MATHEMATICAL CENTRE TRACTS 27

EUROPEAN MEETING 1968

SELECTED STATISTICAL PAPERS 2

2nd printing

MATHEMATISCH CENTRUM AMSTERDAM 1975

AMS(MOS) subject classification scheme (1970): 60J10, 60K05, 62E10, 62F10, 62C05, 62G10, 62J05

ISBN 90 6196 040 1

1st printing 1973

2nd printing 1975

PREFACE

This volume of the series "Mathematical Centre Tracts" is published on the occasion of the European Meeting 1968 on Statistics, Econometrics and Management Science in Amsterdam. With permission of the Organizing Committee of this Meeting, the Statistical Department of the Mathematical Centre has invited some authors of papers on Statistics and Probability Theory to publish their work in the form of this Tract. This second volume contains five papers, on rather varying subjects. Just as in the first volume (Selected Statistical Papers 1, M.C. Tract 26), the papers appear in an almost random order, determined by the date of arrival of the manuscript.

Contents

ROBUST ESTIMATION

By Peter J. Huber (Switzerland)
Swiss Federal Institute of Technology, Zürich

CONTENTS

This survey paper neither claims to be exhaustive nor attempts to be impartial. Because of my personal tastes it is heavily biased toward minimax theory, and several important lines of thought will only be mentioned, but not adequately discussed. In particular, I shall concentrate almost exclusively on the simplest, best known and most important case: that of estimating one single location parameter.

1. HISTORICAL REMARKS, OR: THE DOGMA OF NORMALITY

The dogma that measurement errors should be distributed according to the normal law is still widespread among users of the method of least squares; I hope that the following historical remarks will help to clarify some of the issues. I am much indebted to CH. EISENHART for drawing my attention to two crucial 19th century references.

The theory of estimation originated with problems where almost all of the statistical variability of the data is due to measurement errors. This situation should be clearly distinguished from the opposite case where the data show a large internal variability and where good reasons can be advanced for the use of the sample mean, or of the sample median, as estimates for the corresponding population parameters. But in our case, statistical variability is just a nuisance to get rid of, and one is mainly interested in finding that combination of the observations which lies on the average nearest to the true value.

It is illuminating to witness how the normal, or Gaussian, distribution was introduced by GAUSS himself. I quote GAUSS (1821): "Der Verfasser gegenwärtiger Abhandlung, welcher im Jahr 1797 diese Aufgabe nach den Grundsätzen der Wahrscheinlichkeitsrechnung zuerst untersuchte, fand bald, dass die Ausmittelung der wahrscheinlichsten Werthe der unbekannten Grösse unmöglich sei, wenn nicht die Function, die die Wahrscheinlichkeit der Fehler darstellt, bekannt ist. In so fern sie dies aber nicht ist, bleibt nichts übrig, als hypothetisch eine solche Function anzunehmen. Es schien ihm das natürlichste, zuerst den umgekehrten Weg einzuschlagen und die Function zu suchen, die zum Grunde gelegt werden muss, wenn eine allgemein als gut anerkannte Regel für den einfachsten aller Fälle hervorgehen soll, die nemlich, dass das arithmetische Mittel aus mehreren für eine und dieselbe Grösse durch Beobachtungen von gleicher Zuverlässigkeit gefundenen Werthen als der wahrscheinlichste betrachtet werden müsse. Es ergab sich daraus, dass die Wahrscheinlichkeit eines Fehlers x, einer Exponentialgrösse von der Form e^{-hhxx} proportional angenommen werden müsse, und dass dann gerade diejenige Methode, auf die er schon einige Jahre zuvor durch andere Betrachtungen gekommen war, allgemein nothwendig werde. Diese Methode, welche er nachher besonders seit 1801 bei allerlei astronomische Rechnungen fast täglich anzuwenden Gelegenheit hatte, und auf welche auch LEGENDRE inzwischen gekommen war, ist jetzt unter dem Namen Methode der kleinsten Quadrate im allgemeinen Gebrauch."

Note that GAUSS here introduces the normal distribution to suit the sample mean. It is amusing to observe how the use of the arithmetic

mean became almost sacred over the years -- I believe mostly because one misunderstood the GAUSS-MARKOV theorem ("the best *linear unbiased* estimate of the expected value is the sample mean") and the Central Limit theorem ("the sum of *many small* independent elementary errors is *approximately* normal") in conjunction with the theorem that for independent identically distributed normal observations the sample mean is indeed best in almost every conceivable sense. I have italicized the crucial words in the above paraphrases of the theorems; for instance, there is no reason, except mathematical convenience, to impose linearity or unbiasedness, and one might argue from sad experience that the model should also allow for a *few gross* elementary errors occurring with low probability.

Moreover, one can hardly claim that the sample mean was universally accepted, as Gauss did. There is a charming contemporary paper (ANONYMOUS 1821), which first states that good reasons can be advanced for the use of the sample mean in the case of inherent statistical variability of the data, as opposed to mere measurement errors, but which then continues: "Cependant, dans ce cas même, la méthode vulgaire n'a pas été généralement suivi ni pratiquée sans quelques restrictions. Il est, par exemple, certaines provinces de France où, pour déterminer le revenu moyen d'une propriété territoriale, il est d'usage de considérer ce revenu durant vingt années consécutives, d'en distraire le revenu le plus fort et le plus faible, et de prendre ensuite le dix-huitième de la somme des autres."

The unknown author then continues to remark that a considerable arbitrariness is involved here: why shouldn't one exclude the two greatest and the two smallest observations? But nevertheless he does not believe that all observations should enter with the same weight into the determination of the mean.

BESSEL (1838, p. 67) states that he never rejected an observation for internal reasons, i.e. because it deviated too much from the majority of the observations, and that he gave them all the same weight. He remarks: "Wir haben geglaubt, nur durch die feste Beobachtung dieser Regel, Willkür aus unseren Resultaten entfernen zu können."

It seems to me that this kind of discussion borders on an irrational dogmatism; a more rational action would have been to look at actual error distributions in large samples obtained under controlled conditions, to check whether they were compatible with a normal distribution and, if not, to develop a more appropriate theory of estimation.

Actually, BESSEL himself (1818, p.19f.) had made such a comparison. He notes that all three of his test samples show a slightly higher frequency of large errors than the normal distribution would predict, but he minimizes this discrepancy and failes to recognize its significance (the sample mean is a poor estimate of location for longer-tailed distributions).

Much later, NEWCOMB (1886) notes: "In practice, large errors are more frequent than this equation (the normal law) would indicate them to be". He suggests that the square exponent of the normal density function should be replaced by a less rapidly increasing function. "The management of such an exponent might, however, prove inconvenient, and I shall adopt a law of error founded on the very probable hypothesis that we are dealing with a mixture of observations having various measures of precision." Thus, he adopts an error distribution with density

$$(1.1)\qquad \frac{1}{\sqrt{2\pi}}\left\{\frac{p_1}{\sigma_1}e^{-\frac{x^2}{2\sigma_1^2}} + \ldots + \frac{p_m}{\sigma_m}e^{-\frac{x^2}{2\sigma_m^2}}\right\}$$

and proposes to use the Bayes estimate for a uniform prior distribution. Roughly, this amounts to giving lesser weights to more extreme observations.

There was essentially no progress beyond NEWCOMB in the subsequent 60 years, even though the situation had been quite clearly recognized by eminent statisticians like STUDENT (1927) and JEFFREYS (1932).

But it seems that nobody had realized how bad the classical estimates could be in slightly non-normal situations and how much security and accuracy could be gained by some quite simple alternative procedures, like the French custom, quoted above, of deleting extreme

observations. The procedures of NEWCOMB and of JEFFREYS were excessively laborious; according to the latter "each approximation took about 6 hours' work, using a Marchant calculating machine and the tables of Milne-Thomson and Comrie" (l.c. p. 85).

The turning point came after World War II, when TUKEY and the Statistical Research Group at Princeton began to propagandize the problem, to emphasize the shortcomings of the classical estimates and -- perhaps most important of all -- to establish properties of several really practicable alternatives to them. Much of the material was disseminated in unpublished reports of the Statistical Research Group, which are almost inaccessible now. A survey paper was later published by TUKEY (1960); compare also TUKEY (1962). As a latecomer and outsider I have difficulties in properly assigning the merits for different important innovations, and I will therefore refrain from sketching the history of this period.

2. WHAT IS ROBUST ESTIMATION?

Let me begin with a remark on terminology, namely on the notions "distribution-free" or "nonparametric" estimators. Personally, I think this terminology is ill-conceived and should be abandoned. It makes some sense to speak of distribution-free or nonparametric tests (meaning tests which are invariant under arbitrary order preserving homeomorphisms of the real line),because their performance does not depend on the true underlying distribution under the null hypothesis, and they keep their validity outside of a single real parameter family of distributions. Still, the *power* of such tests depends heavily on the underlying distributions, and since the performance of estimates is intimately connected with the power of certain tests, there are no really distribution-free estimates and confidence sets -- even though the probability that the interval $(X_{(k)}, X_{(n-k+1)})$ covers the true median does not depend on the underlying distribution, the average length of this interval does. Moreover, it is definitely awkward to talk of a nonparametric estimate of a location parameter. Indeed, it seems that "distribution-

free" and "nonparametric" are now gradually being superseded by "robust" in estimation contexts.

Through the work of TUKEY and others, one had recognized by 1960 that

(i) one never has a very accurate knowledge of the true underlying distribution;

(ii) the performance of some of the classical estimators is very unstable under small changes of the underlying distribution;

(iii) some alternative estimators like the α-trimmed mean (i.e. one removes a fixed small fraction α of extreme observations on either side and computes the mean of the rest) lose very little efficiency relative to the sample mean for an exactly normal law, but show a much better and more stable (more "robust") performance for other error distributions F.

The question now arises how to choose in a rational fashion between different robust estimates. This means that one has to make precise the goals one wants to achieve. I cannot report that unanimity has been reached, even if one restricts attention to what TUKEY and McLAUGHLIN (1963) call "robustness of efficiency".

To be specific, consider the problem of estimating a location parameter Θ from n independent observations $X_1, \ldots, X_n$, each distributed according to $P(X_i < x) = F(\frac{x-\Theta}{\sigma})$, where F is not exactly known.

In the course of a discussion at the IMS-Meeting in Washington, D.C., December 1967, at least four distinct goals for robust estimators have emerged.

According to the first, a robust estimator should possess

(1) a high absolute efficiency for *all* suitably smooth shapes F.

While there is hardly any doubt that goal (1) can be achieved asymptotically for large sample sizes, the currently available evidence points to a very slow convergence, much too slow that the estimates would be useable for moderate sample sizes (cf. TAKEUCHI (1967), HÁJEK and ŠIDÁK (1967) p. 264ff.) Thus one relaxes the requirements to one of the following:

(2) a high efficiency *relative* to the sample mean (and some other estimates), and this for all F (cf. BICKEL (1965));

(3) a high absolute efficiency over a strategically selected *finite* set $\{F_i\}$ of shapes (e.g. the normal, logistic, double-exponential, Cauchy and rectangular shapes), cf. BIRNBAUM and LASKA (1965), MIKÉ (1967), YHAP (1967), CROW and SIDDIQUI (1967);

(4) a small asymptotic variance over some *neighborhood* of *one* shape, in particular the normal one.

Personally, I favor approach (4). The rationale behind it is that we usually have quite a good idea of the approximate shape of the true distribution, so that it should suffice to consider the neighborhood of only one shape. On the other hand, we need a rather "full" set, to exclude estimates which are unstable under small changes of the underlying distribution. The following sections will be heavily biased toward this approach.

3. THREE METHODS FOR CONSTRUCTING ESTIMATES

As before, let $X_1, \ldots, X_n$ be independent random variables with distribution $P(X_i < x) = F(\frac{x-\theta}{\sigma})$. We shall assume that F is symmetric, having a density $f = F'$, and that the scale parameter $\sigma = 1$ is known. At the cost of some complications these assumptions could be relaxed, but the basic ideas can be presented more clearly this way. We shall not bother about regularity conditions.

(i) Maximum likelihood type estimators.

Let ρ be a real valued symmetric function of a real parameter, with derivative $\psi = \rho'$. Define a statistic $T_n = T_n(X_1, \ldots, X_n)$ either by

$$(3.1) \qquad \sum_{i \le n} \rho(X_i - T_n) = \inf_t \sum_{i \le n} \rho(X_i - t)$$

or by

(3.2) $$\sum_{i \le n} \Psi(X_i - T_n) = 0.$$

Under quite general conditions T_n is a consistent estimator of Θ, and $n^{\frac{1}{2}}(T_n - \Theta)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance

$$\sigma_M^2(F) = \frac{\int \Psi(x)^2 \, F(dx)}{(\int \Psi'(x) \, F(dx))^2} .$$

If we insert

(3.3) $$\Psi_o(x) = -f_o'(x)/f_o(x)$$

for $\Psi(x)$, then T_n is the maximum likelihood estimator of Θ for the true underlying distribution F_o and will under suitable regularity conditions be asymptotically efficient for $F = F_o$. (HUBER (1964), (1967)).

(ii) Linear combinations of order statistics.

Let $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ be the ordered sample; put

(3.4) $$T_n = \frac{1}{n} \sum_{i \le n} h(\frac{i}{n+1}) \, X_{(i)},$$

where h is some function satisfying $\int_o^1 h(t) \, dt = 1$.

Under quite general (but not yet entirely satisfactory) regularity conditions $n^{\frac{1}{2}}(T_n - \Theta)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance

$$\sigma_L^2(F) = \int_o^1 U(t)^2 \, dt - (\int_o^1 U(t) \, dt)^2 ,$$

where U is an indefinite integral of

$$U'(t) = \frac{h(t)}{f(F^{-1}(t))} .$$

If we choose

$$h(t) = \frac{1}{I(F_o)} \quad \psi_o'(F_o^{-1}(t)),$$

where ψ_o' is the derivative of (3.3) and $I(F_o) = \int \psi_o(x)^2 \, F_o(dx)$ is FISHER's information, then T_n is asymptotically efficient for F_o. (JUNG (1955); the best results to date are those of CHERNOFF, GASTWIRTH and JOHNS (1967); compare also BICKEL (1967).)

(iii) Estimates derived from rank tests.

Consider a 2-sample rank test for shift: let $Y_1,\ldots,Y_n$ and $Z_1,\ldots,Z_n$ be two independent samples with distributions $F(x)$ and $F(x - \Delta)$ respectively. Form the combined sample of size $N = 2n$ and take as test statistic for testing $\Delta = 0$ against $\Delta > 0$

$$W(Y_1,\ldots,Y_n;Z_1,\ldots,Z_n) = \sum_{i \leq n} J(\frac{i}{N+1})V_i$$

where $V_i = 1$ if the ith smallest entry in the combined sample is a Y, and $V_i = 0$ otherwise; assume for simplicity $J(t) = -J(1-t)$.

One can derive estimates of location from such tests: determine $T_n(X_1,\ldots,X_n)$ such that

$$W(X_1-T_n, \ \ldots \ , X_n-T_n; -(X_1-T_n), \ \ldots \ , -(X_n-T_n)) = 0.$$

The asymptotic behavior of T_n can be determined from the asymptotic power of the rank test; it turns out that $n^{\frac{1}{2}}(T_n-\theta)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance

$$\sigma_R^2(F) = \frac{\int J(t)^2 \, dt}{(\int \frac{d}{dx}(J(F(x))) \, f(x) \, dx)^2} \, .$$

The rank test is asymptotically most powerful for F_o if

$$J(t) = \psi_o(F_o^{-1}(t)),$$

and then the estimate T_n is asymptotically efficient for F_o. (CHERNOFF and SAVAGE (1958), HODGES and LEHMANN (1963), HÁJEK and ŠIDÁK (1967)).

4. ASYMPTOTICALLY ROBUST ESTIMATES

Let C be a convex compact set of distribution functions F. The problem is to find a sequence T_n of estimators of location which have a small asymptotic variance over the whole of C; more precisely, the maximum over C of the asymptotic variance should be least possible. We shall restrict attention to symmetric distributions.

Let F_o be the distribution in C having the smallest FISHER information $I(F) = \int(f'/f)^2 f\,dx$; there is one and only one such F_o, and in many interesting cases F_o can be determined explicitly through variational methods (cf. HUBER (1964)).

Thus, for any sequence T_n, the asymptotic variance of $n^{\frac{1}{2}}(T_n - \Theta)$ under F_o will at best be $1/I(F_o)$; our goal is to find a T_n such that the asymptotic variance does not exceed $1/I(F_o)$ for *any* $F \in C$.

In particular, this sequence T_n must be asymptotically efficient for F_o, and we shall therefore have a closer look at the estimators constructed in the preceding section.

Consider first the behavior of $\sigma^2_*(F)$ under infinitesimal variations of F, where the star stands for either M, L or R.
Let $F_\gamma = (1-\gamma)F_o + \gamma F_1$ with $F_1 \in C$, $0 \leq \gamma \leq 1$, then $F_\gamma \in C$ because of convexity. Explicit computations yield that

$$\frac{d}{d\gamma}(1/\sigma^2_*(F_\gamma)) = \frac{d}{d\gamma} I(F_\gamma) \geq 0 \quad \text{for } \gamma = 0.$$

Thus, one can expect that all three estimators will have good robustness properties at least in an immediate neighborhood of F_o. In case (i) one can show that $1/\sigma^2_M(F)$ is a convex function of F, hence $\sigma^2_M(F)$ has not only a local, but also a global maximum at F_o, and the sequence of maximum likelihood estimates for F_o solves the problem.

In the cases (ii) and (iii) I do not know whether the maximum of $\sigma_L^2(F)$ and $\sigma_R^2(F)$ at F_o is a global one, with one exception to be mentioned below.

To give a specific example, assume that C is the set of all distributions of the form $F = (1-\varepsilon)\Phi + \varepsilon H$, where $0 \leq \varepsilon < 1$ is a fixed number, Φ is the standard normal distribution, and H varies over the set of all symmetric probability distributions. Then, the least favorable F_o has the density

$$f_o(x) = \frac{1-\varepsilon}{\sqrt{2\pi}} e^{-\rho_o(x)}, \tag{4.1}$$

where

$$\begin{aligned} \rho_o(x) &= \tfrac{1}{2}x^2 && \text{for } |x| < k \\ &= k|x| - \tfrac{1}{2}k^2 && \text{for } |x| \geq k, \end{aligned} \tag{4.2}$$

with k depending on ε. Thus,

$$\begin{aligned} \psi_o(x) &= x && \text{for } |x| < k \\ &= k.\text{sign}(x) && \text{for } |x| \geq k. \end{aligned} \tag{4.3}$$

The maximum likelihood estimate for this F_o was treated in HUBER (1964). The best linear combination of order statistics for F_o is the α-trimmed mean, with $\alpha = F_o(-k)$, and in this case one can show that $\sigma_L^2(F)$ has a global maximum at F_o. Hence, also the trimmed mean is a solution of our problem. (The close connection between HUBER's estimate and the trimmed mean was recognized by BICKEL (1965), p.850: for any symmetric F, they have the same asymptotic behavior, if α and k determine each other uniquely through $\alpha = F(-k)$.) The estimate corresponding to (iii) does not seem to allow a simple description and has not been investigated so far.

The fact that ψ_o' is discontinuous sometimes causes trouble, and one might prefer to smooth ψ_o near $\pm k$, even though the estimates then lose their asymptotic minimax property. Note in this connection that the

$\psi_0(x) = (1-e^{-x})/(1+e^{-x})$ corresponding to the logistic distribution $F_0(x) = 1/(1+e^{-x})$ behaves much like a smooth version of (4.3) and if we take approach (iii) with the logistic in place of F_0, we obtain the HODGES-LEHMANN estimate -- the median of the pairwise means $\frac{1}{2}(X_i+X_j)$. This perhaps explains in a heuristic way the empirical fact that the trimmed mean, the HODGES-LEHMANN and the HUBER estimate have very similar performances over a rather wide range of distributions.

But it should also be pointed out that our intuition about robust estimates is very unreliable. For instance, I had once falsely believed (HUBER (1964), p.75) that my estimate would be asymptotically equivalent to Winsorizing (= replacing the r leftmost and the r rightmost observations by $X_{(r)}$ and $X_{(n-r+1)}$ respectively, and then taking the mean of the thus modified sample). Another example: GREGERSON (1961) proposed an estimate of the type (3.1), with $\rho(x) = -e^{-\frac{1}{2}x^2}$. This estimate gives small *negative* weights to all observations whose residuals exceed 1: increasing an extreme observation slightly decreases the value of the estimate! While this is not necessarily objectionable, it will come as a surprise to many people.

5. CRITICISMS AND COMPLEMENTS

1. First, there is the usual objection against any asymptotic theory in statistics: one never knows whether it is applicable for any given finite sample size. Direct calculations are not very manageable except for rather small sample sizes (TUKEY and McLAUGHLIN (1963), ANSCOMBE and BARRON (1966), CROW and SIDDIQUI (1967), GASTWIRTH and COHEN (1968)); in addition, several Monte Carlo studies have been reported (LEONE, JAYACHANDRAN and EISENSTAT (1967), DIXON and TUKEY (1968)). While such studies mathematically do not prove anything about the applicability of the asymptotic theory, they furnish convincing evidence that it is applicable for sample sizes 20 and larger, i.e. confidence levels between 1% and 5% derived from the asymptotic theory then seem to be sufficiently reliable.

The evidence to date indicates that there is an analogue to STUDENT's

t, with the robust estimate of location in the numerator, and which can be closely approximated by a t-distribution with the appropriate number of degrees of freedom (m-1, where m is the number of observations left intact by trimming or Winsorizing, and similarly for HUBER's estimate), but the exact form of the denominator giving the best results is somewhat in doubt.

2. One might question the wisdom of restricting the attention to symmetric shapes. It is difficult to avoid some restriction of this kind in an asymptotic theory; the reason for this is that unknown small asymmetries in the distribution F introduce unknown small biases into the estimates. For very large sample sizes these biases would take precedence over the random errors in the estimates, and the unique minimax estimate would thus be the sample median (HUBER (1964), p.83). However, if these biases should ever take precedence, the practicing statistician would presumably conclude that the sample size is unreasonably large. Thus, this problem can only be settled through a genuine finite sample theory (see section 7 below).

