$, and thus

$$\sup_{t\in(0,\ell]} \left| U\left(t + \frac{C}{\sqrt{n}}\right) - U(t) \right| \to 0, \quad n \to \infty. \tag{15}$$

We now want to prove that

$$\sup_{t\in(0,\ell]} |U_n(t) - U(t)| \to 0, \quad n \to \infty. \tag{16}$$

For any $\varepsilon > 0$ we have from (14) that for n large enough

$$\widetilde{F}(y) - \varepsilon \le \widetilde{F}_n(y) < \widetilde{F}(y) + \varepsilon \quad \text{for all} \quad y > 0 \quad \text{a.s.}\ .$$

By Lemma 5.1, U is the generalized inverse of $\widetilde{F}$. It is easy to see that U_n and $\widetilde{F}_n$ are generalized inverses. Hence, we obtain from the above inequalities that

$$U(t-\varepsilon) \le U_n(t) \le U(t+\varepsilon) \text{ for all } \ t \in (0,1) \quad \text{a.s.}\ .$$

Since U is uniformly continuous, there exists $\delta > 0$ such that

$$U(t) - \delta \le U(t-\varepsilon) \le U_n(t) \le U(t+\varepsilon) \le U(t) + \delta \quad \text{for any} \ \ t \in (0,\ell] \ \ \text{a.s.},$$

which immediately yields (16). From (15) and (16) it follows that

$$\sup_{t\in(0,\ell]} \left| U_n\left(t + \frac{C}{\sqrt{n}}\right) - U(t) \right| \to 0 \ \ \text{a.s.}, \quad n \to \infty. \tag{17}$$

Now let us return to the sets given in the statement of the lemma:

- $A_{n,t_0,C}$, the a.s. unique smallest element of $\mathcal{A}$ with $P_n(A_{n,t_0,C}) \ge t_0 + \frac{C}{\sqrt{n}}$ (and hence, $V(A_{n,t_0,C}) = U_n(t_0 + \frac{C}{\sqrt{n}})$),
- A_{t_0}, the unique smallest element of $\mathcal{A}$ with $P(A_{t_0}) = t_0$ (and $V(A_{t_0}) = U(t_0)$).

By (1),

$$P_n(A_{n,t_0,C}) \to t_0 \ \ \text{a.s.}, \quad n \to \infty,$$

and thus by (13)

$$P(A_{n,t_0,C}) \to t_0, \quad \text{a.s.}, \quad n \to \infty.$$

From (17) we have

$$\lim_{n\to\infty} V(A_{n,t_0,C}) = V(A_{t_0}) \ \ \text{a.s.}\ .$$

The sequence $\{A_{n,t_0,C}\}_{n\ge1}$ is uniformly bounded a.s., i.e. for each $\omega \in \Omega_0$, with $\mathbb{P}(\Omega_0) = 1$, there exists a compact set $\mathcal{M}_\omega$, that contains all the $A_{n,t_0,C}$'s. By the Blaschke Selection Principle the sequence $\{A_{n,t_0,C}\}_{n\ge1}$ has at least one limit set. So there exists a subsequence $\{A_{n_k,t_0,C}\}_{k\ge1}$ and a non-empty closed convex set A^* (an element of the indexing class (a), (b) or (c), respectively), such that

$$\lim_{k\to\infty} V(A_{n_k,t_0,C} \,\triangle\, A^*) = 0 \ \ \text{a.s.}\ .$$

Hence, $V(A_{n_k,t_0,C}) \to V(A^*)$, and thus $V(A^*) = U(t_0)$ a.s. Using that P is absolutely continuous with respect to Lebesgue measure, it is easy to see that $P(A^*) = t_0$.

So we have for the limit set A^* that

$$V(A^*) = U(t_0) \text{ and } P(A^*) = t_0 \text{ a.s.,}$$

but by assumption there exists a unique set A_{t_0} satisfying these two equations. Hence, any limit set A^* of the sequence $\{A_{n,t_0,C}\}_{n\geq 1}$ is equal to A_{t_0}, and thus the sequence itself converges to A_{t_0} (a.s.). □

Proof of Theorem 4 We will check the conditions C1)-C4) of Theorem 1. We first prove (3) and (4), for the cases (a), (b) and the restricted case (c). As noted in the proof of Lemma 2 we have that C1) holds. In Section 2 it is shown that C2) holds; C3) follows from (1). The first part of C4) is an assumption of Theorem 4; Lemma 2 yields the second part of condition C4). This completes the proof of (3) and (4) for these cases.

