The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1887–1896

A GENERAL QUALITATIVE DEFINITION OF ROBUSTNESS[1]

BY FRANK R. HAMPEL

University of California, Berkeley, and University of Zurich[2]

Two very closely related definitions of robustness of a sequence of estimators are given which take into account the types of deviations from parametric models that occur in practice. These definitions utilize the properties of the Prokhorov distance between probability distributions. It is proved that weak*-continuous functionals on the space of probability distributions define robust sequences of estimators (in either sense). The concept of the "breakdown point" of a sequence of estimators is defined, and some examples are given.

1. Introduction and motivation. The setup of robust estimation can be described as follows (compare Hampel (1968); for the background, see Tukey (1960), Huber (1964), (1968a), among others). We assume that the process generating the observations under consideration can approximately be described by some parametric model (e.g. the model of independently and identically normally distributed observations), and we want to estimate the parameters of this model (or some function of them), i.e. we want to find a statistic whose distribution is close to these parameter values. However, we know that the parametric model is not quite true; and therefore we require that the distribution of the estimator changes only slightly if the distribution of the observations is slightly altered from that of the strict parametric model with certain parameter values. (In order that this be possible one has to presuppose a certain smoothness or "robustness" of the parametric model itself.) We now have to specify the types of deviations to be allowed for. One may try to distinguish three main reasons for deviations from the parametric model: (i) rounding of the observations; (ii) the occurrence of gross errors; (iii) the model itself may only be an approximation to the underlying chance mechanism, e.g. by virtue of the central limit theorem. It turns out that a literal quantitative description of (i) and (ii) is given by the Prokhorov distance between probability distributions, and this distance, leading exactly to weak*-convergence, also takes care of (iii). Moreover, the Prokhorov distance can be defined on very general spaces. It seems reasonable also to describe the differences between the distributions of the estimators by this distance, as it corresponds roughly to what can be detected in practice (even better than the Lévy distance). Thus one is led to require that the distribution of the estimator be a continuous functional (with respect to Prokhorov distance) of the

Received December 22, 1970.

[1] Revised first part of author's dissertation, which had been written with partial support of NSF Grant GP–7454. Manuscript completed under ONR contract No. N00014–67–0151–0017.

[2] 1970/71 at Princeton University.

underlying distribution "at" the parametric model. However, if we talk about an "estimator," we usually have in mind a whole sequence of estimators, with the number of observations n tending to infinity; and it may be (e.g. as with the arithmetic mean) that for increasing n the true underlying distribution has to be closer and closer to the parametric model in order to keep the distribution of the estimator near that which it would have under this model. Such a sequence of estimators would show quite a bad behavior for larger n, and we therefore require in addition that the continuity be uniform in n. On the other hand, this is all one can require, since (trivial and artificial cases excepted) there will always occur a nonzero asymptotic bias for any sequence of estimators in every neighborhood of any distribution in the parametric model. Thus we make this our formal definition of robustness. Later on, we shall give a very closely related, but slightly stronger definition of robustness, which is based on the Prokhorov distance in the n-fold Cartesian product of the sample space with itself and which therefore allows also for some weak dependence between the observations and for slight changes of the underlying distribution from observation to observation. The sufficient conditions for robustness that are derived are the same for both definitions; essentially it is shown that a weak*-continuous functional on the space of distributions defines a robust sequence of estimators. Either of the two definitions leads to a simple dichotomic classification of sequences of estimators which is widely applicable. After defining the concept of the "breakdown point," which might prove useful in practice, we give some examples of estimators (mainly for a location parameter). For a more detailed investigation of robustness, one ought to consider a quantitative theory; however, there it seems advisable to separate several quantitative aspects of robustness, so that the theory becomes more complicated. Therefore, a quantitative theory is not treated in this paper. For a more extensive discussion, see Hampel (1968).