3. One might also question the appropriateness of a minimax theory, especially of an asymptotic one, since minimax methods generally are too pessimistic. I think there are two answers: first, it seems that sample sizes reasonable for a given amount of not necessarily symmetric contamination will not allow to determine the nature of this contamination, except in rather extreme cases (cf. HUBER (1964), p.82ff.). Second, one might check some of these extreme cases.
ROMANOWSKI and GREEN (1965) have collected some quite impressive empirical error distributions, and it turns out that their largest sample (n = 8688) behaves very much like the least favorable F_o for the 2%-contaminated normal distribution (it lies between the slightly different curves for the least favorable F_o for location (4.1) and the least favorable one for scale (9.1)). In this case, a very good estimate for location would be the 5%-trimmed mean; for their smaller samples the conclusions are less definitive, but also these suggest trimming rates between 1% and 10%.

4. The question of computing these estimates. The trimmed and the

Winsorized mean are probably easiest to compute; they need about $O(n \log n)$ operations for large samples, most of them spent for ordering the sample. Some limited evidence from Monte Carlo computations indicates that HUBER's proposal 2 (HUBER (1964), p.96ff. -- this is a variant of the maximum likelihood estimate (4.3) which simultaneously estimates scale) can be implemented using about the same number of operations. Despite all shortcuts proposed so far, the HODGES-LEHMANN estimate uses $O(n^2)$ operations.

5. Some Monte Carlo experiments of a student of mine (LOCHER (1966)) suggested some disturbing conjectures. He worked with NEWCOMB's model (1.1), taking $m = 2$, $\sigma_1 = 1$, $\sigma_2 = 3$, $p_1+p_2 = 1$, $0 \leq p_2 \leq 0.1$ and sample sizes 5, 10 and 20. He compared sample mean, trimmed mean and "proposal 2". It turned out that for sample size 5 there is almost no difference in performance (variance) for these three estimates.

This raised the question whether the asymptotic optimality theory is at all relevant for small sample sizes, i.e. whether the small sample minimax estimates do bear any resemblance to the asymptotic ones. Fortunately, this question could be settled in the affirmative (cf. section 7 below).

6. An interesting and quite different minimax problem in robust estimation has been solved by DOKSUM (1966).

6. THE NEYMAN-PEARSON LEMMA FOR CAPACITIES

According to the NEYMAN-PEARSON Lemma, the most powerful tests of a simple hypothesis P_o against a simple alternative P_1 are given by likelihood ratio tests: form $h(\underset{\sim}{X}) = \prod_{i=1}^{n} p_1(X_i)/p_o(X_i)$ and reject P_o if $h(\underset{\sim}{X}) > c$ (p_j is a density of P_j).

What happens if the P_j are only approximately known? Clearly, likelihood ratio tests may fail to be robust: a single factor $p_1(X_i)/p_o(X_i)$ equal (or almost equal) to 0 or ∞ might completely change $h(\underset{\sim}{X})$.

If for instance the uncertainty is formalized in terms of total variation, i.e. if we replace P_j by the composite hypothesis $\mathcal{P}_j = \{Q \mid ||Q - P_j|| \leq \varepsilon\}$, then it turns out that there is a least favorable pair $Q_j \in \mathcal{P}_j$ (j=0,1), such that the maximin tests of any level between $\mathcal{P}_o$ and $\mathcal{P}_1$ coincide with the likelihood ratio tests of the same level between Q_o and Q_1. The likelihood ratio

$$\frac{q_1(x)}{q_o(x)} = \min\left(c'', \max\left(c', \frac{p_1(x)}{p_o(x)} \right)\right), \quad c' < c'',$$

is a censored version of $p_1(x)/p_o(x)$ (HUBER (1965)).

The existence of such a least favorable pair (Q_o, Q_1) is indeed quite general; it is essentially equivalent to the fact that each $\mathcal{P}_j$ is the set of all probability measures majorized by some 2-alternating capacity (STRASSEN oral communication; for a somewhat weaker result see STRASSEN (1964)).

7. FINITE SAMPLE MINIMAX ESTIMATES

The generalization of the NEYMAN-PEARSON Lemma just mentioned can be used to construct a finite sample minimax theory for robust estimators.

Assume that the measurement errors $\Delta_i = X_i - \Theta$ are independent random variables whose distribution functions F_i lie anywhere within δ of the standard normal cumulative Φ:

$$(7.1) \qquad \sup_x \left| F_i(x) - \Phi(x) \right| \leq \delta .$$

(The method works for several different neighborhoods; note that the F_i need not be symmetric.)

Let $a > 0$ be a fixed number: the accuracy of any estimate $T = T(X_1, \ldots, X_n)$ of Θ shall be assessed by the least α for which one can guarantee

$$P\{T < \Theta - a\} \leq \alpha$$
$$P\{T > \Theta + a\} \leq \alpha$$

for all Θ and for all distributions satisfying (7.1) -- the smaller α, the better the estimate.

The corresponding minimax solution T^0 can be described explicitly as follows. Let $\psi_0(x)$ be defined by (4.3), where k depends on δ and a (but not on the sample size n) through the relation

$$e^{-2ak} \left[\Phi(a-k) - \Phi(-a-k) \right] = (1 + e^{-2ak})\delta .$$

Let T^* and T^{**} be the smallest and the largest solution T of

$$\sum_{i \leq n} \psi_0(X_i - T) = 0$$

respectively. Then put $T^0 = T^*$ or $T^0 = T^{**}$ at random with equal probability (HUBER (1968)).

The idea behind this result is very simple: one first constructs a maximin test (with level α and power $1-\alpha$) between $\Theta-a$ and $\Theta+a$ according

to the method of the preceding section, then one derives an estimate from this test in the manner of HODGES and LEHMANN (1963).

Note that this estimate is formally identical with an asymptotically robust estimate for symmetric contamination, mentioned in section 4.

8. SOME ROBUST ESTIMATES NOT FITTING IN OUR FRAMEWORK

1. The so-called "quick and dirty" methods are estimates based on a few selected order statistics. One recent proposal (GASTWIRTH (1966)) for instance takes the $33\frac{1}{3}$, 50, and $66\frac{2}{3}$ percentiles with weights 0.3, 0.4, and 0.3 respectively. Apart from being robust, they tend to have quite a high efficiency with respect to the best possible estimate (this one has efficiency approximately 80% or better, simultaneously for the Cauchy, double-exponential, logistic and normal distribution). Compare also CROW and SIDDIQUI (1967).

2. The HODGES-LEHMANN estimate (the median of the pairwise means $\frac{1}{2}(X_i+X_j)$) uses a disproportionate amount of computing time, namely $O(n^2)$, if the sample is large. BICKEL and HODGES (1967) have investigated a simplified version, namely the median of the pairwise means

$$\tfrac{1}{2}(X_{(1)}+X_{(n)}),\ \ldots\ ,\ \tfrac{1}{2}(X_{(k)}+X_{(n-k+1)}),\ \ldots \quad .$$

Here the computing time goes only with O(n log n), most of it spent for ordering the sample. This estimate has a very good performance, as it seems, but its asymptotic distribution is *not* normal (it can be represented as the distribution of the time which Brownian motion spends above some curve).

9. OTHER ESTIMATION PROBLEMS

1. The scale parameter problem can be reduced to that of a location parameter by taking logarithms. However, some difficulties arise since the resulting distributions then tend to be asymmetric, and it is not quite clear what one is estimating, if the underlying distribution is only

approximately known. But, since one has taken logarithms, this uncertainty only acts as an arbitrary additive constant, and it still makes sense to minimize the maximum of the asymptotic variance over some neighborhood of some model distribution. Also here the asymptotically efficient estimates for the least favorable distributions F_o seem to have good robustness -- and probably minimax -- properties.

For example, in the ε-contaminated normal case, the least favorable F_o has density

$$
\begin{aligned}
f_o(x) &= \frac{1-\varepsilon}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} && \text{for } |x| < q \\
&= \frac{1-\varepsilon}{\sqrt{2\pi}}\, e^{-\frac{1}{2}q^2} \left(\frac{q}{|x|}\right)^{(q^2)} && \text{for } |x| \geq q
\end{aligned}
\tag{9.1}
$$

where ε and q are related through

$$
\frac{\varepsilon}{1-\varepsilon} = \frac{2q}{q^2-1} \frac{e^{-\frac{1}{2}q^2}}{\sqrt{2\pi}} - 2\,\Phi(-q).
$$

The corresponding maximum likelihood estimate was treated in HUBER (1964); another asymptotically efficient estimate of σ^2 for the distribution $F_o(\frac{x}{\sigma})$ is the α-trimmed variance, suitably scaled, where $\alpha = F_o(-q)$.

2. The higher dimensional location parameter problem can be treated in very much the same way as the one-dimensional problem, if one assumes that the error distribution is spherically symmetrical. In particular, one can determine a least favorable F_o just as in the one-dimensional case; it is somewhat surprising that $-\log f_o(x)$ fails to be convex even in the simplest contaminated normal case. The maximum likelihood approach works well, compare GENTLEMAN (1965), HUBER (1967).

The higher dimensional analogue to trimming has been called "peeling" by TUKEY; it consists of deleting the extreme points of the convex hull of the observations, and to repeat this operation a fixed number of k times. Nothing is known about the behavior of such a procedure; one can surmise from RÉNYI and SULANKE (1963, 1964) that

it must be a very tricky problem.

A multivariate version of the HODGES-LEHMANN estimate has been considered by BICKEL (1964).

3. Almost nothing is known about robust estimation in the general case, where there is neither translation nor scale invariance. It is a fair guess that a modified maximum likelihood estimate should have good robustness properties: put

$$\psi(x,\theta) = \min(c_2(\theta),\ \max(c_1(\theta),\ \frac{\partial}{\partial\theta} \log f(x,\theta))),$$

where $f(x,\theta)$ is the probability density of the assumed family of distributions, and define an estimator T_n of θ by

$$\sum_{i\leq n} \psi(X_i,T_n) = 0 .$$

The most serious difficulty with this kind of problem is that one does not quite know what one is estimating. Perhaps one should define the parameter to be estimated in terms of the estimator.

Among the classical procedures, the method of moments is obviously non-robust, whereas minimum chi-square methods presumably have good robustness properties. Some investigations of such problems would certainly be desirable.

4. Regression and analysis of variance problems.
Consider the general linear regression problem

$$X_i = \sum_{j=1}^{m} c_{ij}\,\theta_j + b_i + \Delta_i, \quad 1 \leq i \leq n,$$

where the X_i are observed, the θ_j are to be estimated, the c_{ij} and b_i are known coefficients, and the Δ_i are independent random errors whose (symmetric) distribution function F is only approximately known.

The classical least squares method is to minimize

$$\sum_i (X_i - \sum_j c_{ij}\hat{\theta}_j - b_i)^2 ,$$

which generalizes at once to minimizing

$$\sum_i \rho_o(X_i - \sum_j c_{ij}\hat{\theta}_j - b_i) \quad ,$$

with ρ_o as in (4.2). It is not too difficult to develop an asymptotic theory for these estimates if one assumes that m stays fixed while n tends to infinity, but severe difficulties arise in the more realistic case, where both m and n become large comparably fast.

The other two estimators considered in section 3, (ii) and (iii), do not seem to generalize easily. LEHMANN and his students have attacked several regression and analysis of variance problems with the aid of rank tests (LEHMANN (1963a,b), plus a large number of papers by different authors in subsequent volumes of Ann. Math. Statist.), but it seems that each of these problems requires an individual treatment. To illustrate the basic idea, consider a simple regression problem (ADICHIE (1967)):

$$X_i = \alpha + \beta\, c_i + \Delta_i \quad ,$$

where α and β are to be estimated. Every test of the hypothesis $\beta = 0$ furnishes some estimate of β: apply the test to the pseudo-observations $X_i' = X_i - \hat{\beta}c_i$ and adjust the value of $\hat{\beta}$ in such a way that the test is least able to reject the hypothesis. In many cases the asymptotic power of the tests then can be used in a more or less straight-forward way to compute the asymptotic variances and covariances, and hence the asymptotic efficiencies, of these estimates.

Zürich, 4 June 1968

REFERENCES

ADICHIE, J.N. (1967). Estimates of regression parameters based on rank-tests. Ann.Math.Statist. 38, 894-904.

ANONYMOUS (1821). Dissertation sur la recherche du milieu le plus probable. Ann.Math. pures et appl. 12, 181-204.

ANSCOMBE, F. and B.A. BARRON (1966). Treatment of outliers in samples of size three. J. Res. Nat. Bureau of Standards 70B, 141-147.

BESSEL, F.W. (1818). Fundamenta astronomiae. (Königsberg).

BESSEL, F.W. und BAEYER (1838). Gradmessung in Ostpreussen. (Berlin).

BICKEL, P.J. (1964). On some alternative estimates for shift in the p-variate one sample problem. Ann. Math. Statist. 35, 1079-1090.

BICKEL, P.J. (1965). On some robust estimates of location. Ann. Math. Statist. 36, 847-858.

BICKEL, P.J. (1967). Some contributions to the theory of order statistics. Proc. Fifth Berkeley Symp. on Math. Statist. and Probability, Vol. 1, 575-591.

BICKEL, P.J. and J.L. HODGES (1967). The asymptotic theory of Galton's test and a related simple estimate of location. Ann. Math. Statist. 38, 73-89.

BIRNBAUM, A. and E. LASKA (1965). Optimal robustness for estimators and tests. Tech. report No. 334, Courant Inst. of Math. Sciences.

CHERNOFF, H. and I.R. SAVAGE (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. 29, 972-994.

CHERNOFF, H., J.L. GASTWIRTH and M.V. JOHNS (1967). Asymptotic distributions of linear combinations of functions of order statistics with applications to estimation. Ann. Math. Statist. 38, 52-72.

CROW, E.L. and M.M. SIDDIQUI (1967). Robust estimation of location. J. Amer. Statist. Assoc. 62, 353-389.

DIXON, W.J. and J.W. TUKEY (1968). Approximate behavior of the distribution of Winsorized t (Trimming/Winsorization 2). Technometrics 10, 83-98.

DOKSUM, K. (1966). Asymptotically minimax distribution-free procedures. Ann. Math. Statist. 37, 619-628.

GASTWIRTH, J.L. (1966). On robust procedures. J. Amer. Statist. Assoc. 61, 929-948.

GASTWIRTH, J.L. and M.L. COHEN (1968). The small sample behavior of some robust linear estimators of location. Tech. report No. 91, Dept. of Statistics, The John Hopkins University.

GAUSS, C.F. (1821). Göttingische gelehrte Anzeigen, p. 321-327 (reprinted in Werke Bd. 4, p. 98).

GENTLEMAN, W.M. (1965). Ph.D. Thesis, Princeton University (unpublished).

GREGERSON, L.F. (1961). The most probable value of a set of observations. The Canadian Surveyor, 333-339.

HÁJEK, J. and Z. ŠIDÁK (1967). Theory of rank tests. (Academic Press).

HODGES, J.L. and E.L. LEHMANN (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. (1965). A robust version of the probability ratio test. Ann. Math. Statist. 36, 1753-1758.

HUBER, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symp. on Math. Statist. and Probability, Vol 1, 221-233.

HUBER, P.J. (1968). Robust confidence limits. To be published in Z. f. Wahrscheinlichkeitstheorie.

JEFFREYS, H. (1932). An alternative to the rejection of outliers. Proc. Royal Soc. London, Ser. A 137, 78-87.

JUNG, J. (1955). On linear estimates defined by continuous weight function. Arkiv f. Mat. 3, 199-209.

LEHMANN, E.L. (1963a). Robust estimation in analysis of variance. Ann. Math. Statist. 34, 957-966.

LEHMANN, E.L. (1963b). Asymptotically nonparametric inference: an alternative approach to linear models. Ann. Math. Statist. 34, 1494-1506.

LEONE, F.C., T. JAYACHANDRAN and S. EISENSTAT (1967). A study of robust estimators. Technometrics 9, 652-660.

LOCHER, H. (1966). Robuste Schätzungen. Diplomarbeit ETH (unpubl.)

MIKÉ, V. (1967). Contributions to robust estimation. Tech. report IMM 361, Courant Institute of Math. Sciences.

NEWCOMB, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. Amer. J. of Math. 8, 343-366.

RÉNYI, A. and R. SULANKE (1963, 1964). Ueber die konvexe Hülle von n zufällig gewählten Punkten. I, II. Z. f. Wahrscheinlichkeitstheorie 2, 75-84; 3, 138-147.

ROMANOWSKI, M. and E. GREEN (1965). Practical applications of the modified normal distribution Bull. Géodésique N.S. 76, 1-20.

SIDDIQUI, M. and K. RAGHUNANDANAN (1967). Asymptotically robust estimators of location. J. Amer. Statist. Assoc. 62, 950-953.

STRASSEN, V. (1964). Messfehler und Information. Z. f. Wahrscheinlichkeitstheorie 2, 273-305.

STUDENT (1927). Errors of routine analysis. Biometrika 19, 151-164.

TAKEUCHI, K. (1967). Robust estimation and robust parameter. (Unpublished manuscript).

TUKEY, J.W. (1960). A survey of sampling from contaminated distributions. In: Contributions to Probability and Statistics (ed.Olkin), p. 448-485, Stanford Univ. Press.

TUKEY, J.W. (1962). The future of data analysis. Ann. Math. Statist. 33, 1-67.

TUKEY, J.W. and D.H. McLAUGHLIN (1963). Less vulnerable confidence and significance procedures for location based on a single sample. Sankhya Ser. A 25, 331-352.

YHAP, E. (1967). An asymptotic optimally robust linear unbiased estimator of location for symmetric shapes. (Unpublished manuscript).

LIMIT THEOREMS FOR STOCHASTIC PROCESSES OCCURRING IN STUDIES OF THE LIGHT-SENSITIVITY OF THE HUMAN EYE

by J.Th. Runnenburg (Netherlands)
University of Amsterdam

1. INTRODUCTION

Light quanta arrive at times $\alpha\underline{t}_1$, $\alpha\underline{t}_2$, ... , with $0 < \underline{t}_1 < \underline{t}_2 < \ldots$ and α a positive constant, according to a renewalprocess at a given spot on the retina of one of the eyes of an observer. Hence we assume that the differences

$$(1.1) \qquad \underline{y}_n = \underline{t}_n - \underline{t}_{n-1} \quad \text{for } n = 1,2, \ldots \quad (\text{with } \underline{t}_o = 0)$$

are identically distributed positive random variables with a common distribution function $A(y)$. We shall restrict our considerations to those $A(y)$ with finite first moment for which

$$(1.2) \qquad \lim_{y \downarrow 0} y^{-\rho} A(y) = c$$

for some positive constants ρ and c. Upon his arrival at the retina the n^{th} light quantum starts a lifetime of duration $\underline{s}_n$ (after which it is regarded as dead), where the $\underline{s}_n$ are identically distributed positive random variables with a common distribution function $B(s)$ with finite first and ρ^{th} moment. The random variables $\underline{y}_1$, $\underline{s}_1$, $\underline{y}_2$, $\underline{s}_2$, ... are mutually independent.

Our observer has been waiting in the dark for this particular sequence of light quanta to reach his eye and it is his task to record whether and if so how many times he notices light in the time interval $(0, t]$. It is assumed that he does notice light each time that the number of living light quanta on the retina of his eye jumps from less than k to at least k, where k is a natural number and at least 2.

We are interested in the distribution of $\underline{n}_\alpha(t)$, the number of

times the observer notices light in $(0, t]$. This distribution can at best be obtained numerically for exponentially distributed $\underset{\sim}{y}_n$ and $\underset{\sim}{s}_n$ (cf. ten Hoopen en Reuver [1965]). However, if we consider $\underset{\sim}{n}_\alpha(t_\alpha)$ for a suitable choice of t_α as a function of α, we may hope for limiting theorems for $\alpha \to \infty$. This work generalizes and provides an alternative approach to Ikeda [1965]. An earlier version of the results in this paper was obtained jointly with Meyer and is contained in his thesis, cf. Meyer [1967].

2. NEW LIGHT VERSUS LIGHT

In this section we consider the number of times our observer notices light under the restriction that exactly n_α light quanta reach the retina of his eye, where n_α is a natural number for each positive α with

$$\lim_{\alpha\to\infty} n_\alpha \alpha^{-\rho(k-1)} = 1. \tag{2.1}$$

We shall see that in the limit light is observed only on those occasions where $k-1$ consecutively arriving light quanta survive the arrival of the next quantum. If we write

$$A_j^\alpha = \{\text{light quanta } j-k+1, j-k+2, \ldots, j \text{ are simultaneously alive}\}, \tag{2.2}$$

we have for each positive α and each j with $k \le j \le n_\alpha$

$$\begin{aligned} P(A_j^\alpha) &= P\{\underset{\sim}{s}_{j-k+1} > \alpha(\underset{\sim}{y}_{j-k+2}+\ldots+\underset{\sim}{y}_j), \underset{\sim}{s}_{j-k+2} > \alpha(\underset{\sim}{y}_{j-k+3}+\ldots+\underset{\sim}{y}_j), \ldots \\ &\ldots, \underset{\sim}{s}_{j-1} > \alpha\, \underset{\sim}{y}_j\} = \\ &= \int_0^\infty \ldots \int_0^\infty \{1-B(\alpha(y_1+\ldots+y_{k-1}))\}\ldots\{1-B(\alpha y_{k-1})\}\, d\,A(y_1)\ldots d\,A(y_{k-1}) = \\ &= c^{k-1}\alpha^{-\rho(k-1)} \int_0^\infty \ldots \int_0^\infty \{1-B(t_1+\ldots+t_{k-1})\}\ldots\{1-B(t_{k-1})\}\, d\,A^\alpha(t_1)\ldots \\ &d\,A^\alpha(t_{k-1}), \end{aligned} \tag{2.3}$$

where

$$(2.4) \qquad A^{\alpha}(t_i) = \alpha^{\rho}\, c^{-1}\, A(\alpha^{-1} t_i) \quad \text{for } i = 1,2,\ldots, k-1.$$

Because of our assumption (1.2) we have

$$(2.5) \qquad \lim_{\alpha\to\infty} A^{\alpha}(t_i) = t_i^{\rho}$$

and the product measure with respect to which we integrate in the last integral of (2.3) tends to a limit (which is also a measure) as α tends to infinity. If we now take

$$(2.6) \qquad I_{k-1}(\rho) = \int_0^{\infty} \cdots \int_0^{\infty} \{1-B(t_1+\ldots+t_{k-1})\} \ldots \{1-B(t_{k-1})\}\, dt_1^{\rho} \ldots dt_{k-1}^{\rho}$$

it is easy to verify that $I_{k-1}(\rho)$ is finite, if we use the fact that

$$(2.7) \qquad \mathcal{E}s_{-1}^{\rho} = \int_0^{\infty} \{1-B(s)\}\, ds^{\rho} < \infty .$$

A generalization of the proof that a sequence of characteristic functions converges to a characteristic function if the corresponding distribution functions converge to a distribution function shows that

$$(2.8) \qquad \lim_{\alpha\to\infty} \alpha^{\rho(k-1)}\, P(A_j^{\alpha}) = c^{k-1}\, I_{k-1}(\rho).$$

We have now proved that

$$(2.9) \qquad \lim_{\alpha\to\infty} n_{\alpha}\, P(A_j^{\alpha}) = c^{k-1}\, I_{k-1}(\rho)$$

or that the expectation of the number of times a light quantum reaches the retina of our observer's eye at a time that the last k - 1 quanta arriving before him are still alive (out of a total of n_{α} arrivals) tends to a finite positive limit as α tends to infinity.