Now consider the unrestricted case (c). We will prove (3) and (4). Let us first construct a proper circle B, as used in the definition of the restricted class. Let B_{t_0} be a circle with radius r, say, such that $A_{t_0} \subset B_{t_0}$ and $P(B_{t_0}) > t_0 \vee (1-t_0)$. For sake of notation, any space V_γ between two parallel lines in $\mathbb{R}^2$ at distance γ is said to be a γ-strip. Note that for a probability measure P with density f we have that

$$\lim_{\gamma\to 0} \sup_{V_\gamma} P(V_\gamma) = 0,$$

where each supremum runs over all γ-strips. Therefore there exists a γ_0 satisfying the inequality

$$\sup_{V_{\gamma_0}} P(V_{\gamma_0}) \leq \tfrac{1}{2}\, t_0, \tag{18}$$

where the supremum runs over all γ_0-strips. Now choose B to be a circle with the same centre as B_{t_0}, but with radius $R > \frac{8}{\gamma_0} U(t_0) + r$, where γ_0 satisfies (18).

Next we show that $A_{n,t_0,C} = A^*_{n,t_0,C}$ for large n a.s., where $A^*_{n,t_0,C}$ is defined similarly as $A_{n,t_0,C}$ but for the restricted class. In other words we have to show that for n large enough $A_{n,t_0,C} \subset B$ almost surely. Observe that

$$\begin{aligned} \lim_{n\to\infty} P_n(A_{n,t_0,C}) &= t_0 \text{ a.s.,} \\ \lim_{n\to\infty} P_n(B^c_{t_0}) = P(B^c_{t_0}) &< t_0 \wedge (1-t_0) \text{ a.s. .} \end{aligned}$$

So, if there exists with positive probability a subsequence $\{A_{n_k,t_0,C}\}_{k\geq 1}$ such that $A_{n_k,t_0,C} \not\subset B$ for all k, then $A_{n_k,t_0,C}$ contains an element of B_{t_0} as well as an element of B^c eventually. Because the γ_0-strips form a VC class, we have that

$$\lim_{n\to\infty} \sup_{V_{\gamma_0}} P_n(V_{\gamma_0}) \leq \tfrac{1}{2}\, t_0 \text{ a.s. .}$$

Hence, $A_{n_k,t_0,C}$ eventually contains a triangle with area $\frac{\gamma_0}{4}(R-r) > 2U(t_0)$. However, this can not happen because of the Glivenko-Cantelli theorem. This proves (3) and hence (4).

Finally we prove (5) for all three cases. It suffices to show that $\sqrt{n}(t_0 - P(A_{n,t_0,0}))$ is uniformly integrable. It follows from (1) that

$$\begin{aligned}|\sqrt{n}(t_0 - P(A_{n,t_0,0}))| &\leq |\sqrt{n}(P_n(A_{n,t_0,0}) - P(A_{n,t_0,0}))| + |\sqrt{n}(t_0 - P_n(A_{n,t_0,0}))| \\ &\leq \sup_{A\in\mathcal{A}} |\sqrt{n}(P_n(A) - P(A))| + 1 \quad \text{a.s.}\,.\end{aligned}$$

Therefore it suffices to establish uniform integrability of

$$Y_n := \sup_{A\in\mathcal{A}} |\sqrt{n}(P_n(A) - P(A))|.$$

Note that if Y is a non-negative random variable then

$$\mathbb{E}Y = \int_0^\infty \mathbb{P}\{Y > y\}dy.$$

Hence,

$$\mathbb{E}YI_{[Y>a]} = \int_0^\infty \mathbb{P}\{YI_{[Y>a]} > y\}dy = a\mathbb{P}\{Y > a\} + \int_a^\infty \mathbb{P}\{Y > y\}dy.$$

Moreover, for the cases (a) and (b) (as then $\mathcal{A}$ is a VC class), using Theorem 2.11 of Alexander (1984), we have for $\lambda \geq 8$ and $C_1, C_2 \in (0,\infty)$ that

$$\mathbb{P}\{\sup_{A\in\mathcal{A}} |\sqrt{n}(P_n(A) - P(A))| > \lambda\} \leq C_1\lambda^{C_2}\exp(-2\lambda^2). \tag{19}$$

For large enough λ, the right-hand side of (19) is less than $\exp(-\lambda^2)$. Let $\varepsilon > 0$. Then for a large enough:

$$\mathbb{E}Y_nI_{[Y_n>a]} = a\mathbb{P}\{Y_n > a\} + \int_a^\infty \mathbb{P}\{Y_n > y\}dy \leq ae^{-a^2} + \int_a^\infty e^{-y^2}dy < \varepsilon.$$