2. Definition and some properties of the Prokhorov distance. Let $(\Omega, \mathscr{A})$ be a measurable space such that $\Omega = \{\omega, \cdots\}$ is a complete separable metric space and $\mathscr{A}$ denotes the σ-algebra generated by the topology. For $A \subset \Omega$, $A \in \mathscr{A}$ let A^ε denote the set of all points whose distance from A (i.e., from at least one point in A) is less than ε. Let P and Q be two probability measures (or, more generally, two finite measures) on $(\Omega, \mathscr{A})$. Then their Prokhorov distance $\pi(P, Q)$ is defined by $\pi(P, Q) = \inf\{\varepsilon : P(A) \leqq Q(A^\varepsilon) + \varepsilon$ and $Q(A) \leqq P(A^\varepsilon) + \varepsilon$ for all $A \in \mathscr{A}\}$. An equivalent definition requires the two sets of inequalities to hold only for all closed sets A. Moreover, if the two measures have the same total mass (i.e., if $P(\Omega) = Q(\Omega)$), as is the case for probability measures, then the second set of inequalities is implied by the first one and hence can be dropped. It is easy to verify that the Prokhorov distance actually has the properties of a distance taking values between 0 and 1. Prokhorov (1956) has shown that the topology induced by this metric is the one of weak* convergence in the sense of convergence of the integrals of all bounded continuous functions.

Let $\Omega \times \Omega$ be the Cartesian product of Ω with itself, with the distance between two points defined as the maximum of the distances of their respective coordinates, and with the product topology and product σ-algebra induced by this metric. For a given ε let $D(\varepsilon)$ be the set of all points in $\Omega \times \Omega$ whose distance from the diagonal is less or equal to ε. Let P and Q be two probability measures on Ω. Then Strassen (1965) (Corollary to Theorem 11) has shown that $\pi(P, Q) \leqq \varepsilon$ if and only if there exists a probability measure R on $\Omega \times \Omega$ with marginals P and Q such that $R\{D(\varepsilon)\} \geqq 1-\varepsilon$.

If Ω is the real line, then the Lévy distance $\lambda(P, Q)$ between two probability distributions is obtained from the above definition of the Prokhorov distance by simply restricting the sets A to be intervals of the form $(-\infty, x]$ (resp. $[x, \infty)$). From this it follows that always $\lambda(P, Q) \leqq \pi(P, Q)$; in general this inequality is strict. Convergence in Lévy distance is equivalent to convergence in Prokhorov distance. Furthermore, as to the Kolmogorov distance $\kappa(P, Q)$ (which is obtained from the definition of the Lévy distance by replacing A^ε by A), it is well known that always $\lambda(P, Q) \leqq \kappa(P, Q)$. For the Kolmogorov distance and Prokhorov distance no such inequality holds; however, convergence in Kolmogorov distance is a stronger property than convergence in Prokhorov (or Lévy) distance.

3. Sequences of estimators. Let $(\Omega, \mathscr{A})$ be defined as above. Let $\mathscr{F} = \{F, G, \cdots, P, Q, \cdots\}$ denote the set of all probability measures on $(\Omega, \mathscr{A})$, and for each $n \geqq 1$ let $\mathscr{F}_n = \{F_n, \cdots, P_n, \cdots\} \subset \mathscr{F}$ be the set of discrete probability measures whose atoms have probabilities equal to $1/n$ or to a multiple of $1/n$.

Obviously each finite sequence $\{\omega_1, \cdots, \omega_n\}$ of observations in Ω defines an $F_n \in \mathscr{F}_n$ in a natural way; and conversely, each F_n determines a sequence up to permutations. In considering only F_n we neglect the information possibly provided by the permutation (in accordance with the usual models of statistics, in which F_n is "sufficient" for the sequence) being well aware that in practice sometimes this information is very important (which means, of course, that in these cases the usually assumed model of independent identically distributed observations does not hold).

As above, $\pi(F, G)$ denotes the Prokhorov distance between F and G. Let R^k be the k-dimensional Euclidean space, with the distance between two points given by the maximal distance between respective coordinates. A sequence $\{T_n\}$ $(n \geqq 1)$ of estimators is now defined to be a sequence of measurable mappings $T_n : \mathscr{F}_n \to R^k$. (This definition includes simultaneous estimation of location and scale, of the cell means in some replicated experimental design, of some real-valued functional of a continuous stochastic process, etc.) In general, the value $T_n(F)$ (for $F \in \mathscr{F}_n$) depends not only on F but also on n; however, often n does not enter explicitly (as is the case for most estimators in the examples below), so that $F \in \mathscr{F}_n \cap \mathscr{F}_m$ implies $T_n(F) = T_m(F)$. An important special case occurs if a mapping T (with values in R^k) is defined for all $F \in \mathscr{F}$ and if T_n is simply the restriction of T to $\mathscr{F}_n$.