In Ikeda [1965] the integral $I_{k-1}(1)$ is evaluated. One first introduces the substitution $u_1 = t_1 + \ldots + t_{k-1}$, $\ldots$, $u_{k-1} = t_{k-1}$, leading to a (k-1)-fold integral over the region $0 \leq u_{k-1} \leq u_{k-2} \leq \ldots \leq u_1 < \infty$ with $\{1-B(u_1)\} \ldots \{1-B(u_{k-1})\}$ as integrand. If now we

integrate over the set of all $(u_1, u_2, \ldots, u_{k-1})$ with nonnegative components, we obtain $(k-1)!$ times $I_{k-1}(1)$ and the new integral is the product of $k-1$ factors each of which equals $\rho \underline{s}$. Hence

$$I_{k-1}(1) = \frac{(\rho s)^{k-1}}{(k-1)!} \quad . \tag{2.10}$$

We use the following notations

$$B_j^\alpha = \{\text{at least } k-1 \text{ light quanta are alive at the arrival of quantum } j\} \tag{2.11}$$

and

$$C_j^\alpha = B_j^\alpha - A_j^\alpha \quad . \tag{2.12}$$

Clearly B_j^α is the event that *light* is noticed by the observer at the j^{th} arrival. In contrast we shall say that A_j^α is the event that *new light* occurs at the j^{th} arrival. Because $A_j^\alpha \subset B_j^\alpha$, we have

$$P(B_j^\alpha) = P(A_j^\alpha) + P(C_j^\alpha) \quad . \tag{2.13}$$

We first prove a relation between light and new light.

Theorem 2.1

$$\lim_{\alpha\to\infty} P\{\text{exactly i of the } B_k^\alpha, B_{k+1}^\alpha, \ldots, B_{n_\alpha}^\alpha \text{ occur}\} =$$

$$= \lim_{\alpha\to\infty} P\{\text{exactly i of the } A_k^\alpha, A_{k+1}^\alpha, \ldots, A_{n_\alpha}^\alpha \text{ occur}\} \text{ for } i = 0,1,2, \ldots .$$

In the next section we shall compute the second limit for each i and hence this theorem is a major s ep in our considerations as it reduces the complexity of the calculation of the first limit to the comparatively easy computation of the second limit. In this section we assume that the second limit exists for each i.

Proof: it is clear that exactly i of the A_k^α , A_{k+1}^α , ..., $A_{n_\alpha}^\alpha$ occur if and only if either exactly i of the A_k^α , A_{k+1}^α , ... , $A_{n_\alpha}^\alpha$ and none of the C_k^α , C_{k+1}^α , ... , $C_{n_\alpha}^\alpha$ occur or alternatively exactly i of the A_k^α , A_{k+1}^α , ... , $A_{n_\alpha}^\alpha$ and at least one of the C_k^α , C_{k+1}^α , ... , $C_{n_\alpha}^\alpha$ occur. Also exactly i of the B_k^α , B_{k+1}^α , ... , $B_{n_\alpha}^\alpha$ occur if and only if either exactly i of the A_k^α , A_{k+1}^α , ... , $A_{n_\alpha}^\alpha$ and none of the C_k^α , C_{k+1}^α , ... , $C_{n_\alpha}^\alpha$ occur or alternatively exactly i of the B_k^α , B_{k+1}^α , ... , $B_{n_\alpha}^\alpha$ and at least one of the C_k^α , C_{k+1}^α , ... , $C_{n_\alpha}^\alpha$ occur. But then

$$(2.14)\quad \left|P\{\text{exactly i of the } A_k^\alpha, A_{k+1}^\alpha, \dots, A_{n_\alpha}^\alpha \text{ occur}\} - P\{\text{exactly i of the } B_k^\alpha, B_{k+1}^\alpha, \dots, B_{n_\alpha}^\alpha \text{ occur}\}\right| \leq$$
$$\leq P\{\text{at least one of the } C_k^\alpha, C_{k+1}^\alpha, \dots, C_{n_\alpha}^\alpha \text{ occurs}\}.$$

It is thus sufficient to prove

$$(2.15)\quad \lim_{\alpha\to\infty} P\{\text{at least one of the } C_k^\alpha, C_{k+1}^\alpha, \dots, C_{n_\alpha}^\alpha \text{ occurs}\} = 0.$$

This again is a consequence of $\lim_{\alpha\to\infty} \sum_{j=k}^{n_\alpha} P(C_j^\alpha) = 0$. As $P(C_k^\alpha) \leq$ $\leq P(C_{k+1}^\alpha) \leq \dots \leq P(C_{n_\alpha}^\alpha)$ (these inequalities hold trivially for the $P(B_j^\alpha)$ and $P(A_j^\alpha)$ does not depend on j), we need only prove $\lim_{\alpha\to\infty} n_\alpha P(C_{n_\alpha}^\alpha) = 0$ or

$$(2.16)\qquad \lim_{\alpha\to\infty} n_\alpha P(B_{n_\alpha}^\alpha) = c^{k-1} I_{k-1}(\rho).$$

Now

$$(2.17)\qquad P(B_j^\alpha) = P\{\text{at least k-1 of } D_1^\alpha, D_2^\alpha, \dots, D_{j-1}^\alpha \text{ occur}\},$$

where

$$(2.18)\qquad D_i^\alpha = \{\underline{s}_i^* > \alpha(\underline{y}_1^* + \dots + \underline{y}_i^*)\} \qquad \text{for } i = 1,2, \dots, j$$

and $(\underline{s}_1^*, \underline{s}_2^*, \dots, \underline{s}_{j-1}^*, \underline{y}_1^*, \underline{y}_2^*, \dots, \underline{y}_{j-1}^*)$ is just another notation

for $(\underline{s}_{j-1}, \underline{s}_{j-2}, \cdots, \underline{s}_1, \underline{y}_j, \underline{y}_{j-1}, \cdots, \underline{y}_2)$. But then for each j with $k \leq j \leq n_\alpha$

$$(2.19)\quad P(A_j^\alpha) \leq P(B_j^\alpha) \leq \sum_{1 \leq i_1 < i_2 < \ldots < i_{k-1} \leq j-1} P(D_{i_1} \cap D_{i_2} \cap \ldots \cap D_{i_{k-1}}) \leq$$

$$\leq \sum_{j_1 \geq 1, j_2 \geq 1, \ldots, j_{k-1} \geq 1} P(D_{j_1} \cap D_{j_1+j_2} \cap \ldots \cap D_{j_1+\ldots+j_{k-1}}) =$$

$$= \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \cdots \sum_{j_{k-1}=1}^{\infty} \int_0^\infty \cdots \int_0^\infty \{1-B(\alpha(y_1+\ldots+y_{k-1}))\} \ldots \{1-B(\alpha y_{k-1})\} \, d\, A_{j_1}(y_1) \ldots$$

$$d\, A_{j_{k-1}}(y_{j_{k-1}}) =$$

$$= \int_0^\infty \cdots \int_0^\infty \{1-B(\alpha(y_1+\ldots+y_{k-1}))\} \ldots \{1-B(\alpha y_{k-1})\} \, d\, U(y_1) \ldots d\, U(y_{k-1}) =$$

$$= c^{k-1} \alpha^{-\rho(k-1)} \int_0^\infty \cdots \int_0^\infty \{1-B(t_1+\ldots+t_{k-1})\} \ldots \{1-B(t_{k-1})\} \, d\, U^\alpha(t_1) \ldots d\, U^\alpha(t_{k-1}),$$

where we have used an obvious extension of the notation,

$$(2.20)\quad A_j(y) = P\{\underline{y}_1+\ldots+\underline{y}_j \leq y\} \qquad \text{for } j = 1,2, \ldots,$$

$$(2.21)\quad U(y) = \sum_{j=1}^{\infty} A_j(y)$$

and

$$(2.22)\quad U^\alpha(t) = \alpha^\rho c^{-1} U(\alpha^{-1} t).$$

Here again the product measure with respect to which we integrate in the last integral tends to a limit as α tends to infinity. Because of (1.2) there exists a $y_o > 0$ with

$$(2.23)\quad A_1(y) \leq 2\, c\, y^\rho \qquad \text{for } 0 \leq y < y_o .$$

By induction we can then prove

$$(2.24) \qquad A_j(y) \leq \frac{(2c\ \Gamma(\rho+1)\ y^{\rho})^j}{\Gamma(j\rho+1)} \text{ for } 0 \leq y < y_o \text{ and } j = 1,2, \ldots .$$

But then we easily show

$$(2.25) \qquad \lim_{\alpha\to\infty} U^{\alpha}(t) = t^{\rho} .$$

From (2.19) we now conclude

$$(2.26) \qquad \lim_{\alpha\to\infty} n_{\alpha}\ P(B_j^{\alpha}) = c^{k-1}\ I_{k-1}(\rho)$$

and our proof is complete.

Our observer can only notice light (at least k light quanta simultaneously alive) or the absence of light (less than k light quanta simultaneously alive). Hence we have to decide whether the occurrence of exactly i of the B_k^{α}, B_{k+1}^{α}, ..., $B_{n_\alpha}^{\alpha}$ leads the observer to register exactly i light sensations from the n_{α} light quanta. It turns out that in the limit the simplest possible situation occurs. There are then with probability 1 never k light quanta simultaneously alive from one arrival to the next, so every B_j^{α} that occurs is registered separately. This statement is an easy consequence of the following theorem.

Theorem 2.2

$$\lim_{\alpha\to\infty} P\{\text{at least one of the } B_k^{\alpha} \cap B_{k+1}^{\alpha},\ B_{k+1}^{\alpha} \cap B_{k+2}^{\alpha},\ \ldots \ldots,\ B_{n_\alpha-1}^{\alpha} \cap B_{n_\alpha}^{\alpha} \text{ occurs}\} = 0 .$$

Proof: we proceed as in the proof of theorem 2.1. Here $P(B_k^{\alpha} \cap B_{k+1}^{\alpha}) \leq$ $\leq P(B_{k+1}^{\alpha} \cap B_{k+2}^{\alpha}) \leq \ldots \leq P(B_{n_\alpha-1}^{\alpha} \cap B_{n_\alpha}^{\alpha})$ follows at once from the definition of these probabilities in terms of the y_1, s_1, y_2, s_2, ... So we need only prove

$$(2.27) \qquad \lim_{\alpha\to\infty} n_{\alpha}\ P(B_{n_\alpha-1}^{\alpha} \cap B_{n_\alpha}^{\alpha}) = 0.$$

With the * and D_i^{α} notation from the last proof we have for $k+1 \leq j \leq n_{\alpha}$

(2.28) $P(B^{\alpha}_{j-1} \cap B^{\alpha}_{j}) = P\{$at least k-1 of $\underset{\sim}{s}^{*}_{1} > \alpha\ \underset{\sim}{y}^{*}_{1}$, $\underset{\sim}{s}^{*}_{2} > \alpha\ (\underset{\sim}{y}^{*}_{1}+\underset{\sim}{y}^{*}_{2})$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{1}+\ldots+\underset{\sim}{y}^{*}_{j-1})$ and at least k-1 of $\underset{\sim}{s}^{*}_{2} > \alpha\ \underset{\sim}{y}^{*}_{2}$, $\underset{\sim}{s}^{*}_{3} > \alpha(\underset{\sim}{y}^{*}_{2}+\underset{\sim}{y}^{*}_{3})$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{2}+\ldots+\underset{\sim}{y}^{*}_{j-1})$ occur$\}$ =

$= P\{\underset{\sim}{s}^{*}_{1} \leq \alpha\ \underset{\sim}{y}^{*}_{1}$ and at least k-1 of $\underset{\sim}{s}^{*}_{2} > \alpha(\underset{\sim}{y}^{*}_{1}+\underset{\sim}{y}^{*}_{2})$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{1}+\ldots+\underset{\sim}{y}^{*}_{j-1})$ occur$\}$ +

$+ P\{\underset{\sim}{s}^{*}_{1} > \alpha\ \underset{\sim}{y}^{*}_{1}$ and at least k-2 of $\underset{\sim}{s}^{*}_{2} > \alpha(\underset{\sim}{y}^{*}_{1}+\underset{\sim}{y}^{*}_{2})$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{1}+\ldots+\underset{\sim}{y}^{*}_{j-1})$ and

at least k-1 of $\underset{\sim}{s}^{*}_{2} > \alpha\ \underset{\sim}{y}^{*}_{2}$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{2}+\ldots+\underset{\sim}{y}^{*}_{j-1})$ occur$\}$

$\leq P\{$at least k-1 of $\underset{\sim}{s}^{*}_{2} > \alpha(\underset{\sim}{y}^{*}_{1}+\underset{\sim}{y}^{*}_{2})$, ... , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{1}+ \ldots +\underset{\sim}{y}^{*}_{j-1})$ occur$\}$ +

$+ P\{\underset{\sim}{s}^{*}_{1} > \alpha\ \underset{\sim}{y}^{*}_{1}$ and at least k-1 of $\underset{\sim}{s}^{*}_{2} > \alpha\ \underset{\sim}{y}^{*}_{2}$, . , $\underset{\sim}{s}^{*}_{j-1} > \alpha(\underset{\sim}{y}^{*}_{2}+ \ldots +\underset{\sim}{y}^{*}_{j-1})$ occur$\}$ =

$= P\{$at least k-1 of $D^{\alpha}_{2}, D^{\alpha}_{3}, \ldots, D^{\alpha}_{j-1}$ occur$\} + P\{\underset{\sim}{s}_{1} > \alpha\ \underset{\sim}{y}_{1}\}\ P(B^{\alpha}_{j-1})$.

We already know that

(2.29) $\lim_{\alpha\to\infty} n_{\alpha}\ P\{\underset{\sim}{s}_{1} > \alpha\ \underset{\sim}{y}_{1}\}\ P(B^{\alpha}_{j-1}) = c^{k-1}\ I_{k-1}(\rho) \lim_{\alpha\to\infty} P\{\underset{\sim}{s}_{1} > \alpha\ \underset{\sim}{y}_{1}\} = 0$,

so we must now consider

(2.30) $P\{$at least k-1 of $D^{\alpha}_{2}, D^{\alpha}_{3}, \ldots, D^{\alpha}_{j-1}$ occur$\} \leq$

$$\leq \sum_{2\leq i_1<i_2<\ldots<i_{k-1}\leq j-1} P(D^{\alpha}_{i_1} \cap D^{\alpha}_{i_2} \cap \ldots \cap D^{\alpha}_{i_{k-1}}) \leq$$

$$\leq \sum_{j_1\geq 2, j_2\geq 1, \ldots, j_{k-1}\geq 1} P(D^{\alpha}_{j_1} \cap D^{\alpha}_{j_1+j_2} \cap \ldots \cap D^{\alpha}_{j_1+\ldots j_{k-1}}) =$$

$$= c^{k-1}\,\alpha^{-\rho(k-1)} \int_0^\infty \cdots \int_0^\infty \{1-B(t_1+\ \cdots\ +t_{k-1})\}\cdots\{1-B(t_{k-1})\}\,d\,U^\alpha(t_1)\cdots$$

$$\cdots\ d\,U^\alpha(t_{k-1}) +$$

$$- c^{k-1}\,\alpha^{-\rho(k-1)} \int_0^\infty \cdots \int_0^\infty \{1-B(t_1+\ \cdots\ +t_{k-1})\}\ \cdots$$

$$\cdots\ \{1-B(t_{k-1})\}\,d\,A^\alpha(t_1)\ d\,U^\alpha(t_2)\ \cdots\ d\,U^\alpha(t_{k-1}).$$

As both integrals have the same limit for $\alpha \to \infty$, our proof is complete.

3. CALCULATION OF LIMITS FOR NEW LIGHT FROM n_α LIGHT QUANTA

We need a generalization of a theorem from Watson [1954] for 1-dependent stationary sequences of events. His proof is generalized.

Assume that for each $\alpha > 0$ a sequence $E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha$ of stationary 1-dependent events is given with $m_\alpha \to \infty$ as $\alpha \to \infty$. The sequence is stationary for a fixed $\alpha > 0$ if for each j-tuple $i_1, i_2, \ldots, i_j$ with $1 \leq i_1 < i_2 < \ldots < i_j \leq m_\alpha$ (with arbitrary natural j)

$$(3.1) \qquad P(E_{i_1+h}^\alpha \cap E_{i_2+h}^\alpha \cap \ldots \cap E_{i_j+h}^\alpha) = P(E_{i_1}^\alpha \cap E_{i_2}^\alpha \cap \ldots \cap E_{i_j}^\alpha)$$

for all integers h with $1 - i_1 \leq h \leq m_\alpha - i_j$. The sequence is 1-dependent for a fixed $\alpha > 0$ if for each 2j-tuple $i_1, h_1, i_2, h_2, \ldots, i_j, h_j$ with $1 \leq i_1 < i_1 + h_1 < i_2 < i_2 + h_2 < \ldots < i_j < i_j + h_j \leq m_\alpha$ (with arbitrary natural $j \geq 2$), the smallest algebra 's $\mathfrak{F}_1, \mathfrak{F}_2, \ldots, \mathfrak{F}_j$ containing respectively $A_{i_1}, A_{i_1+1}, \ldots, A_{i_1+h_1}$; $A_{i_2}, A_{i_2+1}, \ldots, A_{i_2+h_2}$; $\ldots$; $A_{i_j}, A_{i_j+1}, \ldots, A_{i_j+h_j}$ are independent if $i_2 - i_1 - h_1 > 1$, $i_3 - i_2 - h_2 > 1$, $\ldots$, $i_j - i_{j-1} - h_{j-1} > 1$. Clearly 0-dependent is just another way of saying independent.

Theorem 3.1

If $E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha$ is for each $\alpha > 0$ a sequence of stationary 1-dependent events with $m_\alpha \to \infty$ for $\alpha \to \infty$,

$$(3.2) \qquad P(E_1^\alpha) = \frac{a}{m_\alpha} + o\left(\frac{1}{m_\alpha}\right) \text{ for } \alpha \to \infty$$

for a nonnegative constant a and

$$(3.3) \qquad \max_{2 \le j \le l+1} P(E_1^\alpha \cap E_j^\alpha) = o\left(\frac{1}{m_\alpha}\right) \text{ for } \alpha \to \infty,$$

then

$$(3.4) \qquad P_i^\alpha = P\{\text{exactly i of the } E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha \text{ occur}\}$$

satisfies

$$(3.5) \qquad \lim_{\alpha \to \infty} P_i^\alpha = \frac{a^i e^{-a}}{i!} \quad \text{for } i = 0, 1, 2, \ldots .$$

Proof: consider

$$(3.6) \qquad S_r^\alpha = \sum_{1 \le i_1 < i_2 < \ldots < i_r \le m_\alpha} P(E_{i_1}^\alpha \cap E_{i_2}^\alpha \cap \ldots \cap E_{i_r}^\alpha).$$

It is well-known (cf. Feller [1957], page 100) that

$$(3.7) \qquad P_i^\alpha = \sum_{r=i}^{m_\alpha} (-1)^{r-i} \binom{r}{i} S_r^\alpha$$

and

$$(3.8) \qquad S_r^\alpha = \sum_{i=r}^{m_\alpha} \binom{i}{r} P_i^\alpha ,$$

while for j = 0, 1, 2, ... (and say $i+2j+1 \le m_\alpha$)

$$(3.9) \qquad \sum_{r=i}^{i+2j+1} (-1)^{r-i} \binom{r}{i} S_r^\alpha \le P_i^\alpha \le \sum_{r=i}^{i+2j} (-1)^{r-i} \binom{r}{i} S_r^\alpha .$$

If it is known that

$$(3.10) \qquad \lim_{\alpha \to \infty} S_r^\alpha = \frac{a^r}{r!} \quad \text{for } r = 1, 2, \ldots ,$$

then(3.5) follows from the inequalities (3.9). We have $S_0^\alpha = 1$ by definition.