In case (c), using Corollary 2.4 and Example 3 (p. 1045) of Alexander (1984) with $\psi = \psi_3$, we obtain the uniform integrability similarly as above; see also van der Vaart (1996), p. 2134. □

Recall the notation of Section 2, in particular let $X_1, \ldots, X_n$ and $\mathcal{E}$ be as in that section. Denote with $E_1 \in \mathcal{E}$ the almost surely unique ellipsoid of minimum volume containing *at least* $m \in \{k+1, \ldots, n\}$ (data) points.

Lemma 5.3 *E_1 contains exactly m points, almost surely.*

Proof Assume that E_1 contains $\ell > m$ points and t ($k+1 \leq t \leq k(k+3)/2$ a.s.) of these points are on its boundary. Note that the smallest ellipsoid containing these t boundary points is equal to E_1, see Silverman and Titterington (1980). Consider $t-1$ of the t boundary points (call this set B) and let E_0 be the smallest ellipsoid containing B. Denote the remaining t-th boundary point of E_1 with Y_1. Observe that $Y_1 \notin E_0$. It follows from a conditioning argument that for any subset of size $r > 1$ of the n points, we have a.s. that none of the remaining $n-r$ points is on the boundary of the smallest ellipsoid containing these r points. This yields that a.s. $V(E_0) < V(E_1)$.

Note that the smallest ellipsoid containing a finite set is equal to the smallest ellipsoid containing the convex hull of that set. Denote with Y_0 a point on the boundary of E_0 such that the line through Y_0 and Y_1 intersects the convex hull of B and such that the open interval from Y_0 to Y_1 has an empty intersection with E_0. Set $Y_\lambda = (1-\lambda)Y_0 + \lambda Y_1$, $\lambda \in [0,1]$. Let C_λ be the convex hull of $B \cup \{Y_\lambda\}$. Note that for $\lambda < \lambda'$ we have that $C_\lambda \subset C_{\lambda'}$. Let E_λ be the smallest ellipsoid containing C_λ. So $V(E_\lambda) \leq V(E_{\lambda'})$ for $\lambda \leq \lambda'$.

From the Blaschke Selection Principle it follows that there exists a sequence λ_j (< 1), $j \in I\!N$, converging to 1 and such that

$$\lim_{j\to\infty} V(E_{\lambda_j} \Delta E^*) = 0$$

for some $E^* \in \mathcal{E}$. We have $V(E^*) \leq V(E_1)$, since $V(E_{\lambda_j}) \leq V(E_1)$, $j \in I\!N$. But $C_1 \subset E^*$, so $V(E_1) \leq V(E^*)$. Hence $V(E^*) = V(E_1)$ and E^* and E_1 both contain C_1. But, with probability 1, E_1 is unique, so $E^* = E_1$ and hence

$$\lim_{j\to\infty} V(E_{\lambda_j} \Delta E_1) = 0.$$

So there exists a large j (denote the corresponding λ_j with η) such that E_η contains all the $\ell - t$ points in the interior of E_1 and the points of B and does not contain the $n - \ell$ points in the complement of E_1. If $Y_1 \in E_\eta$, then Y_η is in the interior of E_η, so according to Silverman and Titterington (1980), $E_\eta = E_0$ and hence $V(E_\eta) = V(E_0) < V(E_1)$ a.s., but this can not happen since $C_1 \subset E_\eta$. This yields that $Y_1 \notin E_\eta$. We now see that E_η contains $\ell - 1$ $(\geq m)$ points and $V(E_\eta) \leq V(E_1)$. Since E_1 is the minimum volume ellipsoid containing at least m points, we have that $V(E_\eta) = V(E_1)$. Since $E_\eta \neq E_1$ this contradicts the a.s. uniqueness of the minimum volume ellipsoid. □

Acknowledgements We are grateful to an Associate Editor and a referee for a careful reading of the manuscript and insightful remarks which led to substantial improvements of the paper. Thanks are also due to Jose Agulló for kindly making available to us the computer code for the minimum volume ellipsoid algorithm, to Bart Hartgers for implementing our rectangles algorithm in C++ and linking it to our Mathematica notebook, and to Marko Boon for help with Figures 3 and 5. Finally, we want to express our gratitude to Blood bank de Meierij, Eindhoven, in particular to Harry von Hegedus, Paul van Noord, Anja Verheijden, and Anouk Wesselink, for generously providing us with medical expertise and the blood count data set.