Let $\omega_1, \omega_2, \cdots$ be independently identically distributed according to some distribution F, and let F_n be the (random) measure corresponding to the first n observations. The mapping T_n from $\mathscr{F}_n$ to R^k thus induces a probability measure in R^k (the distribution of T_n under F) which will be denoted by $\mathscr{L}_F(T_n)$. F_n converges weakly to F (a.s.) (see, e.g., the book by K. Parthasarathy: *Probability Measures on Metric Spaces*), and it might be that $T_n(F_n)$ converges in probability to ("is consistent for") some value which will then be denoted by $T_\infty(F)$; in this case we might say that, under F, $\{T_n\}$ is a sequence of estimators for $T_\infty(F)$. (Note that even if $T_n \equiv T \mid \mathscr{F}_n$ for some T, $T_\infty(F)$ does not have to coincide with $T(F)$, although $T_\infty(F)$ would appear to be a more natural value of the mapping at F. If $T_\infty(F)$ exists for all F, we might define a new sequence of estimators $\{\tilde{T}_n\}$ based on the old one by putting $\tilde{T}_n \equiv T_\infty \mid \mathscr{F}_n$.)

4. Robust sequences.

DEFINITION. (i) A sequence of estimators $\{T_n\}$ is robust at a probability measure F iff

(A)
$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall G\, \forall n:$$
$$\{\pi(F, G\} < \delta \Rightarrow \pi(\mathscr{L}_F(T_n), \mathscr{L}_G(T_n)) < \varepsilon\}.$$

(ii) A sequence $\{T_n\}$ is robust in a neighborhood of F iff there is an $\eta > 0$ such that $\pi(F, G) < \eta$ implies $\{T_n\}$ is robust at G. A sequence $\{T_n\}$ is robust at a class $\mathscr{E} \subset \mathscr{F}$ iff it is robust at all $F \in \mathscr{E}$. A sequence $\{T_n\}$ is robust (everywhere) iff it is robust at all $F \in \mathscr{F}$.

The following definition is of a technical nature.

DEFINITION. Let a probability measure F and a sequence of estimators $\{T_n\}$ be given. Then condition (B) is fulfilled for F and $\{T_n\}$ iff

(B)
$$\forall \varepsilon > 0\, \forall \eta > 0\, \exists \delta > 0\, \forall n\, \exists \mathscr{E}_n \subset \mathscr{F}_n : \{F(\mathscr{E}_n)$$
$$> (1-\eta)\} \wedge \{F_n \in \mathscr{E}_n \wedge G_n \in \mathscr{F}_n \wedge \pi(F_n, G_n) < \delta$$
$$\Rightarrow |T_n(F_n) - T_n(G_n)| < \varepsilon\}.$$

($F(\,.\,)$ denotes the probability measure induced by F in the space being considered, respectively.)

LEMMA 1. *Let F and $\{T_n\}$ be given. Then condition* (B) *implies* (A) (*robustness of $\{T_n\}$ at F*).

PROOF. Let $\varepsilon > 0$ for (A) be given. In condition (B), choose the same ε, and choose $\eta = \varepsilon/2$. Let $\delta_B > 0$ and $\{\mathscr{E}_n\}$ ($n \geqq 1$) be the objects whose existence is given by (B) (for the given choice of ε and η). Put $\delta = \min\{\delta_B^2, \frac{1}{4}\varepsilon^2\}$.

Let G be such that $\pi(F, G) < \delta$. Then Strassen's theorem implies the existence of a probability measure Q in the product space $\Omega \times \Omega$ with points $\tilde{\omega} = (\omega_1, \omega_2)$, such that the marginals of Q are F and G and that $Q\{\tilde{\omega}: |\omega_1, \omega_2| \geqq \delta\} < \delta$ (where $|\cdot,\cdot|$ denotes the metric in Ω). Let $D = \{\tilde{\omega}: |\omega_1, \omega_2| \geqq \delta\}$. But $Q(D) < \delta$ implies immediately by a variant of Chebyshev's inequality (for nonnegative random variables) that for n independent observations $\tilde{\omega}_i$ in $\Omega \times \Omega$, $Q\{(\#\text{ of } \tilde{\omega}_i \in D)/n \geqq$