The number $m_\alpha(j,r)$ of r-tuples $i_1, i_2, \dots, i_r$ with $1 \le i_1 < i_2 < \dots < i_r \le m_\alpha$ and exactly j of the r-1 differences $i_2-i_1, i_3-i_2, \dots, i_r-i_{r-1}$ less than l+1 is $\binom{r-1}{j}$ times the number of this kind of r-tuples with moreover

$i_2-i_1 \le l, i_3-i_2 \le l, \dots, i_{j+1}-i_j \le l$ and hence (with $r_1 = i_2-i_1, \dots, r_j = i_{j+1}-i_j$)

$$(3.11)\quad m_\alpha(j,r) = \binom{r-1}{j} \sum_{r_1=1}^{l} \sum_{r_2=1}^{l} \dots \sum_{r_j=1}^{l} \sum_{\substack{1 \le i_1 < i_{j+2} < \dots < i_r \le m_\alpha \\ i_{j+2}-i_1 \ge l+1+r_1+\dots+r_j \\ i_{j+3}-i_{j+2} \ge l+1, \dots, i_r-i_{r-1} \ge l+1}} 1 =$$

$$= \binom{r-1}{j} \sum_{r_1=1}^{l} \sum_{r_2=1}^{l} \dots \sum_{r_j=1}^{l} \sum_{1 \le s_1 < s_2 < \dots < s_{r-j} \le m_\alpha - r_1 - \dots r_j - (r-j-1)l} 1 =$$

$$= \binom{r-1}{j} \sum_{r_1=1}^{l} \sum_{r_2=1}^{l} \dots \sum_{r_j=1}^{l} \binom{m_\alpha - r_1 - \dots - r_j - (r-j-1)l}{r-j}.$$

It is now clear that

$$(3.12)\quad m_\alpha(j,r) = \binom{r-1}{j} l^j \frac{m_\alpha^{r-j}}{(r-j)!} \left\{1 + O\left(\frac{1}{m_\alpha}\right)\right\} \text{ for } \alpha \to \infty \text{ and } j = 0,1,2,\dots, r-1 .$$

We split the summation in (3.6) with respect to the number $j(i_1,i_2,\dots,i_r)$, where $j(i_1,i_2,\dots,i_r) = j$ if exactly j of the differences $i_2-i_1, i_3-i_2, \dots, i_r-i_{r-1}$ are less than l+1. We have from (3.2) and (3.3)

$$(3.13)\qquad P(E_{i_1}^{\alpha} \cap E_{i_2}^{\alpha} \cap \ldots \cap E_{i_r}^{\alpha}) = \begin{cases} (\frac{a}{m_\alpha})^r + o(\frac{1}{m_\alpha^r}) & \text{for } j(i_1,i_2,\ldots,i_r) = 0, \\ o(\frac{1}{m_\alpha^{r-j}}) & \text{for } j(i_1,i_2,\ldots,i_r) = j \\ & \text{and } 1 \leq j \leq r-1. \end{cases}$$

Adding the contributions for different values of j to S_r^{α}, we find

$$(3.14)\qquad S_r^{\alpha} = \frac{a^r}{r!} + o(1) \quad \text{for } \alpha \to \infty \text{ and } r = 1,2, \ldots \ .$$

Our proof is now complete.

If we return to the sequence $A_k^{\alpha}, A_{k+1}^{\alpha}, \ldots, A_{n_\alpha}^{\alpha}$ from section 2, we can show that the conditions of Theorem 3.1 are satisfied.

Theorem 3.2

$$\lim_{\alpha\to\infty} P\{\text{exactly i of the } A_k^{\alpha}, A_{k+1}^{\alpha}, \ldots, A_{n_\alpha}^{\alpha} \text{ occur}\} = e^{-\mu}\frac{\mu^i}{i!}$$

for i = 0,1,2,... with

$$(3.15)\qquad \mu = c^{k-1} I_{k-1}(\rho).$$

Proof: for each $\alpha > 0$ the sequence $E_1^{\alpha}, E_2^{\alpha}, \ldots, E_{m_\alpha}^{\alpha}$ with $E_j^{\alpha} = A_{j+k-1}^{\alpha}$ and $m_\alpha = n_\alpha - k + 1$ is stationary and (k-2)- dependent. The theorem is trivial for k=2, so from now on we assume $k > 2$ in this proof. We know from (2.9) with μ as in (3 15)

$$(3.16)\qquad P(E_1^{\alpha}) = \frac{\mu}{m_\alpha} + o(\frac{1}{m_\alpha}) \quad \text{for } \alpha \to \infty.$$

Furthermore for $2 \leq j \leq k-1$ we find with the by now well-known technique

(3.17)

$$P(E_1^{\alpha} \cap E_j^{\alpha}) = P\{\underline{s}_1 > \alpha(\underline{y}_2+\ldots+\underline{y}_k), \underline{s}_2 > \alpha(\underline{y}_3+\ldots+\underline{y}_k), \ldots, \underline{s}_{j-1} > \alpha(\underline{y}_j+\ldots+\underline{y}_k),$$

$$\underline{s}_j > \alpha(\underline{y}_{j+1}+\ldots+\underline{y}_{k+j-1}),\ \underline{s}_{j+1} > \alpha(\underline{y}_{j+2}+\ldots+\underline{y}_{k+j-1}),\ldots$$

$$\ldots,\ \underline{s}_{k+j-2} > \alpha\ \underline{y}_{k+j-1}\Big\} =$$

$$= \int_0^\infty \ldots \int_0^\infty \{1-B(\alpha(y_2+\ldots+y_k))\}\ldots\{1-B(\alpha(y_j+\ldots+y_k))\}\ \{1-B(\alpha(y_{j+1}+\ldots$$

$$\ldots+y_{k+j-1}),\}\ldots\{1-B(\alpha y_{k+j-1})\}\ d\,A(y_2)\ d\,A(y_3)\ldots d\,A(y_{k+j-1}) =$$

$$= c^{k+j-2}\ \alpha^{-\rho(k+j-2)} \int_0^\infty \ldots \int_0^\infty \{1-B(t_2+\ldots+t_k)\}\ldots\{1-B(t_{k+j-1})\}\ d\,A^\alpha(t_2)\ldots$$

$$\ldots\ d\,A^\alpha(t_{k+j-1}) =$$

$$= O(\alpha^{-\rho(k+j-2)}) \qquad \text{for } \alpha \to \infty.$$

Hence

$$(3.18)\qquad \max_{2\le j\le k-1} P(E_1^\alpha \cap E_j^\alpha) = O(\alpha^{-\rho k})\ o(\frac{1}{m_\alpha}) \quad \text{for } \alpha \to \infty.$$

Our proof is now complete.

Watson's theorem was generalized in Newell [1964] in a different direction from ours. Slightly restricting the generality of the last reference, we have the following theorem.

Theorem 3.3

If $E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha$ is for each $\alpha > 0$ a sequence of stationary 1-dependent events with $m_\alpha \to \infty$ for $\alpha \to \infty$,

$$(3.19)\qquad P(E_1^\alpha \cap \overline{E_2^\alpha} \cap \ldots \cap \overline{E_{l+1}^\alpha}) = \frac{a}{m_\alpha} + o(\frac{1}{m_\alpha}) \text{ for } \alpha \to \infty$$

for a nonnegative constant a and

$$(3.20)\qquad P(E_1^\alpha) = O(\frac{1}{m_\alpha}) \quad \text{for } \alpha \to \infty,$$

then

$$(3.21)\qquad \lim_{\alpha\to\infty} P\{\text{none of the the } E_1^\alpha\ E_2^\alpha, \ldots\ E_{m_\alpha}^\alpha \text{ occurs}\} = e^{-a}.$$

We shall prove a more detailed version of this theorem (although it is not needed at present), which is also a generalization of Theorem 3.1.

Theorem 3.4.

If $E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha$ is for each $\alpha > 0$ a sequence of stationary 1-dependent events with $m_\alpha \to \infty$ for $\alpha \to \infty$ and

$$(3.22)\quad P(E_{i_1}^\alpha \cap E_{i_2}^\alpha \cap \ldots \cap E_{i_j}^\alpha) = \frac{a_j(i_2-i_1, i_3-i_2, \ldots, i_j-i_{j-1})}{m_\alpha} + o(\frac{1}{m_\alpha})$$

for $\alpha \to \infty$ for each j-tuple $i_1, i_2, \ldots, i_j$ with $1 \leq i_1 < i_2 < \ldots < i_j \leq l+1$ and $1 \leq j \leq l+1$ with nonnegative constants $a_j(i_2-i_1, i_3-i_2, \ldots, i_j-i_{j-1})$, then

$$(3.23)\quad P_i^\alpha = P\{\text{exactly i of the } E_1^\alpha, E_2^\alpha, \ldots, E_{m_\alpha}^\alpha \text{ occur}\}$$

satisfies for $i = 0,1,2, \ldots$

$$(3.24)\quad P_i \overset{\text{def}}{=} \lim_{\alpha\to\infty} P_i^\alpha = \frac{e^{b_0}}{i!} \left(\left(\frac{d}{dw}\right)^i \exp\left(\sum_{j=1}^{l+1} b_j w^j\right)\right)_{w=0},$$

where (with $a_0 = 0$ and $a_1 = a_1(\,)$)

$$(3.25)\quad b_h = \sum_{j=h}^{l+1} (-1)^{j-h} \binom{j}{h} a_j \quad \text{for } h = 0,1,\ldots, l+1$$

and

$$(3.26)\quad a_j = \sum_{\substack{l_1, l_2, \ldots, l_{j-1} \geq 1 \\ l_1+l_2+\ldots+l_{j-1} \leq l}} a_j(l_1, l_2, \ldots, l_{j-1}) \quad \text{for } j = 2,3,\ldots, l+1.$$

The $b_1, b_2, \ldots, b_{l+1}$ are nonnegative and $b_0 = -(b_1+b_2+\ldots+b_{l+1})$.

Proof: relation (3.22) holds for all $i_1, i_2, \ldots, i_j$ with $1 \leq i_1 < i_2 < \ldots < i_j \leq m_\alpha$

and $1 \leq j \leq m_\alpha$, as in those cases which were not yet considered we always have $i_j - i_1 > l$ and hence

(3.27) $0 \leq P(E_{i_1}^{\alpha} \cap E_{i_2}^{\alpha} \cap \ldots \cap E_{i_j}^{\alpha}) \leq P(E_{i_1}^{\alpha} \cap E_{i_j}^{\alpha}) = \left\{ \frac{a_1}{m_\alpha} + o(\frac{1}{m_\alpha}) \right\}^2 =$

$= o(\frac{1}{m_\alpha})$ for $\alpha \to \infty$.

We can thus define

(3.28) $a_j(i_2-i_1,\ldots,i_j-i_{j-1}) = 0$ if $j > l+1$ or $i_j-i_1 > l$.

We first compute

(3.29) $$S_r = \lim_{\alpha\to\infty} S_r^{\alpha},$$

where S_r^{α} is given by (3.6).

Consider a fixed sequence of indices $i_1, i_2, \ldots, i_r$ with $1 \leq i_1 < i_2 < \ldots < i_r \leq m_\alpha$.
Write it down from left to right in increasing order. Place partitions in the r-1 spaces between i_1 and i_2, i_2 and i_3, $\ldots$, i_{r-1} and i_r according to the rule: put in a partition between i_{h-1} and i_h if $i_h - i_{h-1} > l$, do not put in a partition otherwise. Do this for each $h \in \{2,3,\ldots, r\}$. If at least one partition is needed, we call the indices to the left of the first partition (counting partitions from the left), as well as those between two successive partitions and those to the right of the last partition, a group of indices; if no partition is needed, we call all indices together a group of indices. Each such group contains at least one and at most r indices. Let the chosen sequence $i_1, i_2, \ldots, i_r$ contain β_1 groups with one index, β_2 groups with 2, $\ldots$, β_r groups with r indices. Then

(3.30) $$\beta_1 + 2\beta_2 + \ldots + r\beta_r = r.$$

If $(j_1,j_2,\ldots,j_h)$ is a group with $h > 1$ indices, the differences $j_2-j_1, j_3-j_2, \ldots, j_h-j_{h-1}$ have definite values. Let $\beta_h(l_1, l_2, \ldots, l_{h-1})$ be the number of groups $(j_1,j_2,\ldots,j_h)$ contained in $i_1, i_2, \ldots, i_r$ with h indices and $j_2-j_1 = l_1$, $j_3-j_2 = l_2$, $\ldots$, $j_h-j_{h-1} = l_{h-1}$. Then

$$(3.31)\qquad \beta_h = \sum_{l_1=1}^{l} \sum_{l_2=1}^{l} \cdots \sum_{l_{h-1}=1}^{l} \beta_h(l_1,l_2,\ldots,l_{h-1}) \text{ for } h = 2,3,\ldots,r.$$

We first compute the contribution to S_r^{α} from those sequences of indices that have the same group characteristics from left to right as $i_1, i_2, \ldots, i_r$. Thus the value for $\beta_1, \beta_2(1), \beta_2(2), \ldots, \beta_2(l), \beta_3(1,1), \beta_3(1,2), \ldots \beta_3(l,l), \beta_4(1,1,1), \ldots, \beta_r(1,1,\ldots,1)$ is the same for these sequences and the j^{th} group from the left contains in each of these sequences the same number of indices with the same successive differences in the same order. As there are $\beta_1 + \beta_2 + \cdots + \beta_r$ groups, there are

$$(3.32)\qquad \frac{m_\alpha^{\beta_1+\beta_2+\ldots+\beta_r}}{(\beta_1+\beta_2+\ldots+\beta_r)!}\left(1 + O\left(\frac{1}{m_\alpha}\right)\right) \quad \text{for } \alpha \to \infty$$

different sequences of this kind (the exact answer is a binomial coefficient as in (3.11)). Each sequence of this kind contributes

$$(3.33)\qquad \frac{a_1^{\beta_1}\left(\prod_{l_1=1}^{l} a_2(l_1)^{\beta_2(l_1)}\right)\cdot\left(\prod_{l_1,l_2,\ldots,l_{r-1}=1}^{l} a_r(l_1,l_2,\ldots,l_{r-1})^{\beta_r(l_1,l_2,\ldots,l_{r-1})}\right)}{m_\alpha^{\beta_1+\beta_2+\ldots+\beta_r}} + o\left(\frac{1}{m_\alpha^{\beta_1+\beta_2+\ldots+\beta_r}}\right) \text{ for } \alpha \to \infty.$$

The total contribution from this kind of sequence follows from multiplying (3.32) by (3.33). Next we allow the groups with the same number of indices to change places, obliging us to multiply further by a factor

$$(3.34)\qquad \frac{\beta_2!}{\beta_2(1)!\ldots\beta_2(l)!}\cdot\frac{\beta_3!}{\beta_3(1,1)!\ldots\beta_3(l,l)!}\cdot\ldots\cdot\frac{\beta_r!}{\beta_r(1,1,\ldots,1)!\ldots\beta_r(l,l,\ldots,l)!}$$

to obtain the total contribution under the new restriction. Now we allow the group characteristics to vary without changing $\beta_1, \beta_2, \ldots, \beta_r$ or the number of indices in each group as they occur from left to right. By now the total contribution to S_r^α is

$$(3.35)\quad \frac{a_1^{\beta_1}(a_2(1)+\ldots+a_2(1))^{\beta_2}\ldots(a_r(1,1,\ldots,1)+\ldots+a_r(1,1\ldots,1))^{\beta_r}}{(\beta_1+\beta_2+\ldots+\beta_r)!} +$$

$+ o(1)$ for $\alpha \to \infty$.

Next we allow the groups with different numbers of indices to change places. Now we have to multiply by

$$(3.36)\quad \frac{(\beta_1+\beta_2+\ldots+\beta_r)!}{\beta_1!\beta_2!\ldots\beta_r!} .$$

Finally we add all contributions for different values of $\beta_1, \beta_2, \ldots, \beta_r$. We obtain (with (3.26) extended to arbitrary j)

$$(3.37)\quad S_r^\alpha = \sum_{\substack{\beta_1\geq 0,\beta_2\geq 0,\ldots,\beta_r\geq 0 \\ \beta_1+2\beta_2+\ldots+r\beta_r=r}} \frac{a_1^{\beta_1}}{\beta_1!}\cdot\frac{a_2^{\beta_2}}{\beta_2!}\cdot\ldots\cdot\frac{a_r^{\beta_r}}{\beta_r!} + o(1) \text{ for}$$

$\alpha \to \infty$.

As $a_r = 0$ for $r > l+1$, we finally obtain

$$(3.38)\quad S_r = \sum_{\substack{\beta_1\geq 0,\beta_2\geq 0,\ldots,\beta_{l+1}\geq 0 \\ \beta_1+2\beta_2+\ldots+(l+1)\beta_{l+1}=r}} \frac{a_1^{\beta_1}}{\beta_1!}\frac{a_2^{\beta_2}}{\beta_2!}\cdot\ldots\cdot\frac{a_{l+1}^{\beta_{l+1}}}{\beta_{l+1}!} \text{ for}$$

$r = 1,2,\ldots$.

With $S_0 = 1$ as before, we find for the generating function

$$(3.39)\quad S(z) = \sum_{r=0}^{\infty} S_r z^r$$

for all finite values of z

$$(3.40) \qquad S(z) = \exp \sum_{j=1}^{l+1} a_j z^j .$$

From the inaqualities (3.9) follows easily, that $\lim_{\alpha\to\infty} P_i^{\alpha}$ exists with

$$(3.41) \qquad P_i = \left(\left(\frac{d}{dz}\right)^i S(z)\right)_{z=-1} \text{ for } i = 0,1,2, \ldots .$$

But then for all finite values of w

$$(3.42) \quad P(w) \overset{\text{def}}{=\!=} \sum_{i=0}^{\infty} P_i w^i = \sum_{i=0}^{\infty} \frac{w^i}{i!} S^i(-1) = S(w-1) .$$

If now (with $a_0 = 0$ by definition)

$$(3.43) \quad b_i = \sum_{j=1}^{l+1} (-1)^{j-i} \binom{j}{i} a_j \quad \text{for } i = 0,1,2, \ldots \quad l+1 ,$$

then for all finite values of w

$$(3.44) \qquad \sum_{j=0}^{l+1} a_j (w-1)^j = \sum_{j=0}^{l+1} b_j w^j$$

and

$$(3.45) \qquad P(w) = \exp \sum_{j=0}^{l+1} b_j w^j .$$

We have now shown that (3.24) holds. The relation $b_0 = -(b_1+b_2+\ldots+b_{l+1})$ follows from (3.44) by taking w = 1.

Because $P(1) = S(0) = 1$ the P_i form a probability distribution, say of a random variable $\underline{i}$. This distribution is now shown to be infinitely divisible. Consider $E_1^{\alpha}, E_2^{\alpha}, \ldots, E_{\left[\frac{m_\alpha}{j}\right]}^{\alpha}$ for an integer $j \geq 2$. Then by what we have already proved, the limit of

$$(3.46) \qquad P_{i\,j}^{\alpha} = P\{\text{exactly } i \text{ of } E_1, E_2^{\alpha}, \ldots, E_{\left[\frac{m_\alpha}{j}\right]}^{\alpha} \text{ occur}\}$$

for $\alpha \to \infty$ exists. If we take

$$(3.47) \qquad P_j(w) = \sum_{i=0}^{\infty} w^i \lim_{\alpha\to\infty} P_{i,j}^{\alpha} ,$$

then $P_j(w)$ is again the generating function of a probability distribution and also

$$(3.48) \qquad P_j(w) = \exp \sum_{h=0}^{l+1} \frac{b_h}{j} w^j ,$$

because (3.22) now applies with m_α replaced by $j[\frac{m_\alpha}{j}]$. But then $P(w) = \{P_j(w)\}^j$ and $\underline{i}$ is infinitely divisible. From Feller [1957], page 271, we then conclude that $b_1, b_2, \ldots, b_{l+1}$ are nonnegative. This last result is not trivial, because there exist probability distributions with generating function $\exp \sum_{j=0}^{l} c_j w^j$ with two or more negative c_j (cf. Levy [1937]). Our proof is now complete.

4. LIGHT SENSATIONS IN A FIXED TIME-INTERVAL

Now that we know what happens when a fixed number n_α of light quanta reaches the retina of our observer's eye, it is not hard to extend the results to what happens in a fixed time-interval $(o, t_\alpha]$ with a conveniently chosen t_α for $\alpha \to \infty$

. The expected number of arrivals in $(O, t_\alpha]$ is given by

$$(4.1) \qquad \sum_{n=1}^{\infty} n\, P\{\alpha(\underline{y}_1+\ldots+\underline{y}_n) \leq t_\alpha < \alpha(\underline{y}_1+\ldots+\underline{y}_{n+1})\} = \sum_{n=1}^{\infty} P\{\alpha(\underline{y}_1+\ldots+\underline{y}_n) \leq t_\alpha\} = U(\frac{t_\alpha}{\alpha}) .$$

If this expectation is to be close to n_α and as is known from renewal theory

$$(4.2) \qquad U(t) = \frac{t}{\mathcal{E}\underline{y}} + o(t) \quad \text{for } t \to \infty,$$

it is convenient to assume

$$\lim_{\alpha\to\infty} t_\alpha \alpha^{-\rho(k-1)-1} = \mathcal{E}_{\underline{y}} . \tag{4.3}$$

For any ε with $0 < \varepsilon < 1$ we have

$$\lim_{\alpha\to\infty} P\left\{\underline{y}_1+\ldots+\underline{y}_{\lfloor(1-\varepsilon)n_\alpha\rfloor} \leq \frac{t_\alpha}{\alpha}\right\} = 1 \tag{4.4}$$

and

$$\lim_{\alpha\to\infty} P\left\{\underline{y}_1+\ldots+\underline{y}_{\lfloor(1+\varepsilon)n_\alpha\rfloor} \leq \frac{t_\alpha}{\alpha}\right\} = 0 \tag{4.5}$$

by the weak law of large numbers for identically distributed independent random variables with first moment.

If we write again $\underline{t}_n$ for $\underline{y}_1 + \underline{y}_2 + \ldots + \underline{y}_n$ as we did in section 1, we note that

$$\begin{aligned} &P\{\text{at least i light sensations in } (0,t_\alpha]\} = \\ &= P\{\text{at least i light sensations in } (0,t_\alpha] \text{ and } \underline{t}_{\lfloor(1+\epsilon)n_\alpha\rfloor} > \alpha^{-1}t_\alpha\} + \\ &+ P\{\text{at least i light sensations in } (0,t_\alpha] \text{ and } \underline{t}_{\lfloor(1+\epsilon)n_\alpha\rfloor} \leq \alpha^{-1}t_\alpha\} \leq \\ &\leq P\{\text{at least i light sensations in } (0,\underline{t}_{\lfloor(1+\epsilon)n_\alpha\rfloor}]\} + \\ &+ P\{\underline{t}_{\lfloor(1+\varepsilon)n_\alpha\rfloor} \leq \alpha^{-1}t_\alpha\} . \end{aligned} \tag{4.6}$$

But then we know from the theory developed in sections 2 and 3 and (4.5)

$$\begin{aligned} &\limsup_{\alpha\to\infty} P\{\text{at least i light sensations in } (0,t_\alpha]\} \leq \\ &\leq \lim_{\alpha\to\infty} P\{\text{at least i light sensations in } (0,\underline{t}_{\lfloor(1+\varepsilon)n_\alpha\rfloor}]\} = \\ &= \sum_{j=i}^{\infty} e^{-(1+\varepsilon)\mu}\frac{\{(1+\varepsilon)\mu\}^j}{j!} \qquad \text{for } i = 0,1,2, \ldots \end{aligned} \tag{4.7}$$

with $0 < \varepsilon < 1$, ε otherwise arbitrary and $\mu = c^{k-1} I_{k-1}(\rho)$

To find a lower bound we note that

$$(4.8)\quad P\{\text{at least } i \text{ light sensations in } (0,t_\alpha]\} =$$

$$= P\{\text{at least } i \text{ light sensations in } (0,t_\alpha] \text{ and } \underline{t}_{[(1-\varepsilon)n_\alpha]} \leq \alpha^{-1}t_\alpha\} +$$

$$+ P\{\text{at least } i \text{ light sensations in } (0,t_\alpha] \text{ and } \underline{t}_{[(1-\varepsilon)n_\alpha]} > \alpha^{-1}t_\alpha\} \geq$$

$$\geq P\{\text{at least } i \text{ light sensations in } (0,\underline{t}_{[(1-\varepsilon)n_\alpha]}] \text{ and } \underline{t}_{[(1-\varepsilon)n_\alpha]} \leq \alpha^{-1}t_\alpha\} \geq$$

$$\geq P\{\text{at least } i \text{ light sensations in } (0,\underline{t}_{[(1-\varepsilon)n_\alpha]}]\} - P\{\underline{t}_{[(1-\varepsilon)n_\alpha]} > \alpha^{-1}t_\alpha\}.$$

With (4.4) we conclude

$$(4.9)\quad \liminf_{\alpha\to\infty} P\{\text{at least } i \text{ light sensations in } (0,t_\alpha]\} \geq$$

$$\geq \lim_{\alpha\to\infty} P\{\text{at least } i \text{ light sensations in } (0,\underline{t}_{[(1-\varepsilon)n_\alpha]}]\} =$$

$$= \sum_{j=i}^{\infty} e^{-(1-\varepsilon)\mu} \frac{\{(1-\varepsilon)\mu\}^i}{i!} \quad \text{for } i = 0,1,2,\ \ldots$$

and arbitrary ε with $0 < \varepsilon < 1$.