References

Ackermann, H. (1983). Multivariate nonparametric tolerance regions: a new construction technique. *Biometrical J.*, **25**, 351–359.

Agulló, J. (1996). Exact iterative computation of the multivariate minimum volume ellipsoid estimator with a branch and bound algorithm. In A. Prat (editor), *COMPSTAT '96 Barcelona*, pages 175–180, Heidelberg. Physica Verlag.

Aitchison, J., and I. R. Dunsmore (1975). *Statistical prediction analysis*. Cambridge University Press.

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, **12**, 1041–1067.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.

Beirlant, J., and J. H. J. Einmahl (1995). Asymptotic confidence intervals for the length of the shortt under random censoring. *Statist. Neerlandica*, **49**, 1–8.

Bolthausen, E. (1978). Weak convergence of an empirical process indexed by the closed convex subsets of I^2. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **43**, 173–181.

Carroll, R. J., and D. Ruppert (1991). Prediction and tolerance intervals with transformation and/or weighting. *Technometrics*, **33**, 197–210.

Chatterjee, S. K., and N. K. Patra (1980). Asymptotically minimal multivariate tolerance sets. *Calcutta Statist. Assoc. Bull.*, **29**, 73–93.

Davies, L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.*, **20**, 1828–1843.

Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.*, **6**, 899–929.

Dudley, R. M. (1982). Empirical and Poisson processes on classes of sets or functions too large for central limit theorems. *Z. Wahrsch. Verw. Gebiete*, **61**, 355–368.

Einmahl, J. H. J., and D. M. Mason (1992). Generalized quantile processes. *Ann. Statist.*, **20**, 1062–1078.

Eppstein, D. (1992). New algorithms for minimum area k-gons. In *Proceedings of the Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 83–88, New York. ACM.

Eppstein, D., M. Overmars, G. Rote, and G. Woeginger (1992). Finding minimum area k-gons. *Discrete Comput. Geom.*, **7**, 45–58.

Fraser, D. A. S. (1953). Nonparametric tolerance regions. *Ann. Math. Stat.*, **24**, 44–55.

Gaenssler, P. (1983). *Empirical Processes*. IMS, Hayward, California.

Grübel, R. (1988). The length of the shorth. *Ann. Statist.*, **16**, 619–628.

Guttman, I. (1970). *Statistical tolerance regions: classical and Bayesian*. Charles Griffin, London.

Jílek, M. (1981). A bibliography of statistical tolerance regions. *Math. Operationsforsch. Statist. Ser. Statist.*, **12**, 441–456.

Jílek, M., and H. Ackermann (1989). A bibliography of statistical tolerance regions. II. *Statistics*, **20**, 165–172.

Nolan, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.*, **39**, 348–371.

Rousseeuw, P., and A. Leroy (1988). A robust scale estimator based on the shortest half. *Statist. Neerlandica*, **42**, 103–116.

Rousseeuw, P., and B. C. van Zomeren (1991). Robust distances: simulations and cutoff values. In W. Stahel and S. Weisberg (editors), *Directions in Robust Statistics and Diagnostics, part II*, volume 34 of *The IMA Volumes in Mathematics and Its Applications*, pages 195–203. Springer Verlag, New York.

Shephard, G. C., and R. J. Webster (1965). Metric sets of convex bodies. *Mathematika*, **12**, 73–88.

Shorack, G. R., and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Silverman, B. W., and D. M. Titterington (1980). Minimum covering ellipses. *SIAM J. Sci. Stat. Comput.*, **1**, 401–409.

Tukey, J. W. (1947). Nonparametric estimation, II. Statistical equivalent blocks and tolerance regions - the continuous case. *Ann. Math. Stat.*, **18**, 529–539.

Tukey, J. W. (1948). Nonparametric estimation, III. Statistical equivalent blocks and multivariate tolerance regions - the discontinuous case. *Ann. Math. Stat.*, **19**, 30–39.

Vaart, A. W. van der (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.*, **4**, 501–504.

Vaart, A. W. van der (1996). New Donsker classes. *Ann. Probab.*, **24**, 2128–2140.

Wald, A. (1943). An extension of Wilks' method for setting tolerance limits. *Ann. Math. Stat.*, **14**, 45–55.

Webster, R. (1994). *Convexity*. Oxford University Press.

Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *Ann. Math. Statist.*, **12**, 91–96.