$\delta^{\frac{1}{4}}\} < \delta^{\frac{1}{2}}$, for all n (this bound is quite crude, but sufficient for our purposes). Now let F_n resp. G_n be the distributions determined by the first resp. second coordinates of the $\tilde{\omega}_i$ $(i \leqq n)$; then $(\#$ of $\tilde{\omega}_i \in D)/n < \delta^{\frac{1}{2}}$ implies $\pi(F_n, G_n) < \delta^{\frac{1}{2}}$. Therefore $Q\{F_n \in \mathscr{E}_n \wedge \pi(F_n, G_n) < \delta^{\frac{1}{2}}\} > 1-\varepsilon/2-\delta^{\frac{1}{4}} \geqq 1-\varepsilon$; hence (B) implies $Q\{|T_n(F_n)-T_n(G_n)| < \varepsilon\} > 1-\varepsilon$ and by Strassen's theorem $\pi(\mathscr{L}_F T_n, \mathscr{L}_G T_n) < \varepsilon$, for all n.

DEFINITION. A sequence of estimators $\{T_n\}$ is continuous at F iff

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \exists\, n_0\ \forall\, n, m \geqq n_0\ \forall\, F_n, F_m:$$
$$\{F_n \in \mathscr{F}_n \wedge F_m \in \mathscr{F}_m \wedge \pi(F, F_n) < \delta \wedge \pi(F, F_m) < \delta$$
$$\Rightarrow |T_n(F_n) - T_m(F_m)| < \varepsilon\}.$$

This definition is analogous to that of a weak*-continuous functional T (which henceforth will be called "continuous" for short), making use of the Prokhorov distance.

Obviously, if $\{T_n\}$ is such that $T_n \equiv T|\mathscr{F}_n$ for some T and all n, then continuity of T at some F implies continuity of $\{T_n\}$ at F (but not conversely). Furthermore, in this case $\{T_n\}$ is consistent for $T(F)$ under F. In general we have:

LEMMA 2. *If $\{T_n\}$ is continuous at F, then, under F, $\{T_n\}$ is consistent for some $T_\infty(F)$, i.e., $\mathscr{L}_F(T_n) \to T_\infty(F)$ (weakly).*

PROOF. For a sequence $\varepsilon_i \downarrow 0$ choose $\delta_i \downarrow 0$ and $n_i \uparrow \infty$ such that the continuity condition is fulfilled for $n, m > n_i$ (for each i). Let $A_i \subset R^k$ be the closure of the set $\{T_n(F_n) : \pi(F_n, F) < \delta_i, n > n_i\}$ for each i. The A_i form a monotone decreasing sequence of compact sets whose diameters ($\leqq 2\varepsilon_i$) tend to 0, hence they converge to a single point $T_\infty(F)$. But $F\{F_n : T_n(F_n) \in A_i\} \to 1$ as $n \to \infty$ for each i (because F_n converges a.s. to F), hence $T_n(F_n) \to T_\infty(F)$ in probability. (We even get $T_n(F_n) \to T_\infty(F)$ a.s., but we will not need this result.)

In the following theorem, let Ω^n be the n-fold Cartesian product of Ω with itself, endowed with the product measure and with the metric given by the maximum of the coordinate distances. Then T_n can be considered as a point function from Ω^n to R^k (invariant under permutations of the coordinates).

THEOREM 1. *Let a sequence of estimators $\{T_n\}$ be such that*

(i) T_n is continuous as a point function on Ω^n for every n;

(ii) $\{T_n\}$ is continuous at F.

Then $\{T_n\}$ is robust at F.

PROOF. We shall show that condition (B) is fulfilled. Let $\varepsilon > 0$ and $\eta > 0$ be given (for (B)). By Lemma 2, there exists $T_\infty(F)$, and there exists a $\delta_0 > 0$ and n_1 such that $\pi(F, F_n) < 2\delta_0$ implies $|T_\infty(F) - T_n(F_n)| < \varepsilon/2$ for $n \geqq n_1$; and there exists $n_0 \geqq n_1$ such that $F\{\pi(F, F_n) \geqq \delta_0\} < \eta$ for $n \geqq n_0$. Hence we can take $\mathscr{E}_n = \{F_n : \pi(F, F_n) < \delta_0\}$ for $n \geqq n_0$.