Combining our results we have proved the following theorem.

Theorem 4.1.

$$\lim_{\alpha\to\infty} P\{\text{exactly } i \text{ light sensations in } (0,t_\alpha]\} = e^{-\mu}\frac{\mu^i}{i!}$$

for $i = 0,1,2,\ \ldots$ with μ as in (3.15).

It is not hard to extend this result to a finite number of non-overlapping intervals.

LITERATURE

Feller [1957], An Introduction to Probability Theory and its Applications, Volume 1, Wiley (second edition),

ten Hoopen en Reuver [1965], On a waiting time problem in physiology, Stat Neerl. 19 (1965) 27-34,

Ikeda [1965], On Bouman-Velden-Yamamoto's asymptotic evaluation formula for the probability of visual response in a certain experimental research in quantum-biophysics of vision, Ann. Inst. Stat. Math. 17 (1965) 295-310,

Lévy [1937], Sur les exponentielles de polynomes et sur l'arithmétique des produits de lois de Poisson, Annales Ecole Normale Supérieure 54 (1937) 231-292,

Meyer [1967], Some Poisson-type limi theorems for sequences of dependent rare events, with applications, Institute of Statistics Mimeo Series No. 529, Univ. of North Carolina,

Newell [1964], Asymptotic extremes for m-dependent random variables, Ann. Math. Stat. 35 (1964) 1322-1325,

Watson [1954], Extreme values in samples from m-dependent stationary stochastic processes, Ann. Math. Stat. 25 (1954) 798-800.

SOME RECENT DEVELOPMENTS IN RELIABILITY THEORY

by Richard E. Barlow (USA)
University of California, Berkeley

1. MODELS FOR LIFE DISTRIBUTIONS

Reliability theory is largely concerned with questions about coherent structures, i.e., structures which can be represented as combinations of various series and parallel networks (allowing the possibility of component replication). Various attempts have been made to delimit the class of life distributions of interest for such structures. Extreme value theory provides an answer, in part, to the question: Given a coherent structure with a finite number of components and some procedure to increase the number of components without bounds, what are the possible limiting distributions for the structure lifetime for given component lifetime distributions? A series structure has a lifetime corresponding to the minimum of its component lifetimes and the limiting procedure adds one component at a time to the structure. The only possible relevant limiting distributions for this case are the Weibull and double exponential distributions [12]. If we add k-out-of n structures to the structures of interest we obtain the normal and lognormal distributions and distributions of the form

$$G_{\alpha,k}(x) = 0 \qquad \text{for } x \leq 0$$

$$= \frac{1}{(k-1)!} \int_0^{x^\alpha} e^{-y} y^{k-1} dy \qquad \text{for } x > 0,\ \alpha > 0$$

$$\Lambda_k(x) = \frac{1}{(k-1)!} \int_0^{e^x} e^{-y} y^{k-1} dy \qquad -\infty < x < \infty$$

[8], [27]. Results for more general models have been obtained by Harris [14].

There is a vast literature concerned with the estimation of parameters for distributions in these classes. For a recent comprehensive survey of results for the Weibull distribution, see Mann [18].

We might call the preceding approach "the limiting distribution approach." A more recent approach to delimit the class of life distributions of interest has been to start with the exponential distribution and to ask the question: What is the smallest family of life distributions containing the exponential distribution which is closed under the formation of coherent structures and limits in distribution? This question was recently answered by Birnbaum, Esary and Marshall[5]. The class of distributions in question is precisely the class with increasing failure rate on the average (IFRA distributions), i.e., $\int_0^t r(x)dx/t$ is nondecreasing in $t \geq 0$ where $r(t)$ is the failure rate function. Research on estimation procedures for IFRA distributions and tests of hypotheses concerning such distributions is still underway [3], [9], [10].

A great deal of attention has been focused on models for life distributions based on properties of the failure rate function. For example, we could consider the class of distributions having failure rate $r(t)$ for which $r(t)/\psi(t)$ is increasing (or convex increasing) for specified $\psi(t) > 0$. If $\psi(t)$ is constant, we have the class of IFR (for increasing failure rate) distributions. The case of more general $\psi(t)$ has been investigated by Saunders [26] in connection with models for fatigue failure.

Some multivariate models have been proposed to describe the joint lifetimes of components in a coherent structure. The multivariate exponential distribution studied by Marshall and Olkin [19], [20] generalizes the univariate exponential distribution in a natural way. Harris [15] has proposed a class of multivariate distributions with IFR marginals which satisfy additional reasonable restrictions.

Esary, Proschan and Walkup [11] study a new concept of positive

dependency describing the joint lifetimes of components in a coherent structure. They say that random variables $T_1, T_2, \ldots, T_n$ are *associated* if

$$\mathrm{Cov}\{f(\underline{T}), g(\underline{T})\} \geq 0$$

for all nondecreasing functions f and g where $\underline{T} = (T_1, T_2, \ldots, T_n)$. Random variables jointly distributed according to the multivariate exponential distribution, for example, are associated. This concept is partly motivated by the idea that working components tend to reinforce the contribution of each other to system performance. Stated in another way, a failed component may, if anything, stress the remaining working components so as to cause them to perform poorly. Also, components subject to a common environment tend to be associated. They obtain lower bounds on system reliability in terms of component marginal reliabilities when components have associated lifetimes. Lehmann [17] considers related ideas of positive dependency for the bivariate case and examines tests of independence versus positive dependency.

2. TESTS OF HYPOTHESES

Considerable attention has been focused on the problem of testing the validity of the IFRA and IFR models [1], [3], [24]. The problem usually posed is that of testing

$$H_o : F \text{ exponential}$$

versus (1)

$$H_1 : F \text{ IFRA (or IFR)}.$$

If $G(x) = 1 - e^{-\lambda x}$ for $x \geq 0$ and $\lambda > 0$, then $G^{-1}F(x)/x$ is nondecreasing in $x \geq 0$ if and only if F is IFRA. This observation suggests that we consider the following partial ordering on the space of life distributions; namely, $F_1 \underset{*}{<} F_2$ if $F_2^{-1}F_1(x)/x$ is nondecreasing in $x \geq 0$.

Note that $F_1 \underset{*}{\leq} F_2 \underset{*}{\leq} G$ suggests that a test for (1) should have greater power at F_1 than at F_2 no matter what significance level we choose for our test. Marshall, Walkup and Wets [22] have characterized the class of differentiable test statistics which produce tests having monotone power with respect to $\underset{*}{\leq}$ ordering. These are just the tests based on functions $h(x_1, x_2, \ldots, x_n)$ having the properties:

(a) h is homogeneous;

(b) $$\sum_{i=0}^{j} x_{n-i} \frac{\partial h(x_1 \ldots, x_n)}{\partial x_{n-i}} \geq 0 \text{ for } j = 0,1, \ldots, n-2$$

and all $0 \leq x_1 \leq x_2 \leq \ldots \leq x_n$. The test associated with h would reject exponentiality if $h(X_1, X_2 \ldots, X_n) \geq c$ where c is some suitable critical number and $0 \leq X_1 \leq X_2 \leq \ldots \leq X_n$ are order statistics from F For example,

$$h(X_1, \ldots, X_n) = \sum_{1}^{n} (X_i - \bar{X})^2/(\bar{X})^2$$

where $\bar{X} = \sum_{1}^{n} X_i/n$ is such a function.

Let n(u) equal the number of items exposed to risk at time u and $T(X_i) = \int_0^{X_i} n(u)du$. Tests based on

$$h(X_1, \ldots, X_n) = \sum_{i=1}^{n-1} \Delta_i T(X_i)/T(X_n) \qquad (2)$$

for $\Delta_i \geq 0$ are unbiased and have monotone power with respect to $\underset{*}{\leq}$ ordering. These tests reject exponentiality in favor of IFRA for large values of $h(X_1, X_2, \ldots, X_n)$. This statistic has a natural analogue for the case of incomplete data [2]. The corresponding test remains unbiased but we can no longer prove that it has monotone power with respect to $\underset{*}{\leq}$ ordering.

Tests based on (2) have been extensively studied by Bickel and Doksum [3]. They call such tests studentized linear spacings tests and consider the following "pencils' of alternative densities to the exponential:

$$f_\theta^{(1)}(x) = \left[1 + \theta(1 - e^{-x})\right] \exp\left\{-\left[x + \theta(x + e^{-x} - 1)\right]\right\}$$

$$f_\theta^{(2)}(x) = (1 + \theta x) \exp\left\{-(x + \tfrac{1}{2}\theta x^2)\right\} \qquad \text{(Linear Failure Rate)}$$

$$f_\theta^{(3)}(x) = (1 + \theta)x^\theta \exp\left\{-x^{(1+\theta)}\right\} \qquad \text{(Weibull)}$$

$$f_\theta^{(4)}(x) = \left[x^\theta e^{-x}\right]/\Gamma(1 + \theta) \qquad \text{(Gamma)}$$

For each density $x \geq 0$, $\theta \geq 0$, and the null hypothesis is obtained for $\theta = 0$. The asymptotically best studentized linear spacings test against $f_\theta^{(1)}$ corresponds to weights $\Delta_i = 1$ $(i = 1,2, \ldots, n-1)$. This statistic is also called the *total time on test* statistic. The asymptotic Pitman efficiency of this test compared to the asymptotically most powerful test for this reduced problem is only $\frac{1}{4}$. However, there is a great deal of evidence supporting the "robustness" of this test. The asymptotically best studentized linear spacings test for $f_\theta^{(3)}$ --the Weibull density-- seems to be a robust competitor to the total time on test statistic. They also consider the corresponding asymptotically best linear rank tests which are asymptotically equivalent to the best linear spacings tests. However, the Monte Carlo power of the asymptotically best linear spacings tests is much superior for small sample sizes.

The Monte Carlo power of the total time on test statistic is computed in [1] against the likelihood ratio test for truncated exponentiality versus IFR distributions. Again, the total time on test statistic seems to be decidedly superior.

In [28], van Soest studied an omnibus Cramer-Von Mises-Smirnov type of statistic. The statistic is

$$C_n = n \int_0^\infty \left\{F_n(x) - \hat{F}(x)\right\}^2 \, d\hat{F}(x)$$

$$= 1/12n + \sum_{j=1}^{n} \left\{\hat{F}(x_j) - \frac{2j - 1}{n}\right\}^2$$

where F_n is the empirical distribution and $\hat{F}$ is the maximum likelihood estimate of F under H_o. The null hypothesis is rejected for large values of C_n. He computes the Monte Carlo power of this test and compares it with a two sided test based on the total time on test statistic. The power curves below compare this test with the one sided test based on the total time on test statistic (2).

Doksum [10] has recently investigated the two sample problem for IFRA distributions. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be two independent random samples from populations with continuous IFRA distributions $F(.)$ and $F(./\Delta)$, respectively, and let $s_1, \ldots, s_n$ denote the ranks of the Y's in the combined sample. For testing $H_o : \Delta \leq 1$ versus $H_1 : \Delta > 1$, it is shown that the error probabilities of each monotone rank test ϕ are bounded by the error probabilities for exponential alternatives, i.e., if $G(t) = 1 - \exp(-t)$, $t \geq 0$, then

$$P[\text{reject } H_o \mid F(\cdot), F(\cdot/\Delta)] \leq P[\text{reject } H_o \mid G(\cdot), G(\cdot/\Delta)]$$

for $\Delta \leq 1$, and

$$P[\text{accept } H_o \mid F(\cdot), F(\cdot/\Delta)] \leq P[\text{accept } H_o \mid G(\cdot), G(\cdot/\Delta)]$$

for $\Delta > 1$.

The Savage test which rejects for small values of

$$S_N = \sum_{i=1}^{n} J_o(s_i) \quad \text{where } J_o(k) = \sum_{j=N+1-k}^{N} \frac{1}{j} \qquad (N = m+n)$$

is locally minimax for IFRA scale alternatives within the class of rank tests. For the two sample problem with indifference region, i.e., $H_o : \Delta \leq 1$ versus $H_1 : \Delta \geq \Delta_1 > 1$, the Lehmann test which rejects for small values of

$$L_N = \sum_{j=1}^{N} \log\left[j - \frac{(\Delta_1 - 1)}{\Delta_1} s'_j \right]$$

where s'_j is the number of Y's greater than or equal to the (N+1-j)th order statistic in the combined sample, is minimax at $\Delta = \Delta_1$ within the class of rank tests for IFRA scale alternatives.

In an earlier paper [9], Doksum considers the asymptotic efficiency of the best test for exponential models relative to the Savage, S_N, test. For the exponential distribution, the uniformly most powerful level α test ϕ_N^* of $H_o : \Delta = 1$ against $H_1 : \Delta > 1$ rejects when

$$T = m^{-1} \sum_{i=1}^{m} X_i / n^{-1} \sum_{i=1}^{n} Y_i > F_{2m,2n}(\alpha)$$

where $F_{2m,2n}(\alpha)$ is obtained from the tables of the F distribution with 2m and 2n degrees of freedom. Doksum considers the modified test $\hat{\phi}$ so as to have an asymptotic level α test when the distribution is IFRA and not exponential. The test $\hat{\phi}$ rejects H_o for large values of

$$N^{\frac{1}{2}}(T - 1)/(\hat{\sigma}/\hat{\mu})$$

where

$$\hat{\mu} = N^{-1}\left(\sum_{1}^{n} X_i + \sum_{1}^{n} Y_i\right)$$

and

$$(\hat{\sigma})^2 = N^{-1}\left(\sum_{1}^{n} X_i^2 + \sum_{1}^{n} Y_i^2\right) - (\hat{\mu})^2$$

The efficiency of the Savage test with respect to $\hat{\phi}$ goes from 1 to ∞ as the shape parameter, b, of the Weibull distribution, $G(t) = 1 - \exp[-\lambda t^b]$, goes from 1 to ∞ (or from 1 to 0). He also treats the case of censored sampling.

3. ESTIMATION PROCEDURES

The failure rate function is perhaps the most useful characterization of a life distribution. Parametric and nonparametric methods for estimating the failure rate are discussed in detail by Grenander [13]. He also characterizes the maximum likelihood estimate (MLE) of the failure rate function under the IFR assumption. The MLE can be easily computed even for very incomplete data (withdrawals may be allowed for example). If a total of n items are exposed to risk, failures are observed at times

$$Z_1 \leq Z_2 \leq \ldots \leq Z_k \qquad (k \leq n)$$

and n(u) is the number of items exposed to risk at time u, then the MLE estimate for the failure rate, r(t), can be expressed as a step function, where

$$\hat{r}_n(t) = \begin{cases} 0 & 0 \leq t < Z_1 \\ \hat{r}_n(Z_i) & Z_i \leq t \leq Z_{i+1} \\ \infty & t \geq Z_k \end{cases}$$

and

$$\hat{r}_n(Z_i) = \underset{s \leq i}{\text{Max}} \ \underset{t \geq i+1}{\text{Min}} \left\{ \frac{t - s}{\sum_{j=s}^{t-1} \int_{Z_j}^{Z_{j+1}} n(u)\,du} \right\}.$$

Marshall and Proschan [21] proved that $\hat{r}_n(t)$ is strongly consistent in the complete sample case. Since the life distribution is determined by the failure rate, we can also determine the MLE for the life distribution under the IFR assumption. Monte Carlo investigations, however, indicate that it is badly biased in the tails. For samples of size 100 or so, the empirical distribution appears to be a better estimate of the life distribution in the tails.

Rao [25] has characterized the limiting distribution of $\hat{r}_n(t)$ assuming r(t) increasing and r'(t) > 0. He shows that

$$\mathcal{L}\left[n^{1/3}\left\{\frac{r'(t)r^2(t)}{2f(t)}\right\}^{-1/3}\{\hat{r}_n(t) - r(t)\}\right] \to H(x)$$

where H(x) is a distribution whose density is determined implicitly as a solution to the heat equation. This result enables us to make asymptotic efficiency comparisons with other nonparametric estimators of the failure rate.

Parzen [23] and Weiss and Wolfowitz [30] investigate window estimators of the density f(t) , i.e.,

$$\bar{f}_n(t) = N/2n\varepsilon_n$$

where N is the number of observations out of a sample of size n in $(t - \varepsilon_n , t + \varepsilon_n)$. If $\varepsilon_n = n^{-\alpha}$ and we assume only that f'(t) exists, then Weiss and Wolfowitz show that $\alpha = 1/3$ provides an asymptotically efficient estimate in a certain sense. A natural nonparametric estimate of the hazard rate is then

$$\bar{r}_n(t) = \frac{\bar{f}_n(t)}{\int_t^\infty \bar{f}_n(x)dx} .$$

Watson and Leadbetter [29] show that

$$\mathcal{L}\left[\left[1 - F(t)\right]\left[2n\varepsilon_n/f(t)\right]^{\frac{1}{2}}\left[\bar{r}_n(t) - r(t)\right]\right] \to \Phi(x)$$

where $\Phi(x)$ is the N(0,1) distribution. Following Hodges and Lehmann [4], we define the asymptotic efficiency of $\bar{r}_n(t)$ relative to $\hat{r}_n(t)$ as

$$e(\bar{r}_n(t),\hat{r}_n(t)) = \frac{\left[\frac{r'(t)r^2(t)}{2f(t)}\right]^{2/3}\sigma_o^2}{\left[\frac{f(t)}{2}\right]\left[\frac{1}{1-F(t)}\right]^2}$$

where σ_o satisfies

$$\inf_{\sigma} d[\Phi(x),H(x/\sigma)] = d[\Phi(x),H(x/\sigma_o)]$$

and

$$d[\Phi(x),H(x)]_\infty = \sup_{-\infty<x<\infty} |\Phi(x) - H(x)| .$$

In this computation, we have let $\varepsilon_n = n^{-1/3}$ for the window estimator It seems clear from (3) that the MLE estimator will do best when $r'(t) \neq 0$ and the window estimator will be better when r'(t) is very large, or in other words, when the failure rate is increasing very rapidly.

Since $\bar{r}_n(t)$ is not necessarily increasing, a more acceptable competitor to $\hat{r}_n(t)$ under the IFR assumption would be

$$\hat{\bar{r}}_n(t_i) = \underset{s\leq i}{\text{Max}}\ \underset{w\geq i}{\text{Min}} \left\{\frac{\sum_{j=s}^{w} \bar{r}_n(t_j)}{w - s + 1}\right\}$$

which is nondecreasing in t_i . ($0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq \ldots$ are midpoints of intervals over which the window estimator is constant) The results of Brunk [7] show that $\hat{\bar{r}}_n(t)$ will inherit the consistency property of $\bar{r}_n(t)$ and will improve on $\bar{r}_n(t)$ in a least squares sense.

Similar nonparametric MLE estimates for U-shaped failure rate functions have been considered by Bray, Crawford and Proschan [6] . There is a very useful general smoothing technique behind all of these estimates. Brunk [7] has observed that $\hat{r}_n(Z_i)$ is the 'isotonic regression"

on the index i of the sequence

$$r_n^{*}(Z_i) = 1/\int_{Z_i}^{Z_{i+1}} n(u)du$$

with respect to a suitable measure, namely,

$$\mu\{i\} = \int_{Z_i}^{Z_{i+1}} n(u)du$$

in this case. The isotonic regression $\hat{r}_n(Z_i)$ has a useful least squares property; namely,

$$\sum_{i=1}^{n-1}\left[r_n^{*}(Z_i) - r(Z_i)\right]^2 \mu\{i\}$$

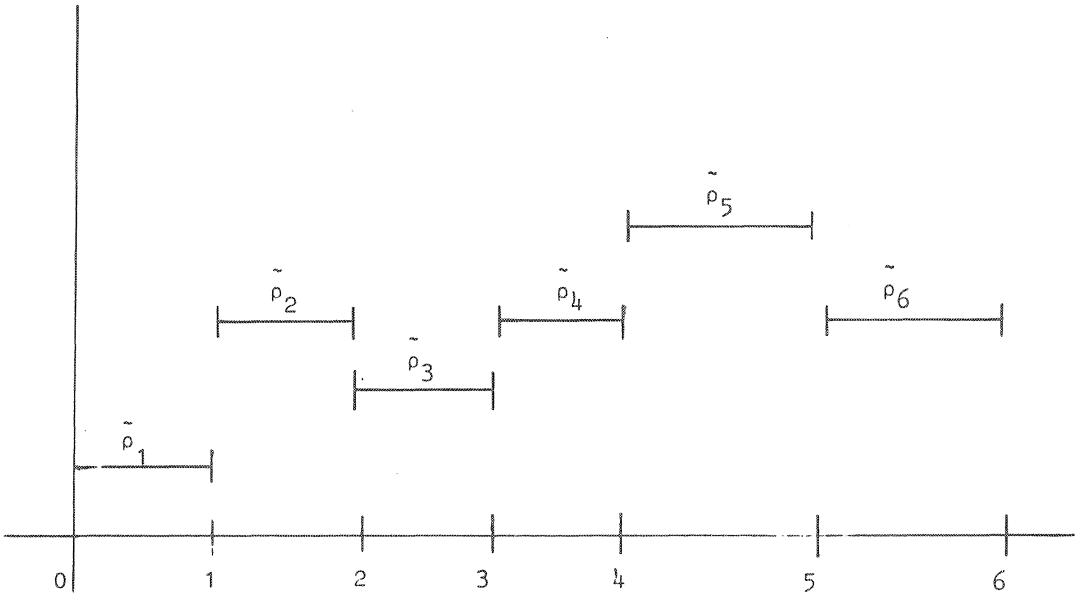
$$\geq \sum_{i=1}^{n-1}\left[r^{*}(Z_i) - \hat{r}_n(Z_i)\right]^2 \mu\{i\} + \sum_{i=1}^{n-1}\left[\hat{r}_n(Z_i) - r(Z_i)\right]^2 \mu\{i\}$$

holds for every increasing function $r(Z_i)$.