For each $n < n_0$ consider T_n as a function on $\Omega^n = \{\omega, \omega', \cdots\}$; denote the metric on Ω^n by $|\cdot,\cdot|$ and the product measure by F^n. Then for each $\omega \in \Omega^n$ there is a

$\delta = \delta(\omega) > 0$ such that $|\omega, \omega'| < \delta$ implies $|T_n(\omega) - T_n(\omega')| < \varepsilon/2$; hence ω', $\omega'' \in U_\delta(\omega)$ (open δ-neighborhood) implies $|T_n(\omega') - T_n(\omega'')| < \varepsilon$. Let $\{\alpha_i\}$, $\alpha_i > 0$, $\alpha_i \downarrow 0$ be some sequence; define $A_i = \{\omega : \delta(\omega) > \alpha_i\}$ and $B_i = \bigcup\{U_{\delta(\omega)/2}(\omega) : \omega \in A_i\}$ (open). Since $\bigcup A_i = \Omega = \bigcup B_i$, there is a j with $F^n\{\bigcup_1{}^j B_i\} > 1 - \eta$. Define $\mathscr{E}_n{}' = \bigcup_1{}^j B_i$, and $\delta_n = \min\{\alpha_j/2, 1/n_0\}$. Then for $\omega' \in \mathscr{E}_n{}'$, $|\omega', \omega''| < \delta_n$ there is an ω with $\delta(\omega)/2 > \delta_n$ and $\omega' \in U_{\delta(\omega)/2}(\omega)$, hence $\omega'' \in U_{\delta(\omega)}(\omega)$ and $|T_n(\omega') - T_n(\omega'')| < \varepsilon$. Let $\mathscr{E}_n{}''$ (open) be the symmetrization of $\mathscr{E}_n{}'$, i.e., the set of all points which by means of a permutation of their coordinates can be transformed into a point of $\mathscr{E}_n{}'$. Then $\mathscr{E}_n{}''$ enjoys the same properties as mentioned for $\mathscr{E}_n{}'$, and it corresponds to a set $\mathscr{E}_n \subset \mathscr{F}_n$. Let $F_n \in \mathscr{E}_n$, $G_n \in \mathscr{F}_n$, $\pi(F_n, G_n) < \delta_n$ $(\leqq 1/n_0)$, then for some corresponding ω', ω'' we have $\omega' \in \mathscr{E}_n{}''$, $|\omega', \omega''| < \delta_n$, hence $|T_n(F_n) - T_n(G_n)| < \varepsilon$. Now define $\delta = \min\{\delta_i : 0 \leqq i < n_0\}$; this δ and the $\mathscr{E}_n (n \geqq 1)$ satisfy condition (B).

COROLLARY. *Let $T : \mathscr{F} \to R^k$ be such that T is continuous at F and $T_n \equiv T \mid \mathscr{F}_n$, considered as a function on Ω^n, is continuous with respect to the metric in Ω^n, for all n. Then $\{T_n\}$ is robust at F.*

THEOREM 1a. *Let $\{T_n\}$ be such that $\{T_n\}$ is continuous at F and that for every n, T_n is continuous as a point function on Ω^n, except for a set of F^n-measure* 0 (*where again F^n denotes the product measure on Ω^n, determined by F on Ω*). *Then $\{T_n\}$ is robust at F.*

(This slightly stronger version of Theorem 1 follows directly from the proof of the theorem.)

Continuity of $\{T_n\}$ or T at F is not implied by robustness and consistency at F. However, the following simple statement holds:

LEMMA 3. *Assume $\{T_n\}$ is robust at F and consistent at all G in a neighborhood of F. Then $T_\infty(G)$ is continuous at F.*

PROOF. Assume there is an $\varepsilon < \frac{1}{2}$ and a sequence $\{G_i\}$ with $G_i \to F$ and $|T_\infty(F) - T_\infty(G_i)| \geqq 2\varepsilon$, hence $\pi(T_\infty(F), T_\infty(G_i)) \geqq 2\varepsilon$ (for all i). Robustness at F implies that for some j and $G_j = G$, $\pi(\mathscr{L}_F(T_n), \mathscr{L}_G(T_n)) < \varepsilon$ for all n. But $\mathscr{L}_F(T_n) \to T_\infty(F)$ and $\mathscr{L}_G(T_n) \to T_\infty(G)$, which leads to a contradiction.

COROLLARY. *Assume $\{T_n\}$ is robust and consistent at all $F \in \mathscr{F}$. Then T_∞ is continuous at all F, $T_\infty \mid \mathscr{F}_n$ yields a continuous point function on Ω^n, and $\{\tilde{T}_n\}$ defined by $\tilde{T}_n \equiv T_\infty \mid \mathscr{F}_n$ is robust and consistent, tending to $T_\infty(F)$, at all F.*

THEOREM 2. *Let $T : \mathscr{F} \to R^k$ and $\{T_n\}$ be defined by $T_n \equiv T \mid \mathscr{F}_n$ for all n. Then the following three conditions are equivalent*:

(i) *T is continuous at all F*;

(ii) *$\{T_n\}$ is robust and consistent, tending to $T(F)$, at all F*;

(iii) *$\forall K \subset \mathscr{F}$, K relatively compact, $\forall \varepsilon > 0\ \exists \delta > 0\ \forall F \in \mathscr{F}\ \forall G \in \mathscr{F}$*:

$$\{F \in K \wedge \pi(F, G) < \delta \Rightarrow |T(F) - T(G)| < \varepsilon\}.$$

PROOF. (i) $\Leftrightarrow$ (ii) has been shown since the pseudo-metric induced by the Prokhorov distance in Ω^n via $\mathscr{F}_n$ is numerically always less or equal to the (product) metric in Ω^n (defined above), as follows directly from the definition of the Prokhorov distance. (iii) $\Rightarrow$ (i) by putting $K = F$. Conversely, (i) implies uniform continuity on the closure of K, and the usual proof shows that even (iii) is implied. ((iii) gives a technical criterion somewhat similar to, but not identical with condition (B).)

5. An alternative formalization of the intuitive concept of robustness. Let $\{(\Omega^n, \mathscr{A}^n)\}$ $(n \geqq 1)$ be the sequence of n-fold Cartesian products of $(\Omega, \mathscr{A})$ with itself. P_n, Q_n, $\cdots$ denote probability measures on Ω^n. If F is a probability measure on Ω, then F^n denotes its product measure on Ω^n.

Let $\tilde{\Omega}^n$ be the space Ω^n modulo the group of permutations of the coordinates, which stands in a natural 1:1 correspondence to $\mathscr{F}_n$; let the metric on $\tilde{\Omega}^n$, called subsequently π-metric, be the metric thus induced by the Prokhorov distance in $\mathscr{F}_n$. The topology (and σ-algebra) generated by this new metric on $\tilde{\Omega}^n$ coincides with the topology (resp. σ-algebra) on Ω^n modulo coordinate permutations. Considering measures on $\tilde{\Omega}^n$ amounts to considering measures that are invariant under permutations of the coordinates; however, every measure on Ω^n can be symmetrized to yield this property, and two measures that give rise to the same invariant measure cannot be distinguished anyway by the type of estimators used here. The probability measures on $\tilde{\Omega}^n$ corresponding to those on Ω^n (symmetrized if necessary)will be denoted by $\tilde{\mathscr{F}}^n$, $\tilde{\mathscr{P}}_n$, $\tilde{\mathscr{Q}}_n$, etc. For convenience of printing, $\mathscr{L}_F T_n$ may be replaced by $\mathscr{L}[F]T_n$.

DEFINITION. A sequence of estimators $\{T_n\}$ is Π-robust at F iff $\forall \varepsilon > 0 \, \exists \delta > 0 \, \forall n \, \forall Q_n : \{\pi(\tilde{F}^n, \tilde{Q}_n) < \delta$ (using the π-metric in $\tilde{\Omega}^n$) $\Rightarrow \pi(\mathscr{L}[F^n]T_n, \mathscr{L}[Q_n]T_n) < \varepsilon\}$. A sequence of estimators is Π-robust (everywhere) iff it is Π-robust at all $F \in \mathscr{F}$.

This definition allows for some dependence within the n-tuples of observed outcomes, and it also allows for slight changes of the distributions underlying the different observations. Thus we get rid of the strict assumption of independent identically distributed observations.

THEOREM 3. *Π-robustness (at some F, or everywhere) implies robustness (as defined earlier).*

PROOF. It is sufficient to observe that, for F, G on Ω, $\pi(F, G) < \delta$ implies $\pi(F^n, G^n) < \delta^{\frac{1}{2}}$ for all n, as has been shown in the proof of Lemma 1.