To illustrate another use of the isotonic regression, consider the following estimation problem. Suppose that items are inspected at random times and that failure is observed only through inspection. Our failure data, on each item which fails, consists of intervals (t_i, t_{i+1}) where it is only known that the item survived to time t_i and failed sometime in the interval (t_i, t_{i+1}) . Harris, Meier and Tukey [16] study the MLE estimate of the failure rate under these conditions. The approach is non-parametric in that it is assumed only that the failure rate is constant (but unknown) over specified time intervals. The fact that the authors seek rates rather than probabilities for intervals produces certain differences in their treatment of observations extending over parts of intervals as compared to the "actuarial" treatment. The isotonic regression technique can be applied to the Harris, Meier, Tukey estimate,

for example, if we assume that the failure rate first decreases and then increases.

Let $\tilde{\rho}_i$ denote the Harris-Meier-Tukey MLE of the failure rate in interval i . Let (0,1) denote the first interval; (1,2) the second interval, etc. Suppose the MLE looks as follows:



To obtain an estimate, say $\hat{\rho}$, which decreases and then increases and which is closest to $\tilde{\rho}$ in a least squares sense, proceed as follows:

(1) Choose in turn, $i_o = 0,1,2, \ldots , n$ for possible change points of the failure rate.

(2) Suppose $i_o = 0$. Then

$$\hat{\rho}_{1,0} = \underset{v \geq 1}{\text{Min}} \frac{\sum_{j\ 1}^{v} \tilde{\rho}_j}{v + 1 - 1}$$

and

$$\hat{\rho}_{i,0} = \underset{v \geq i}{\text{Min}} \; \underset{u \leq i}{\text{Max}} \; \frac{\sum_{j=u}^{v} \tilde{\rho}_j}{v + 1 - u} \quad .$$

(3) If $0 < i_o < n$, then

(3a) $$\hat{\rho}_{i,i_o} = \underset{i_o - 1 \geq v \geq i}{\text{Max}} \; \underset{u \leq i}{\text{Min}} \; \frac{\sum_{j=u}^{v} \tilde{\rho}_j}{v + 1 - u} \quad \text{for } 1 \leq i \leq i_o - 1$$

and

(3b) $$\hat{\rho}_i = \underset{v \geq i}{\text{Min}} \; \underset{i_o \leq u \leq i}{\text{Max}} \; \frac{\sum_{j=u}^{v} \tilde{\rho}_j}{v + 1 - u} \quad \text{for } i_o \leq i \leq n \; .$$

(4) For each choice i_o, compute

$$\sum_{i=1}^{n} \left(\tilde{\rho}_i - \hat{\rho}_{i,i_o} \right)^2 \; .$$

(5) Take as your estimate that $\hat{\rho}_{i,i_o}$ corresponding to the change point which minimizes the sum of squares.

It can be shown that the estimate $\hat{\rho}_{i,i_o}$ so obtained is closest to the Harris-Meier-Tukey estimate $\tilde{\rho}_i$ in the sense that it minimizes the sum of squared errors relative to all competing estimates which first decrease and then increase.

REFERENCES

[1] Barlow, R.E., Likelihood Ratio Tests for Restricted Families of Probability Distributions, Annals of Mathematical Statistics, Vol. 39, pp. 547-560, (1968).

[2] Barlow, R.E. and F. Proschan, A Note on Tests for Monotone Failure Rate Based on Incomplete Data, ORC 68-7, Operations Research Center, University of California, Berkeley, (1968).

[3] Bickel, P.J. and K. Doksum, Tests for Monotone Failure Rate Based on Normalized Spacings, ORC 67-37, Operations Research Center, University of California, Berkeley, (1967).

[4] Bickel, P.J. and J.L. Hodges, The Asymptotic Theory of Galton's Test and a Related Simple Estimate of Location, Annals of Mathematical Statistics, Vol. 38, pp. 73-89, (1967).

[5] Birnbaum, A.W., J.E. Esary and A.W. Marshall, A Stochastic Characterization of Wear-Out for Components and Systems, Annals of Mathematical Statistics, Vol. 37, pp. 816-825, (1966).

[6] Bray, T.A., G.B. Crawford and F. Proschan, Maximum Likelihood Estimation of a U-Shaped Failure Rate Function, Boeing Scientific Research Laboratories Document D1-82-0660, (1967).

[7] Brunk, H.D., Conditional Expectation Given a Lattice and Applications, Annals of Mathematical Statistics, Vol. 36, pp. 1339-1350, (1965).

[8] Chibisov, D.M., On Limit Distributions for Order Statistics, Theory of Probability Applications, Vol. 9, pp. 142-148, (1964).

[9] Doksum, K., Asymptotically Optimal Statistics in Some Models with Increasing Failure Rate Averages, Annals of Mathematical Statistics, Vol. 38, pp. 1731-1739, (1967).

[10] Doksum, K., Minimax Results for IFRA Scale Alternatives, ORC 67-31, Operations Research Center, Univer ity of California, Berkeley, (1967).

[11] Esary, J.D., F. Proschan and D.W. Walkup, Association of Random Variables with Applications, Anna s of Mathematical Statistics, Vol. 38, pp. 1466-1474.

[12] Gnedenko, B.V., Sur la Distribution Limite du Terme Maximum d'une Serie Aléatoire, Annals of Mathematics, Vol. 44, pp. 423-454, (1943).

[13] Grenander, Ulf, On the Theory of Mortality Measurement, Part II, Skand. Aktuarietidskr.,Vol. 39, pp 125-153, (1956).

[14] Harris, R., Contributions to the Theory of Extreme Values, ORC 68-9, Operations Research Center, Univers ty of California, Berkeley, (1968).

[15] Harris, R., A Multivariate Definition f r Increasing Hazard Rate Distribution Functions and Some Related Multivariate Dependence Relations, ORC 68-15, Operations Research Center, University of California, Berkeley, (1968).

[16] Harris, T.E , P. Meier and J.W. Tukey, Timing of the Distribution of Events Between Observations, Human Biology, Vol. 22, pp. 249-270, (1950).

[17] Lehmann, E.L., Some Concepts of Dependence, Annals of Mathematical Statistics, Vol. 37, pp. 1137-1153.

[18] Mann, N., Point and Interval Estimation Procedures for the Two-Parameter Weibull and Extreme-Value Distributions, Technometrics, Vol. 10, pp. 231-256, (1968).

[19] Marshall, A.W. and I. Olkin, A Multivariate Exponential Distribution, Journal of the American Statistical Association, Vol. 62, pp. 30-44, (1967).

[20] Marshall, A.W. and I. Olkin, A Generalized Bivariate Exponential Distribution, Journal of Applied Probability, Vol. 4, pp. 291-302, (1967).

[21] Marshall, A.W. and F. Proschan, Maximum Likelihood Estimation for Distributions with Monotone Failure Rate, Annals of Mathematical Statistics, Vol. 36, pp. 69-77, (1965).

[22] Marshall, A.W., D.W. Walkup and R.J.B. Wets, Order-Preserving Functions; Applications to Majorization and Order Statistics, Pacific Journal of Mathematics, Vol. 23, pp. 569-584, (1967).

[23] Parzen, E., On Estimation of a Probability Density and Mode, Annals of Mathematical Statistics, Vol. 33, p 1065, (1962).

[24] Pyke, R., Spacings, Journal of the Royal Statistical Society, Series B, Vol. 27, pp. 395-449, (1965).

[25] Rao, B.L.S. Prakasa, Asymptotical Distributions in Some Nonregular Statistical Problems, Technical Report No. 9, Department of Statistics and Probability, Michigan State University, (9 April 1966).

[26] Saunders, S., On the Determination of a Safe Life for Classes of Distributions Classified by Failure Rate, Technometrics, Vol. 10, pp. 361-377, (1968).

27 Smirnov, N.V., Limit Distributions for the Terms of a Variational Series, American Mathematical Society Transactions, Series I, No. 67, (1952).

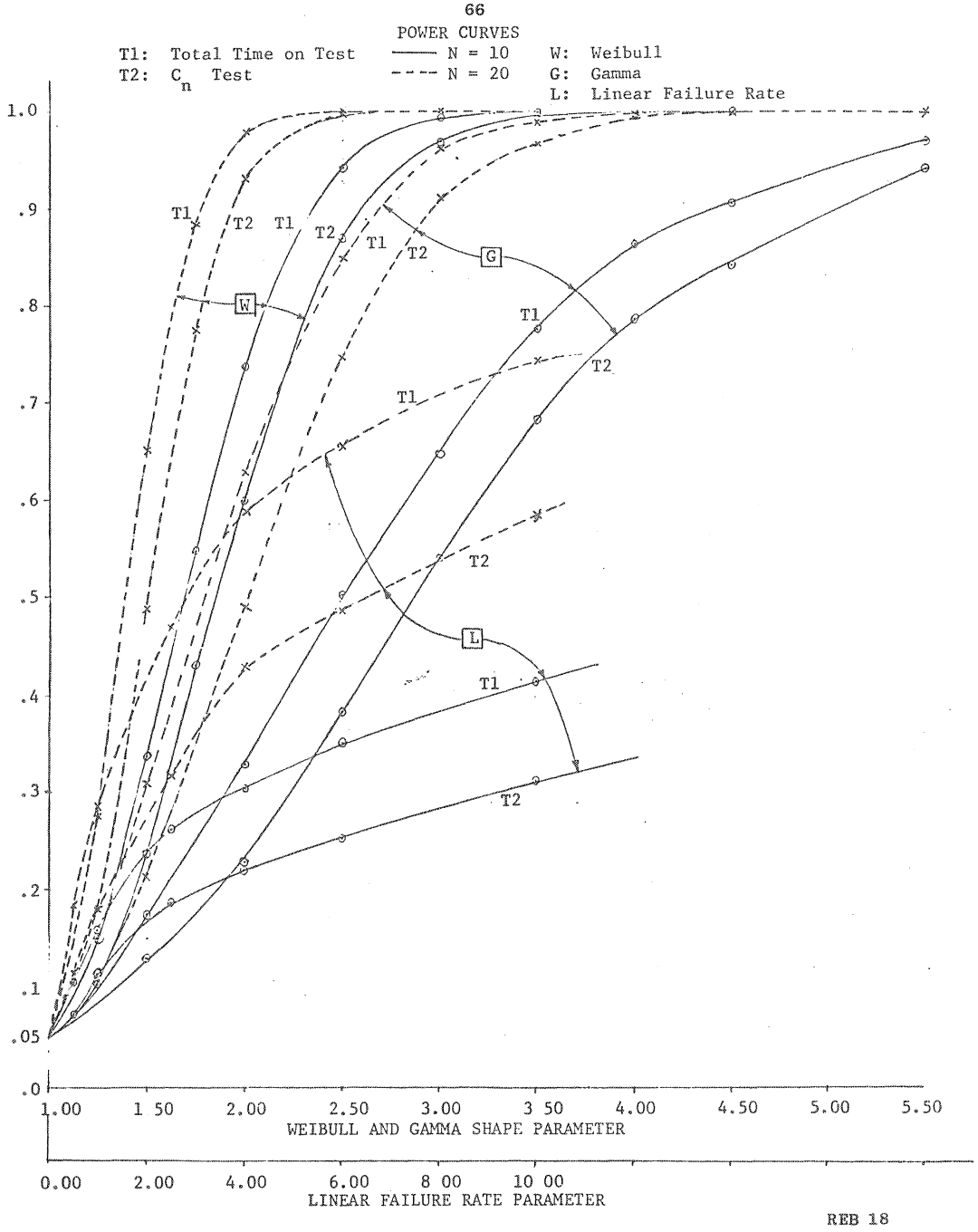
28 van Soest, J., Some Goodness of Fit Tests for Exponential Distributions, Technological University, Delft, Netherlands, mimeographed.

29 Watson, G.S. and M.R. Leadbetter, Hazard Analysis II, Sankhyã, Vol. 26, pp. 101-116, (1964).

30 Weiss, L. and J. Wolfowitz, Estimation of a Density Function at a Point, Z. Wahrscheinlichkeitstheorie verw. Geb., Vol. 7, pp. 327-335, (1967).

ACKNOWLEDGEMENTS

The author would like to acknowledge very helpful conversations with H.D. Brunk, F. Proschan, K. Doksum and, especially, W. van Zwet in connection with the section on estimation procedures. Vsant Ubhaya computed the power curves displayed in Section 2.


POWER CURVES
T1: Total Time on Test
T2: C_n Test
N = 10
N = 20
W: Weibull
G: Gamma
L: Linear Failure Rate
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.05
.0
T1
T2
W
G
L
1.00
1 50
2.00
2.50
3.00
3 50
4.00
4.50
5.00
5.50
WEIBULL AND GAMMA SHAPE PARAMETER
0.00
2.00
4.00
6.00
8 00
10 00
LINEAR FAILURE RATE PARAMETER

REB 18

FINITE AND ASYMPTOTIC OPTIMALITY OF RANK TESTS

by H. Witting (W. Germany)
University of Münster (Westf.)

It is well known, that there are two main kinds of nonparametric tests: rank tests and permutation tests. Although similar or even stronger statements can be made for permutation tests, to-day I will restrict myself to rank tests.

Rank tests originally were introduced by more or less intuitive reasoning; later on the question was raised, in what respect these procedures can be justified by mathematical reasoning; i.e. whether these rank tests can be derived in a certain sense as optimal procedures. There are two main directions of doing this, originated by Hoeffding [3] and Hájek [1], respectively.

Hoeffding [3] derived the general most powerful rank test against a certain alternative and pointed out how to simplify the problem by regarding alternatives in the neighbourhood of the hypotheses; in particular he indicated how to come to the Fisher-Yates test. Later on Terry and Lehmann among others characterized the most important rank tests like Wilcoxon test, Spearman test, Wilcoxon's matched pairs signed rank test and so on in this way as "locally most powerful" against an appropriate class of alternatives. Here "locally most powerful" means, that these tests maximize the slope of the power function at the hypothesis, differentiated along an appropriate one-parametric subclass of alternatives. More precisely, it was proved that these well known rank tests are optimal among the *class of all rank tests*.

Whereas the just described approach to optimality is based on a fixed, *finite* sample size Hajek [1] - using LeCam's [4] concept of contiguity - discussed asymptotic properties and proved, that the above mentioned rank tests are *asymptotically optimal* among the *class of all tests* (of the same asymptotic level) against an appropriate class of

alternatives, which is shown to be strongly related to that class of alternatives, against which the rank test is "locally most powerful".

To-day I will make some remarks concerning these both directions of research and the relations between them. Because of scarcity of time let me concentrate to some special questions which in my opinion have up to now not been discussed in a satisfactory manner or are at least known enough. In this connection I should mention the names of my co-workers G.Nölle and K.Behnen, who did research in this field and to whom I am also obliged for a couple of stimulating discussions during the preparation of this paper.

To avoid the introduction of an appropriate, sufficiently general terminology and framework, let me restrict myself to a special case: The well known comparison of two independent samples. A complete and unified treatment of optimality of rank tests including for instance this two sample problem, the matched pairs problem (tests of symmetry) and the tests of independence (rank correlation) you will find in a forthcoming book [7] of Mr. Nölle and mine.

Let us use the following terminology:

$z := (x,y) := (x_1,\ldots,x_k,y_1,\ldots,y_m)$	sample of k+m observations (x_i corresponding treatment I, y_j corresponding treatment II).
$z_{[]} := (z_{[1]},\ldots,z_{[k+m]})$	ordered pooled sample
$X_1,\ldots,X_k,Y_1,\ldots,Y_m$	corresponding random variables,
$Z_{[\ ]} := (Z_{[1]},\ldots,Z_{[k+m]})$	i.e. $X_1,\ldots,X_k$, $Y_1,\ldots Y_m$ independent r.v.'s with continuous distribution functions F and G, respectively

$\theta = (F,G)$	parameter of the common distribution of the r.v.'s.
$w_\theta = \mathcal{L}_\theta(X_1,\ldots,X_k,Y_1,\ldots,Y_m)$	common distribution of the r.v.'s under $\theta = (F,G)$
$(r,s) := (r_1,\ldots,r_k,s_1,\ldots,s_m)$	sample of ranks of $x_1,\ldots,x_k,y_1,\ldots,y_m$ with respect to the pooled sample
$R_1,\ldots,R_k,S_1,\ldots,S_m$	corresponding r.v.'s
$\phi(x,y) = \psi(r,s)$	rank tests, i.e. tests which depend only on ranks
$H : F \dot{\geq} G$	hypothesis to be tested at significance level α
$K : F \dot{\leq} G$	alternative. Here $F \dot{\leq} G$ means, that the r.v.'s X_i are stochastically larger than the r.v.'s Y_j, i.e. that treatment I is better than treatment II
$J : F = G$	boundary of the hypotheses H and K
$K_1 \subset \tilde{K}_1 \subset K$	$\tilde{K}_1$ is a given subclass of K, against which the test is locally optimal; then K_1 is the subclass of $\tilde{K}_1$, against which the test is shown to be optimal
$\bar{r}-\bar{s}$, $\sum r_i$, $\sum r_i - \sum s_j$	possible expressions for Wilcoxon's test statistic
$\forall\, n \in \mathbb{N}$	for all integers n = 1,2,...

As is well known optimality for finite sample sizes is proved in two steps. The first step justifies mathematically the replacement of observations by their ranks. This can be done by means of a reduction by invariance. The argument is as follows: Let τ be an arbitrary continuous strictly increasing transformation of $\mathbb{R}_1$ to $\mathbb{R}_1$. Since the observations x_i, y_j are as 'informative" as the observations τx_i, τy_j we restrict ourselves to tests, which are invariant against such transformations. As is discussed in the book of Lehmann [5], the class of all these invariant tests is exactly the class of the above defined rank tests.

The second step characterizes the Wilcoxon test or Fisher-Yates test among the class of all rank tests of the same level α; for this the rank test of level α is determined which is optimal against a specific alternative $\theta_\Delta \in \tilde{K}_1$, where $\tilde{K}_1$ is an one-parametric subclass of distributions w_{θ_Δ}, $0 < \Delta < \Delta_1$, dominated by w_{θ_o}, which may or may not depend on some nuisance parameters. Then the best level α rank test against the alternative w_{θ_Δ}, $0 < \Delta < \Delta_1$, is a solution ψ_Δ^* of

(1) $\quad E_{\theta_o} \psi = \alpha$

(2) $\quad E_{\theta_\Delta} \psi = \sup \quad , \qquad \theta_\Delta \in \tilde{K}_1 \quad .$

Here the first condition is equivalent to $E_\theta \psi = \alpha \quad \forall \theta \in J$, because the distribution of the rank statistic (R,S) under each $\theta \in J$ is the discrete uniform distribution over the $(k+m)!$ possible values of (r,s). The distribution of (R,S) under $\theta_\Delta = (F_\Delta, F_o) \in K_1$ is given by Hoeffding's formula

$$w_{\theta_\Delta}(R=r, S=s) = \frac{1}{(k+m)!}\, \pi_\Delta(r,s) \quad ,$$

$$(3) \qquad \pi_\Delta(r,s) := E_{w_o} \frac{p_\Delta(Z_{[r_1]}) \cdots p_\Delta(Z_{[r_k]})}{p_o(Z_{[r_1]}) \cdots p_o(Z_{[r_k]})} \quad ,$$

$$p_\Delta := \frac{dw_{\theta_\Delta}^{X_i}}{d\mu} = \text{density of } X_i \text{ under } \theta_\Delta \quad .$$

Therefore, according to the Neyman-Pearson Lemma the critical region of the best rank test ψ_Δ^* is

$$\pi_\Delta(r,s) > c\,(\Delta) \quad , \tag{4}$$

i.e. consists of those $[\alpha(k+m)!]$ points which have the largest value of the test statistic $\pi_\Delta(r,s)$. (In case that these $[\alpha(k+m)!]$ points are not uniquely determined, we have to randomize appropriately for $\pi_\Delta\,(r,s) = c\,(\Delta)$.)

It is well known in the literature (cf, for instance, Hájek [2]) that for certain "natural" levels of significance, namely $\frac{j}{(k+m)!}$, $(j=0,1,\ldots,(k+m)!$, but not necessarily all of them), and under some regularity conditions (which ensure among others the exchangeability of differentiation with respect to Δ and integration with respect to x) this test ψ_Δ^* is independent of Δ for $0 < \Delta < \Delta_1$ and is identical with the usually used "locally most powerful" test with the critical region

$$\tilde{\pi}(r,s) > \tilde{c} \quad , \quad \tilde{\pi}(r,s) := \frac{\partial}{\partial\Delta}\ \pi_\Delta\,(r,s)\Big|_{\Delta = 0} \quad , \tag{5}$$

namely the Wilcoxon test and the Fisher-Yates test for translation alternatives with F_0 corresponding to the logistic distribution and the normal distribution, respectively, and the Wilcoxon test for nonparametric alternatives.

On the other hand, the following simple example shows that ψ_Δ^* and $\tilde{\psi}$ are not necessarily identical for other suitable levels of significance, although ψ_Δ^* is independent of Δ for $0 < \Delta < \Delta_1$.

Example: Let $\tilde{K}_1$ be a class of nonparametric alternatives $(F_\Delta, F) \in K$, $F_\Delta := (1-\Delta)\,F+\Delta F^2$, $\Delta \in (0,1)$ and let $k=m=2$, $n:=k+m=4$. Then $(r,s) = (r_1, r_2, s_1, s_2)$ and

$$\pi_\Delta(r,s)=1+2\Delta\left(\frac{r_1+r_2}{n+1}-1\right)+\Delta^2\left(4\,\frac{r_1(r_2+1)}{(n+1)(n+2)}-2\,\frac{r_1+r_2}{n+1}+1\right)$$

$$\tilde{\pi}(r,s)=2\left(\frac{r_1+r_2}{n+1}-1\right)\quad\text{(Wilcoxon test statistic)}\quad .$$

Because of sufficiency, the ordering of (r_1,r_2) and of (s_1,s_2) is irrelevant, i.e. there are only $\binom{n}{k}=\binom{4}{2}=6$ essentially different points (r,s), characterized already by the ordered ranks $r_{[1]}$ and $r_{[2]}$ of r_1 and r_2, respectively.

$(r_{[1]},r_{[2]})$	$\pi_\Delta(r,s)$	$\tilde{\pi}(r,s)$
(1,2)	$1-\frac{4}{5}\Delta+\frac{1}{5}\Delta^2$	$-\frac{4}{5}$
(1,3)	$1-\frac{2}{5}\Delta+\frac{1}{15}\Delta^2$	$-\frac{2}{5}$
(1,4)	$1-\frac{1}{3}\Delta^2$	0
(2,3)	$1+\frac{1}{15}\Delta^2$	0
(2,4)	$1+\frac{2}{5}\Delta-\frac{1}{15}\Delta^2$	$+\frac{2}{5}$
(3,4)	$1+\frac{4}{5}\Delta+\frac{1}{5}\Delta^2$	$+\frac{4}{5}$

From this table you can see the following: Whereas the 6 pairs $(r_{[1]},r_{[2]})$ are strictly ordered by $\pi_\Delta(r,s)$ for $0<\Delta\leq 1$, this is not true for the test statistic $\tilde{\pi}(r,s)$, because $\tilde{\pi}(r,s)$ does not distinguish between the rank tuples (1,4) and (2,3). This implies that for $1/3<\alpha<2/3$ the Wilcoxon test is not locally uniformly most powerful.

Hence, for an arbitrary level of significance, there are two questions:

1) When does exist a $\Delta_1>0$, such that ψ^*_Δ is independent of Δ, $0<\Delta<\Delta_1$?

2) Which is the relation between ψ^*_Δ , $0 < \Delta < \Delta_1$, and $\tilde{\psi}$?

Under the condition, that $\pi_\Delta(r,s)$ can be expanded into a powerseries about $\Delta=0$ and differentiation and integration are exchangeable (this condition is somewhat stronger than the condition in the literature), Dr. Nölle gave a quite simple proof of the fact, that for *every* level of significance there exists a number $\Delta_1 > 0$, such that ψ^*_Δ is independent of Δ, $0 < \Delta < \Delta_1$. This means, that under these stronger conditions $\psi^* := \psi^*_\Delta$, $0 < \Delta < \Delta_1$, is locally uniformly most powerful against $\tilde{K}_1$, i.e. uniformly most powerful against an appropriate subclass $K_1 \subset \tilde{K}_1$.

Moreover, he could prove that the region of randomization is enlarged, when the uniformly most powerful test ψ^* is replaced by the usually used "locally most powerful" test $\tilde{\psi}$. In other words, if the test $\tilde{\psi}$ decides uniquely, then does the test ψ^*. Of course, this proof shows also that $\tilde{\psi}$ and ψ^* are identical for the above mentioned "natural" levels α.

The imposed stronger conditions can be proved to be valid for those classes $\tilde{K}_1$ which are used in connection with the Wilcoxon test and the Fisher-Yates test, i.e. for translation alternatives of normal and logistic distributions as well as for nonparametric alternatives used in the above example.

As indicated at the beginning of this paper asymptotically a stronger statement concerning optimality can be proved than in the case of a fixed sample size, namely the optimality among the class of all tests and not only among the class of all rank tests. This implies, that the "loss of information" by using ranks instead of the original observations is asymptotically negligible.

The fact, that asymptotically stronger statements can be proved, is familiar to statisticians, for instance in connection with comparing two means based on independent normally distributed r.v.'s with the same variance. In case that $\sigma^2=\sigma_o^2$ is known, for testing H:a $\leq$ b against K:a > b there exists an uniformly most powerful test, namely the Gauss test

with the test statistic

$$\sqrt{\frac{km}{k+m}}\ \frac{\bar{x}-\bar{y}}{\sigma_o} \quad .$$

But in applications usually the value of σ^2 is not known. Therefore, instead of Gauss's test usually Student's t-test is used with the test statistic

$$\sqrt{\frac{km}{k+m}}\ \frac{\bar{x}-\bar{y}}{s} \quad , \qquad s^2 := \frac{1}{k+m-2}\left(\sum (x_i-\bar{x})^2 + \sum (y_j-\bar{y})^2\right) \quad ,$$

which does not depend on the assumption of a known value $\sigma^2=\sigma_o^2$. Since s^2 is "less informative" than σ^2, Student's t-test is not as good as the Gauss test in case that σ^2 is known. But if k and m tend to infinity, the sample variance s^2 converges in probability (or even almost surely) to the population variance σ^2; therefore, it can be expected that asymptotically Student's t-test is as good as the Gauss test.
For making this precise, let us define two sequences of tests ϕ_n, ϕ_n^* of the same level of significance to be asymptotically equivalent if $E_{\theta_n}(\phi_n - \phi_n^*) \to 0$ holds for all (or an appropriate class of) sequences $\{\theta_n\} \subset K$ with $\theta_n \to \theta \in J$. This is fulfilled for the Gauss test ϕ^*_{n,σ_o} and the t-test ϕ_n for *all* sequences $\theta_n = (a_n, b_n, \sigma_o^2) \in K$ with $\theta_n \to \theta \in J$, i.e. with $a_n - b_n \to 0$, $a_n > b_n$, which can easily be verified. Here it should be mentioned, that $E_{\theta_n}(\phi_n - \phi_n^*) \to 0$ imposes the strongest condition (i.e. is most informative) for sequences θ_n with $a_n - b_n = \eta\sqrt{\frac{km}{k+m}} + o\left(\sqrt{\frac{km}{k+m}}\right)$, for which the power tends to β, $\alpha < \beta < 1$.

Hence the situation is as follows: for fixed $\sigma^2 = \sigma_o^2$ we have a uniformly most powerful test, namely the Gauss test ϕ^*_{n,σ_o} which depends on σ_o^2. On the other hand, the t-test ϕ_n, which is asymptotically equivalent (in the indicated sense) to the Gauss test ϕ^*_{n,σ_o} for each σ_o^2, is applicable for the larger class of hypotheses with an arbitrary $\sigma^2 > 0$.

Exactly the same situation holds concerning rank tests compared with the best test, i.e. the best test depends on the specific simple alternative, whereas the rank test, which is asymptotically equivalent to the best test, does not depend on the specific alternative within a certain subclass of alternatives. For proving this, two things will be needed:

1) Appropriate tools for proving asymptotic statements of the indicated kind. These are developed by LeCam [4] and Hájek [1] with the help of their theory of contiguity.

2) Definition of the corresponding hypotheses of the asymptotic testing problem and of asymptotically uniformly most powerful tests. These concepts were introduced by Neyman [6].

For introducing contiguity, let us remind that the basic problem of testing theory is that of testing two simple hypotheses. Therefore, let us assume that for each $n \in \mathbb{N}$ we have two simple hypotheses

$$H_n : u_n \quad , \quad K_n : v_n$$

where u_n and v_n are probability measures on the sample space $(\mathcal{X}_{(n)}, \mathcal{B}_{(n)})$, for instance the distributions of $n=k+m$ independent random variables $X_1,\ldots,X_k$, $Y_1,\ldots,Y_m$ with distributions, characterized by parameters $\theta_n' \in H$ and $\theta_n \in K$, respectively, i.e.

$$u_n = \mathcal{L}_{\theta_n'}(X_1,\ldots,X_k,Y_1,\ldots,Y_m),\ \theta_n' \in H;\ v_n = \mathcal{L}_{\theta_n}(X_1,\ldots,X_k,Y_1,\ldots,Y_m),\ \theta_n \in K.$$

Then the sequence $\{v_n\}$ is said to be contiguous to the sequence $\{u_n\}$, if

$$(6) \qquad u_n(B_n) \to 0 \quad \text{implies} \quad v_n(B_n) \to 0$$

for any sequence $\{B_n\}$ with $B_n \in \mathcal{B}_{(n)}$ $\forall n \in \mathbb{N}$. It is a trivial but important consequence of this property, that any sequence of r.v.'s Z_n, which converges to 0 in u_n-probability , converges to 0 in v_n-probability,

too, i.e.

$$Z_n \xrightarrow[u_n]{} 0 \text{ implies } Z_n \xrightarrow[v_n]{} 0 \quad .$$

Here $Z_n \xrightarrow[u_n]{} 0$ is defined by $u_n(|Z_n| > \varepsilon) \to 0 \quad \forall \varepsilon > 0$ for $n \to \infty$. Conversely, this last property implies (6).

In most of the important applications of contiguity, a property is fulfilled, which is somewhat stronger than contiguity. In order to explain this, let us introduce the test statistic of the Neyman-Pearson test for testing $H_n : u_n$ against $K_n : v_n$, i.e.

$$(7) \qquad L_n := \begin{cases} q_n/p_n & , \ p_n > 0 \\ \\ \infty & , \ p_n = 0 \end{cases} \quad .$$

Here p_n and q_n are densities of u_n and v_n with respect to a dominating measure. In case that $X_1,\ldots,X_k,Y_1,\ldots,Y_m$ are independent r.v.'s, $L_n(X_{(k)},Y_{(m)})$ is a product, i.e.
$\log L_n := \log L_n(X_{(k)},Y_{(m)})$ is a sum of independent random variables. This implies, according to the central limit theorem, that very often the limiting distribution of $\log L_n$ is normal with mean $-\mathfrak{X}^2/2$ and variance $\mathfrak{X}^2$. In other words, very often there exists a $\mathfrak{X}^2 \geq 0$, such that

$$(8) \qquad \mathcal{L}_{u_n}(\log L_n) \to \mathfrak{N}(-\mathfrak{X}^2/2, \mathfrak{X}^2) \quad .$$

As is proved by LeCam, this property (8) entails

1) contiguity and therefore:

$$(9) \qquad Z_n \xrightarrow[u_n]{} 0 \quad \text{implies} \quad Z_n \xrightarrow[v_n]{} 0 \quad .$$

2) $\log L_n$ is under v_n asymptotically normal with the same variance $\mathfrak{X}^2$ as under u_n, but with mean $+\mathfrak{X}^2/2$, i.e.

$$(10) \qquad \mathcal{L}_{v_n}(\log L_n) \to \mathfrak{N}(+\mathfrak{X}^2/2, \mathfrak{X}^2) \quad .$$

To see the importance of property (8), let us assume $\mathfrak{K} > 0$ and that there exists a statistic S_n, such that

$$(11) \qquad S_n - \frac{\log L_n + \mathfrak{K}^2/2}{\mathfrak{K}} \xrightarrow[u_n]{} 0 \quad .$$

According to contiguity, (9) implies

$$(12) \qquad S_n - \frac{\log L_n + \mathfrak{K}^2/2}{\mathfrak{K}} \xrightarrow[v_n]{} 0 \quad .$$

Since L_n and therefore $(\log L_n + \mathfrak{K}^2/2)/\mathfrak{K}$ is the test statistic of the Neyman-Pearson test for $H_n : u_n$ against $K_n : v_n$, this means, that S_n is asymptotically as good for testing H_n against K_n as the best test statistic is. On the other hand, S_n does not as much depend on the two simple hypotheses $H_n : u_n$ and $K_n : v_n$ as L_n actually does. Therefore it is often the case, for instance if S_n is the standardized test statistic $\tilde{\pi}$ of the Wilcoxon or Fisher-Yates test, that S_n is independent of u_n and v_n within an appropriate class of distributions, i.e. S_n is asymptotically optimal for an appropriate class of such sequences.

According to the basic properties of convergence in distribution, (8), (10), (11) and (12) imply

$$(13) \qquad \mathfrak{L}_{u_n}(S_n) \to \mathfrak{N}(0,1), \qquad \mathfrak{L}_{v_n}(S_n) \to \mathfrak{N}(\mathfrak{K},1) \ ,$$

i.e. S_n is not only an asymptotically optimal test statistic but also an asymptotically normally distributed r.v.. Therefore critical value and power along the sequences v_n can asymptotically easily be determined.

For making precise the kind of asymptotic optimality of rank tests we have to formulate the hypotheses of the asymptotic testing problem. To this end let me remind you of the above indicated finite optimality of rank tests in the two sample problem. There we started (in the second

step) with rank tests and determined (by Hoeffding's formula) that one which was "locally most powerful" against the class of translation alternatives (F_Δ, F) with $F_\Delta(z)=F(z-\Delta)$ and a specific distribution function F or against the class of nonparametric alternatives (F_Δ, F) with $F_\Delta(z)=(1-\Delta)F(z)+\Delta F^2(z)$ and an arbitrary continuous distribution function F. For proving asymptotic optimality we now start with the class of all tests and determine for each $n \in \mathbb{N}$ the best test for the two simple hypotheses $\theta_o=(F,F)$ and, for instance, a translation alternative, say $\theta=(F_{\Delta_{1n}}, F_{\Delta_{2n}})$ with

$$F_{\Delta_{in}}(z)=F(z-\Delta_{in}),\ i=1,2;\ \Delta_{1n}-\Delta_{2n}=\eta\sqrt{\frac{k+m}{km}}+o\left(\sqrt{\frac{k+m}{km}}\right) .$$

If u_n and v_n denote the common distribution of $X_1,\ldots,X_k,Y_1,\ldots,Y_m$ under (F,F) and $(F_{\Delta_{1n}}, F_{\Delta_{2n}})$, respectively, and if L_n is the above defined Neyman-Pearson test statistic, then under certain regularity conditions (like differentiability of the corresponding densities)

$$\mathcal{L}_{u_n}(\log L_n) \to \mathcal{N}(-\sigma^2/2, \sigma^2) \text{ with } \sigma^2=\eta^2 J(F)$$

holds, where $J(F)$ is Fisher's information with respect to the underlying one-parametric class of alternatives. Moreover, you can prove that under these regularity conditions there always exists a rank statistic S_n with

$$S_n-\frac{\log L_n+\sigma^2/2}{\sigma} \xrightarrow[u_n]{} 0 ,$$

for instance the standardization of $\tilde{\pi}$. But S_n is a rank statistic, i.e. its distribution is independent of $\theta=(G,G) \in J$. In other words: though we started with the class of all tests, we can prove asymptotic optimality of a rank test.

To make these things precise, we have to define the concept of an

asymptotically uniformly most powerful test. Given asymptotic composite hypotheses, i.e. sets $\hat{H}$ and $\hat{K}$ of sequences $\{u_n\}$ and $\{v_n\}$, respectively, with u_n and v_n as defined above.

A sequence $\{\phi_n\}$ of tests ϕ_n is called an asymptotically uniformly most powerful test of level α for $\hat{H}$ against $\hat{K}$, if

$$\limsup_{n \to \infty} E_{u_n} \phi_n \leq \alpha \qquad \forall \{u_n\} \in \hat{H} \quad ,$$

and if for all sequences $\{\tilde{\phi}_n\}$ with this property, it holds that

$$\liminf_{n \to \infty} (E_{v_n} \phi_n - E_{v_n} \tilde{\phi}_n) \geq 0 \qquad \forall \{v_n\} \in \hat{K}.$$

Then the main result is, as indicated above, that a rank test ϕ_n is asymptotically uniformly most powerful against the sequences $\{v_n\}$ constructed above, raising from that class of alternatives, against which the rank test ϕ_n is "locally most powerful" for a fixed finite sample size n.

Unfortunately, it is impossible for me to go into details and to indicate how we could simplify the proofs, which were originally given by Hájek [1]. But I should mention that Mr. Behnen could extend these results to nonparametric alternatives and rank correlation tests and that, based on some work of Mr. Behnen and Mr.Nölle it is possible to give a unified theory of asymptotic optimality including the two sample rank tests, the tests of symmetry and the rank correlation tests, which you will find in the above mentioned book of Mr. Nölle and mine [7].

At the end, I would like to mention that there is an intimate connection of this result with the concept of Pitman efficiency. In particular, the indicated methods of contiguity enabled Hájek [1] to give a simple and very elegant proof of the interesting fact, that the Pitman efficiency of a rank test ϕ_n, which is asymptotically uniformly most powerful against alternatives $(F_{\Delta_{1n}}, F_{\Delta_{2n}})$, compared with the best (parametric)

test $\phi^{*}_{n\Delta}$ of (F,F) against $(F_{\Delta_{1n}},F_{\Delta_{2n}})$ equals one, if this comparison is done under $(F_{\Delta_{1n}},F_{\Delta_{2n}})$. For the special case of the Fisher-Yates test and normal translation alternatives this was proved earlier by Chernoff and Savage.

Moreover Hájek [1] could easily derive the Pitman efficiency $e_p(\tilde{\phi}_n : \phi_n | v_n)$ of a rank test $\tilde{\phi}_n$ with respect to the rank test ϕ_n for the underlying sequence $\{v_n\}$, where v_n is the distribution of $X_1,\ldots,X_k,Y_1,\ldots,Y_m$ under $(F_{\Delta_{1n}},F_{\Delta_{2n}})$. Let ϕ_n and $\tilde{\phi}_n$ be derived as asymptotically optimal against the alternatives $(F_{\Delta_{1n}},F_{\Delta_{2n}})$ and $(\tilde{F}_{\Delta_{1n}},\tilde{F}_{\Delta_{2n}})$ respectively. Then it holds

$$e_p(\tilde{\phi}_n : \phi_n | v_n) = \rho^2 \tag{14}$$

where ρ is the correlation-coefficient between the test statistics S_n and $\tilde{S}_n$ of ϕ_n and $\tilde{\phi}_n$, respectively, under (F,F). Since S_n and $\tilde{S}_n$ are rank statistics, ρ is the correlation-coefficient under $(\tilde{F},\tilde{F})$, too. Therefore, if $\tilde{v}_n$ is the distribution of $X_1,\ldots,X_k,Y_1,\ldots,Y_m$ under $(\tilde{F}_{\Delta_{1n}},\tilde{F}_{\Delta_{2n}})$, the Pitman-efficiency of ϕ_n with respect to $\tilde{\phi}_n$ under $\tilde{v}_n$ is ρ^2, too, i.e.

$$e_p(\phi_n : \tilde{\phi}_n | \tilde{v}_n) = \rho^2 \quad . \tag{15}$$

For instance, the Pitman-efficiency of the Wilcoxon test with respect to the Fisher-Yates test for underlying normal translation alternatives is $3/\pi = 0{,}955$ and the same holds for the Pitman efficiency of the Fisher-Yates test with respect to the Wilcoxon test for underlying logistic translation alternatives.

[1] Hájek, J.: Asymptotically most powerful rank-order-tests. Ann.Math.Stat.33 (1962)

[2] Hájek, J.: Theory of Rank Tests. Prag, New York (1967)

[3] Hoeffding, W.: "Optimum" nonparametric tests. Proc.2nd Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, Univ. of Calif. Press (1951)

[4] LeCam, L.: Locally asymptotically normal families of distributions. Univ. of Calif. Publ. in Stat. 3 (1960)

[5] Lehmann, E.L.: Testing Statistical Hypotheses. Wiley, New York (1959)

[6] Neyman, J.: Optimal asymptotic tests of composite statistical hypotheses. The Harald Cramér Volume, Stockholm, New York (1959)

[7] Witting, H. und G. Nölle: Angewandte Mathematische Statistik. Teubner, Stuttgart (1969)

ON THE IMPORTANCE OF COMPONENTS IN A SYSTEM *)

by Z.W. Birnbaum (USA)

University of Washington

1. INTRODUCTION: DEFINITIONS AND NOTATIONS

1.1 In a system whose performance depends on the performance of its components, some of these components may play a more important part than others. For example, in the system indicated in Figure 1, component c_1 would seem intuitively more important than $c_2, c_3, c_4, \ldots c_n$.

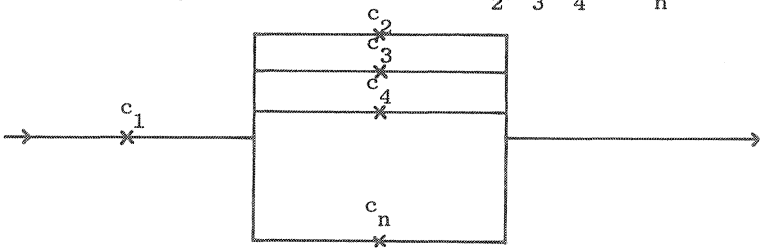


figure 1

In the following we propose a quantitative definition of this notion of importance, and discuss some properties and applications of this concept.

1.2 We assume that every device, whether it is a single component or a system consisting of components, can be in one and only one of two states: it functions or it fails.

When a system consists of components $c_1, c_2, \ldots c_n$, we ascribe to each of them a binary indicator variable

$$x_i = \begin{cases} 1 & \text{when } c_i \text{ functions} \\ 0 & \text{when } c_i \text{ fails.} \end{cases}$$

Each n-tuple of 0's or 1's

$$(x_1, x_2, \ldots x_n) = \underline{x}$$

---

*) This research was supported by the Office of Naval Research.

is called a *vector of component states* or in short a *state vector*. It can assume any one of the 2^n values represented by the vertices of the unit cube in n-dimensional space:
$(0,0,\ldots,0)$, $(1,0,\ldots,0)$, $(1,1,0,\ldots,0)$, ... $(1,1,\ldots,1)$.
We shall use the following notations:

$$\underline{x} < \underline{y} \quad \text{when} \quad x_i \leq y_i \qquad \text{for} \quad i=1,\ldots,n$$

$$\underline{x} = \underline{y} \quad \text{when} \quad x_i = y_i \qquad \text{for} \quad i=1,\ldots,n$$

$$\underline{x} \leq \underline{y} \quad \text{when} \quad \underline{x} < \underline{y} \qquad \text{and} \quad \underline{x} \neq \underline{y}$$

$$(0_k,\underline{x}) = (x_1,x_2,\ldots,x_{k-1},0,x_{k+1},\ldots,x_n)$$

$$(1_k,\underline{x}) = (x_1,x_2,\ldots,x_{k-1},1,x_{k+1},\ldots,x_n)$$

$$\underline{0} = (0,0,\ldots,0), \quad \underline{1} = (1,1,\ldots,1).$$

We ascribe to the system a binary indicator variable

$$u = \begin{cases} 1 & \text{when the system functions} \\ 0 & \text{when the system fails.} \end{cases}$$

When the design of a system is known, then the state vector $\underline{x}$ determines the state of the system so that

$$u = \phi(\underline{x})$$

where ϕ is a function with values 0 or 1, called the *structure function* of the system.