Counterexample. Robustness and Π-robustness are not equivalent. Take, e.g., $\Omega =$ the real line. Consider a point $\tilde{\omega}_n \in \tilde{\Omega}^n$ (or, equivalently, a sequence of n observations, disregarding the order, or a point in Ω^n up to coordinate permutations) whose coordinates are all different from each other. The highest probability this point can have under a product probability measure $\tilde{F}^n$ obviously is achieved only if the corresponding measure F on Ω is carried by those n different values, i.e., if it is a multinomial distribution; and then the probability is $n! \prod_{i=1}^n p_i$ (with $\sum p_i = 1$). This is maximized if $p_i = 1/n$ for all i, so the maximal probability $\tilde{\omega}_n$

can have is $n!/n^n$ which tends to zero as $n \to \infty$. Now choose a point $\omega_0 \in \Omega$ and choose a sequence of points $\tilde{\omega}_n \in \tilde{\Omega}^n$ such that for each n all coordinates are different from each other and from ω_0 and such that their maximum distance from ω_0 tends to zero as $n \to \infty$. Define F as the point mass 1 in ω_0; and define $\tilde{Q}_n$ as the point mass 1 in $\tilde{\omega}_n$, for each n. Finally, define $\{T_n\}$, $T_n : \mathscr{F}_n \to R^1$, by: $T_n = 1$ if the distribution corresponding to $\tilde{Q}_n$ occurs, $T_n = 0$ otherwise. Then $\{T_n\}$ is robust at F, however, as not only product measures may be close to $\tilde{F}^n$ (take $\tilde{Q}_n$), $\{T_n\}$ is not Π-robust at F.

THEOREM 4. *All the theorems, lemmas and corollaries proved in Section* 4 *about robustness remain true if "robust" is replaced by "*Π*-robust".*

PROOF. It is sufficient to show that condition (B) implies Π-robustness. But this follows immediately from the last sentence of the proof of Lemma 1: Given ε, choose the quantities in condition (B) $\varepsilon_B = \varepsilon$, $\eta_B = \varepsilon/2$; take a $\delta_B = \delta_B(\varepsilon_B, \eta_B)$ and define $\delta = \min(\delta_B, \varepsilon/2)$; then replace $\delta^{\frac{1}{2}}$ by δ in the aforementioned sentence.

6. Definition of the breakdown point. Before mentioning some examples, we shall state a definition which is useful in practice by telling us, loosely speaking, "how far" the robustness of an estimator extends.

DEFINITION. Let $\{T_n\}$ be a sequence of estimators. The breakdown point δ^* of $\{T_n\}$ at some probability measure F is defined as follows: $\delta^* = \delta^*(\{T_n\}, F) = \sup\{\delta \leqq 1 : \exists$ a compact set $K = K(\delta)$ which is a proper subset of the parameter space such that $\pi(F, G) < \delta \Rightarrow G\{T_n \in K\} \to 1$ as $n \to \infty\}$.

The breakdown point tells us up to which Prokhorov distance from the parametric model (or, typically, up to what fraction of gross errors) the estimator still gives us *some* indication of the original distribution *within* the parametric model in the sense of excluding part of the parameter space. (Compare also the similar, but less general concept defined in Hodges (1967).)

7. Examples. The following examples have been investigated in more or less detail and are given briefly without proofs. Ω is the real line in all cases.

(i) The arithmetic mean is nowhere robust (and nowhere continuous). The same holds both for the standard deviation and the mean deviation; however, there still is a difference in the "non-robustness" of these two estimators which sheds some light on the old dispute between Fisher and Eddington about their use as measures of dispersion (see Tukey (1960) and Hampel (1968), page 92). The median is robust (and continuous) at F if and only if $F^{-1}(\frac{1}{2})$ contains no more than one point; its breakdown point is $\frac{1}{2}$. It is not robust in any open set of distributions. The α-trimmed mean $(0 < \alpha < \frac{1}{2})$, defined as $(\int_{\alpha}^{1-\alpha} F^{-1}(t)\,dt)/(1-2\alpha)$ (in the obvious way), is robust and continuous at all distributions, with breakdown point α. Perhaps it should also be remarked that the linear functions of the order statistics which are asymptotically optimal for the logistic respectively for the Cauchy distribution are nowhere robust and have breakdown point 0. In general, the breakdown point of a linear function of the order statistics is equal to the smaller one of the two fractions of mass at either end of the distribution which receive weights identically equal to zero.

(ii) The following definition yields a class of location parameter estimators (compare Huber (1964); see also Hampel (1968) for generalizations): Let $\psi(\cdot)$ be a monotone non-decreasing function on the real line which takes on negative and positive values; define the estimator θ as the solution (if it exists) of $\int \psi(x-\theta)F(dx) = 0$ (where F may be the empirical distribution function) or more generally as a value such that $\int \psi(x-\theta+\varepsilon)F(dx) \geqq 0$ and $\int \psi(x-\theta-\varepsilon)F(dx) \leqq 0$ for all $\varepsilon > 0$ (if the solutions fill an interval, one may take its midpoint to enforce uniqueness). Then ψ defines a robust sequence of estimators at F if and only if ψ is bounded and the solution of the defining equations for θ (with F) is unique. An important class of examples are the Huber-estimators $H_1(k)$ $(0 < k < \infty)$ for the mean of a normal distribution with variance 1, defined by $\psi_k(x) = x$ for $|x| \leqq k$, $\psi_k(x) = k \cdot \operatorname{sign}(x)$ for $|x| > k$. They are robust at any normal distribution (and many others) with breakdown point $\frac{1}{2}$. Another example is the "ϕ-estimator" (see Hampel (1968), page 66) with $\psi(x) = \phi(x)-\frac{1}{2}$, where ϕ is the cumulative distribution function of the standard normal distribution. It is robust at all distributions, and at normal distributions with variance 1 it behaves locally like the Hodges-Lehmann-estimator (see (iii), below), but its breakdown point is $\frac{1}{2}$. One may note that arithmetic mean and median can also serve here as examples with $\psi(x) = x$ resp. $\psi(x) = \operatorname{sign}(x)$.

(iii) Another class of location parameter estimators can be obtained from two-sample rank tests in the following way (compare Huber (1968b); see also Jaeckel (1969) and Hodges and Lehmann (1963)): Let $J(\cdot)$ be a monotone non-decreasing function on the open interval (0, 1) which takes on negative and positive values and which is integrable with $\int_0^1 J(t)\,dt = 0$. For given F define $G_\theta(x) = \frac{1}{2}[F(x)+1-F(2\theta-x-0)]$, then the estimator θ is the (or a) solution (if it exists) of $\int J(G_\theta(F^{-1}(t)))\,dt = 0$, or again more generally a value θ such that $\int J(G_{\theta+\varepsilon}(F^{-1}(t)))\,dt \leqq 0$ and $\int J(G_{\theta-\varepsilon}(F^{-1}(t)))\,dt \geqq 0$ for all $\varepsilon > 0$. If the solution is unique for F, then the sequence of estimators defined by J is robust (and continuous) at F. An example is again the median, with $J(t) = \operatorname{sign}(t-\frac{1}{2})$. Another example is the Hodges-Lehmann-estimator which can be defined by $J(t) = 2t-1$; it is robust (and continuous) at the normal (and many other) distributions with breakdown point $1-2^{-\frac{1}{2}} \approx 0.29$ (see also Hodges (1967)). Finally, $J(t) = \phi^{-1}(t)$ (where ϕ denotes again the standard normal distribution) defines an estimator which is asymptotically equivalent to the estimators derived from the Fisher-Yates (normal scores) test and the van der Waerden test. It is robust (and continuous) at the normal (and many other) distributions with breakdown point $2(1-\phi((2\log 2)^{\frac{1}{2}})) \approx 0.24$. (It may be noted that the investigation of robustness and breakdown point both in (ii) and (iii) is simplified by the location invariance and by the monotonicity of the functions defining the estimators. Hence, e.g. the evaluation of the breakdown point amounts to determining the smallest contaminating mass that when moved towards infinity, can carry the value of the estimator beyond all bounds. In the case of (iii), this leads to the equation $\int_{\delta/2}^{1/2} J(t)\,dt + \int_{1-\delta/2}^{1} J(t)\,dt = 0$ for the breakdown point δ.)

Acknowledgments. Many thanks for providing the opportunity of working in this field, and credit for insisting on my submitting this paper for publication, belong (mainly) to K. Jacobs, L. LeCam, E. L. Lehmann, P. J. Huber, and V. Strassen. Some helpful suggestions have been given by a referee.

REFERENCES

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Unpublished dissertation, Univ. of California, Berkeley.

HODGES, J. L., Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 163–186. Univ. of California Press.

HODGES, J. L., Jr. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1968a). Robust confidence limits. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 269–278.

HUBER, P. J. (1968b). Robust estimation. *Mathematical Centre Tracts* **27** 3–25. Mathematisch Centrum, Amsterdam.

JAECKEL, L. A. (1969). Robust estimates of location. Unpublished dissertation, Univ. of California, Berkeley.

PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1** 157–214.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, (Ingram Olkin, ed.) 448–485. Stanford Univ. Press.