A structure function ϕ is called *coherent* [1] when it fulfills the conditions: $\phi(\underline{0}) = 0$, $\phi(\underline{x}) \leq \phi(\underline{y})$ for $\underline{x} \leq \underline{y}$, and $\phi(\underline{1}) = 1$. We shall consider only coherent structure functions. One verifies immediately that

$$(1.2.1) \quad \phi(\underline{x}) = x_j\left[\phi(1_j,\underline{x}) - \phi(0_j,\underline{x})\right] + \phi(0_j,\underline{x}) = x_j\delta_j(\underline{x}) + \mu_j(\underline{x}), \quad j = 1,2,\ldots,n$$

where

$$(1.2.2) \qquad \delta_j(\underline{x}) = \phi(1_j,\underline{x}) - \phi(0_j,\underline{x}) = \frac{\partial\phi(\underline{x})}{\partial x_j}$$

(1.2.3) $$\mu_j(\underline{x}) = \phi(0_j,\underline{x}),$$

and $\delta_j(\underline{x})$ as well as $\mu_j(\underline{x})$ do not depend on the state x_j of component c_j.

1.3 If the state of c_j is determined by chance, so that the value actually assumed by x_j is a binary random variable X_j with the probability distribution

$$P\{X_j = 1\} = p_j$$

(1.3.1) $$j=1,2,\ldots,n$$

$$P\{X_j = 0\} = q_j=1-p_j$$

then p_j is called the *reliability* of c_j. In the following we shall assume that $X_1,X_2,\ldots,X_n$ are totally independent. The n-tuple of component reliabilities determines a point

(1.3.2) $$(p_1,p_2,\ldots,p_n) = \underline{p}$$

in the n-dimensional unit cube $(p_1,\ldots,p_n) : 0 \leq p_i \leq 1 = J_n$.

For a given structure function $\phi(\underline{x})$, the values of component reliabilities $(p_1,\ldots,p_n) = \underline{p}$ determine the probability that the system will function

(1.3.3) $$P\{\phi(\underline{x}) = 1 \mid \underline{p}\} = E[\phi(\underline{x}) \mid \underline{p}] = h_\phi(\underline{p}).$$

This function $h_\phi(\underline{p})$, defined on J_n, is the *reliability function* for ϕ.

There are situations when only the design of a system is known, i.e. $\phi(\underline{x})$ is given, but no information is available about the component reliabilities. We shall consider the relative importance of various components in such situations, and shall call it *structural importance*.

In other instances, both the structure function ϕ and the component reliabilities $\underline{p}$ are known. The concept of importance which will be introduced for such situations will be referred to as *reliability importance*.

A third, more complicated way of considering the importance of components will be briefly mentioned in Section 6.

2. STRUCTURAL IMPORTANCE

2.1 A component c_j is *relevant* for structure ϕ *at* the *state vector* (vertex of unit cube) $\underline{x}$ when

$$\delta_j(\underline{x}) = \phi(1_j,\underline{x}) - \phi(0_j,\underline{x}) = 1; \tag{2.1.1}$$

c_j is *relevant at* $\underline{x}$ *for the functioning of* ϕ when

$$(1-x_j)\ \delta_j(\underline{x}) = 1 \tag{2.1.2}$$

and c_j is *relevant at* $\underline{x}$ *for the failure* of ϕ when

$$x_j\ \delta_j(\underline{x}) = 1. \tag{2.1.3}$$

Clearly, if c_j is relevant at $\underline{x}$ then it is either relevant for functioning or for failure, depending on whether the vertex $\underline{x}$ has its coordinate x_j equal to 0 or to 1.

We define the *structural importance* of c_j *for the functioning* of ϕ as

$$I_j(\phi,1) = 2^{-n} \sum_{(\underline{x})} (1-x_j)\delta_j(\underline{x}) \tag{2.1.4}$$

where the sum is extended over all 2^n vertices of the unit cube (state vectors), and similarly the *structural importance* of c_j *for failure of* ϕ as

$$I_j(\phi,0) = 2^{-n} \sum_{(\underline{x})} x_j\delta_j(\underline{x}) . \tag{2.1.5}$$

Finally, the *structural importance* of c_j for ϕ is defined as

(2.1.6) $I_j(\phi) = I_j(\phi,0) + I_j(\phi,1) = 2^{-n} \sum_{(\underline{x})} \delta_j(\underline{x})$.

One verifies that if c_j is relevant at $\underline{x}$ for the functioning of ϕ then c_j is relevant at $(1_j,\underline{x})$ for failure and if c_j is relevant at $\underline{x}$ for failure of ϕ then it is relevant at $(0_j,\underline{x})$ for functioning. Therefore, from (2.1.4),(2.1.5),(2.1.6) follows

(2.1.7) $I_j(\phi,1) = I_j(\phi,0) = \frac{1}{2} I_j(\phi)$,

and there is no purpose in distinguishing between structural importance for functioning and for failure. We shall see, however, that for reliability importance a similar distinction is meaningful.

2.2 Examples

2.2.1 k-out-of-n structures.

A structure $\phi(\underline{x})$ with n components is called "k-out-of-n" when it functions whenever at least k of its components function. One verifies that for such ϕ

(2.2.1.1) $I_j(\phi) = 2^{-n} \cdot 2\binom{n-1}{k-1}$, $j = 1,2,\ldots,n$.

All components have the same structural importance, and this importance is greatest for $k = \frac{n}{2}$ if n is even, and for $k = \left[\frac{n}{2}\right]$ and $k = \left[\frac{n}{2}\right] + 1$ if n is odd. The importance of every component is smallest in the case of n components in series (n-out-of-n structure) and of n components in parallel (1-out-of-n structure) when $I_j(\phi) = 2^{-n} \cdot 2$.

2.2.2 k components in series, in series with n-k in parallel.

Let

(2.2.2.1) $\phi(\underline{x}) = x_1x_2\ldots x_k \cdot \left[1-(1-x_{k+1})\ldots(1-x_n)\right]$.

This structure may be represented by the diagram in Figure 2.

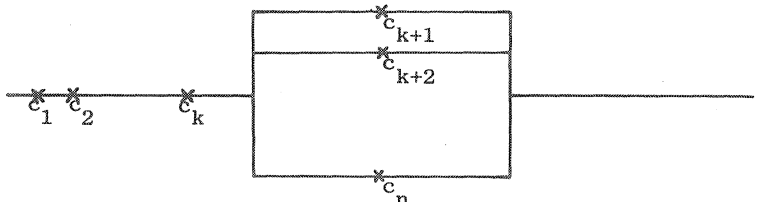


figure 2

One computes

$$(2.2.2.2)\quad \delta_j = \prod_{\substack{r=1\\ r\neq j}}^{k} x_r \left[1 - \prod_{t=k+1}^{n} (1-x_t)\right] \qquad \text{for } j=1,\ldots,k$$

$$(2.2.2.3)\quad \delta_j = \prod_{r=1}^{k} x_r \prod_{\substack{t=k+1\\ t\neq j}}^{n} (1-x_t) \qquad \text{for } j=k+1,\ldots,n$$

hence

$$(2.2.2.4)\quad I_j(\phi) = 2.2^{-n}(2^{n-k}-1) = 2(2^{-k}-2^{-n}) \qquad \text{for } j=1,\ldots,k$$

$$(2.2.2.5)\quad I_j(\phi) = 2.2^{-n} \qquad \text{for } j=k+1,\ldots,n\,.$$

We see that $c_1,\ldots,c_k$ have each importance $2(2^{-k}-2^{-n})$, much greater than the importance 2.2^{-n} of each of $c_{k+1},\ldots,c_n$, which agrees with what one would intuitively expect.

3. RELIABILITY IMPORTANCE

3.1 From (1.2.1) and (1.3.4) one obtains immediately for the reliability function the expression

$$(3.1.1)\quad h_\phi \underline{p} = p_j E\left[\delta_j(\underline{X})\right] + E\left[\mu_j(\underline{X})\right]$$

for every $j=1,2,\ldots,n$, and from (3.1.1) and (1.2.1) follows

$$(3.1.2)\quad \frac{\partial h_\phi(\underline{p})}{\partial p_j} = E\left[\delta_j(\underline{X})\right] = E\left[\frac{\partial \phi(\underline{x})}{\partial x_j}\right], \quad j=1,2,\ldots,n.$$

One also proves by straightforward algebra [2] the identity

$$(3.1.3)\quad \operatorname{cov}\left[X_j, \phi(\underline{X})\right] = p_j(1-p_j)\, E\left[\delta_j(\underline{X})\right], \quad j=1,2,\ldots,n.$$

3.2 We define the *reliability importance* of c_j *for the functioning* of ϕ as

$$(3.2.1)\quad R_j(\phi,1;\underline{p}) = P\left\{\phi(\underline{X}) = 1 \mid X_j = 1\ ;\ \underline{p}\right\} - P\left\{\phi(\underline{X}) = 1\ ;\ \underline{p}\right\}$$

and, similarly, the *reliability importance* of c_j *for failure of* ϕ as

$$(3.2.2)\quad R_j(\phi,0;\underline{p}) = P\left\{\phi(\underline{X}) = 0 \mid X_j=0;\ \underline{p}\right\} - P\left\{\phi(\underline{X}) = 0;\ \underline{p}\right\}$$

and the *reliability importance* of c_j for ϕ as

$$(3.2.3)\quad R_j(\phi;\underline{p}) = R_j(\phi,1;\underline{p}) + R_j(\phi,0;\underline{p}).$$

The following identities will be frequently used:

$$(3.2.4)\quad R_j(\phi,1;\underline{p}) = (1-p_j)\frac{\partial h(\underline{p})}{\partial p_j} = E\left[(1-X_j)\ \delta_j(\underline{X})\right]$$

$$(3.2.5)\quad R_j(\phi,0;\underline{p}) = p_j \frac{\partial h(\underline{p})}{\partial p_j} = E\left[X_j\ \delta_j(\underline{X})\right]$$

$$(3.2.6)\quad R_j(\phi;\underline{p}) = \frac{\partial h(\underline{p})}{\partial p_j} = E\left[\delta_j(\underline{X})\right].$$

Proof: using (3.1.3)

$$P\left\{\phi(\underline{X}) = 1 \mid X_j = 1\right\} = \frac{P\left\{\phi(\underline{X})=X_j=1\right\}}{P\left\{X_j=1\right\}} = \frac{E\left[X_j\phi(\underline{X})\right]}{p_j} =$$

$$= \frac{\operatorname{cov}\left[X_j, \phi(\underline{X})\right] + E(X_j)E\left[\phi(\underline{X})\right]}{p_j} = (1-p_j)E\left[\delta_j(\underline{X})\right]$$

$$+ E\left[\phi(\underline{X})\right]$$

and from (3.2.1) and (1.3.3) one obtains (3.2.4). A similar argument yields (3.2.5), and (3.2.6) follows by adding (3.2.4) and (3.2.5).

3.3 If nothing is known about the reliabilities of the components and, for lack of better knowledge, it is assumed that all vertices $\underline{x}$ are equally probable i.e. each has probability 2^{-n}, then (3.2.4), (3.2.5), and (3.2.6) reduce to (2.1.4), (2.1.5) and (2.1.6), the corresponding structural importances.

3.4. Examples

3.4.1 k-out-of-n structures.

For a k-out-of-n structure we have $\delta_j(\underline{x}) = \phi(1_j,\underline{x}) - \phi(0_j,\underline{x}) = 1$ if and only if exactly k-1 of the n-1 components different from c_j function. Therefore,

$$(3.4.1\ 1)\quad R_j(\phi;\underline{p}) = \sum p_{j_1} p_{j_2} \ldots p_{j_{k-1}} (1-p_{j_k})(1-p_{j_{k+1}}) \ldots (1-p_{j_{n-1}})$$

where the sum is extended over all permutations $(j_1,j_2,\ldots,j_{n-1})$ of the subscripts $(1,2,\ldots,j-1,j+1,\ldots,n)$.

3.4.2 k components in series, in series with n-k in parallel.

For the structure function (2.2.2.1) one computes

$$R_j(\phi,\underline{p}) = \frac{1}{p_j} \prod_{r=1}^{k} p_r \left[1 - \prod_{t=k+1}^{n} (1-p_t)\right], \quad \text{if } j=1,2,\ldots,k,$$

(3.4.2.1)

$$R_j(\phi,\underline{p}) = \prod_{r=1}^{k} p_r \cdot \frac{1}{1-p_j} \prod_{t=k+1}^{n} (1-p_t), \quad \text{if } j=k+1,\ldots,n,$$

and corresponding expressions are immediately obtained for $R(\phi,1,\underline{p})$ and $R(\phi,0,\underline{p})$.

3.4.3 In the examples above, only the reliability importance defined by (3.2.3) has been computed. Expressions (3.2.1) and (3.2.2) can be useful when a cost function is given which ascribes different cost to the failure and to the functioning of a structure.

4. STRUCTURES WITH MODULES

4.1. In designing multi-component systems one often proceeds step-by-step, first constructing a system of fewer components and then replacing some of these components by sub-systems, known as *modules*, each consisting of several components. Some properties of coherent systems constructed of coherent modules have been studied in [3]. For our present purpose we shall use the following definitions:

Let

$$(4.1.1)\quad \phi(\underline{x}) = \phi(x_1,x_2,\ldots,x_n) = x_1\,\delta_{x_1}(\phi;\underline{x}) + \mu_{x_1}(\phi;\underline{x})$$

and

$$(4.1.2)\quad \psi(\underline{y}) = \psi(y_1,y_2,\ldots,y_m)$$

be two coherent structures. We shall say that the structure

$$(4.1.3)\quad \chi(y_1,y_2,\ldots,y_m,x_2,\ldots,x_n) = \phi\left[\psi(y_1,\ldots,y_m),x_2,\ldots,x_n\right] =$$

$$= \phi\left[\psi_1(\underline{y}),\underline{x}\right] = \psi(\underline{y})\delta_{x_1}(\phi;\underline{x}) + \mu_{x_1}(\phi;\underline{x})$$

was obtained by replacing component x_1 in $\phi(\underline{x})$ by the module $\psi(\underline{y})$.

4.2 From (4.1.3) one obtains

$$\delta_{y_1}(\chi;y_1,\ldots,y_m,x_2,\ldots,x_n) = \chi(1,y_2,\ldots,y_m,x_2,\ldots,x_n) -$$

$$- \chi(0,y_2,\ldots,y_m,x_2,\ldots,x_n) = \left[\psi(1\text{-}\underline{y}) - \psi(0_1,\underline{y})\right]\delta_{x_1}(\phi;\underline{x}) =$$

$$= \delta_{y_1}(\psi;\underline{y}) \cdot \delta_{x_1}(\phi;\underline{x}).$$

From the so obtained identity

(4.2.1) $$\delta_{y_1}(\chi) = \delta_{x_1}(\phi;\underline{x})\ \delta_{y_1}(\psi;\underline{y})$$

and from (3.2.6) follows

(4.2.2) $$R_{y_1}(\chi;y_1,\ldots,y_m,x_2,\ldots,x_n) = R_{x_1}(\phi;\underline{x}).R_{y_1}(\psi;\underline{y}).$$

This "chain-rule" property (which could also have been obtained by the chain rule for differentiation using $E\left[\delta_j(\underline{x})\right] = \frac{\partial h(\underline{p})}{\partial p_j}$) makes it possible to compute the importance of each component of a module ψ for the entire system χ, and to repeat this step-by-step as modules are substituted for components. The computation of $R_{y_1}(\chi,1;y_1,\ldots,y_m,x_2,\ldots,x_n)$ and of $R_{y_1}(\chi,0;y_1,\ldots,y_m,x_2,\ldots,x_n)$ is then a simple matter, according to (3.2.4) and (3.2.5).

5. AN APPLICATION

If components with known reliabilities $(p_1,\ldots,p_n) = \underline{p}$ are available, and the known structure $\phi(\underline{x})$ has the reliability $h(\underline{p}) = E\left[\phi(\underline{X});p\right]$, then the problem may arise to decide on which components additional research and development should be done to improve their reliabilities, so that the greatest gain is achieved in system reliability.

Let us assume that improving the reliability of c_j from p_j to $p_j + \Delta_j$ can be achieved at cost $\lambda_j(p_j).\Delta_j$, for $j=1,\ldots,n$. In practical situations $\lambda_j(p_j)$ will be an increasing function, such that $\lambda_j(0) = 0$, $\lambda_j(p) \xrightarrow[p \to 1]{} \infty$. The total cost of improving all components will be

$$(5.1) \qquad C(\underline{p}, \underline{\Delta}) = \sum_{j=1}^{n} \lambda_j(p_j)\, \Delta_j,$$

and the gain in system reliability per unit of cost

$$(5.2) \qquad \frac{h_\phi(\underline{p} + \underline{\Delta}) - h_\phi(\underline{p})}{C(\underline{p}, \underline{\Delta})} .$$

We shall look for the direction of steepest ascent of this gain, in the following sense:

Let

$$(5.3) \qquad \Delta_j = \alpha_j t, \quad j=1,2,\ldots,n$$

with

$$(5.4) \qquad \sum_{j=1}^{n} \alpha_j^2 = 1.$$

We wish to determine the vector of direction cosines $(\alpha_1,\ldots,\alpha_j,\ldots,\alpha_n) = \underline{\alpha}$ so that, for all Δ_j small, (5.2) is maximized. Since (5.2) now is

$$\frac{h_\phi(\underline{p}+\underline{\alpha}t) - h_\phi(\underline{p})}{t \sum_{j=1}^{n} \lambda_j(p_j)\alpha_j} \xrightarrow[t \to 0]{} \frac{\frac{d}{dt} h(\underline{p}+\underline{\alpha}t)\,|_{t=0}}{\sum_{j=1}^{n} \lambda_j(p_j)\alpha_j}$$

our problem is to maximize

$$\frac{\sum_{j=1}^{n} \frac{\partial h(\underline{p})}{\partial p_j}\, \alpha_j}{\sum_{j=1}^{n} \lambda_j(p_j)\alpha_j} = \frac{\sum_{j=1}^{n} R_j(\phi,\underline{p})\alpha_j}{\sum_{j=1}^{n} \lambda_j(p_j)\alpha_j} ,$$

under the restriction (5.4). It can be shown that, except for degenerate cases, the maximum is attained by selecting that component c_{j_0} for which the importance-to-cost ratio $R_j(\phi,\underline{p})/\lambda_j(p_j)$ is maximum, and setting $\alpha_{j_0} = 1$, $\alpha_j = 0$ for $j \neq j_0$.

6. CONCLUDING REMARKS

We have considered situations where only the structure function of $\phi(\underline{x})$ of the system was known, and situations where also the reliabilities $\underline{p} = (p_1, p_2, \ldots, p_n)$ of the components were known, and for each of these situations we proposed a quantitative definition of importance of components. A third possibility should be considered, when the coherent structure ϕ is known and each component c_i has a life length T_i, with a known probability distribution $F_i(t) = P\{T_i \leq t\}$. Under these assumptions the system has a life length [4] T, with a probability distribution $P\{T \leq t\} = F(t)$ which depends on ϕ and on all the $F_i(T)$, $i=1,2,\ldots,n$. Again, intuitively some of the components are more important than others for the life distribution $F(t)$, and their importance depends on their location within the structure as well as on all the life distributions. To our knowledge, a study of the problem arising in this context has not even been initiated.

The author wishes to express his appreciation to Dr. J.D. Esary for many helpful discussions.

REFERENCES

[1] Z.W. Birnbaum, J.D. Esary, and S.C. Saunders, Multicomponent systems and structures and their reliabilities, Technometrics, Vol. 3, 1961, pp. 55-77.

[2] J.D. Esary and F. Proschan, Coherent structures of non-identical components, Technometrics, Vol. 5, 1963, pp. 191-209.

[3] Z.W. Birnbaum and J.D. Esary, Modules of coherent binary systems, J. Soc. Indust. Appl. Math., Vol. 13, 1965, pp. 444-462.

[4] J.D. Esary and A.W. Marshall, System structure and the existence of a system life", Technometrics, Vol. 6, pp. 459-462.

OTHER TITLES IN THE SERIES MATHEMATICAL CENTRE TRACTS

A leaflet containing an order-form and abstracts of all publications mentioned below is available at the Mathematical Centre, 2e Boerhaavestraat 49, Amsterdam-1005, The Netherlands. Orders should be sent to the same address.

MCT 1 T. VAN DER WALT, *Fixed and almost fixed points*, 1963.

MCT 2 A.R. BLOEMENA, *Sampling from a graph*, 1964.

MCT 3 G. DE LEVE, *Generalized Markovian decision processes, part I: Model and method*, 1964.

MCT 4 G. DE LEVE, *Generalized Markovian decision processes, part II: Probabilistic background*, 1964.

MCT 5 G. DE LEVE, H.C. TIJMS & P.J. WEEDA, *Generalized Markovian decision processes, Applications*, 1970.

MCT 6 M.A. MAURICE, *Compact ordered spaces*, 1964.

MCT 7 W.R. VAN ZWET, *Convex transformations of random variables*, 1964.

MCT 8 J.A. ZONNEVELD, *Automatic numerical integration*, 1964.

MCT 9 P.C. BAAYEN, *Universal morphisms*, 1964.

MCT 10 E.M. DE JAGER, *Applications of distributions in mathematical physics*, 1964.

MCT 11 A.B. PAALMAN-DE MIRANDA, *Topological semigroups*, 1964.

MCT 12 J.A.TH.M. VAN BERCKEL, H. BRANDT CORSTIUS, R.J. MOKKEN & A. VAN WIJNGAARDEN, *Formal properties of newspaper Dutch*, 1965.

MCT 13 H.A. LAUWERIER, *Asymptotic expansions*, 1966, out of print; replaced by MCT 54.

MCT 14 H.A. LAUWERIER, *Calculus of variations in mathematical physics*, 1966.

MCT 15 R. DOORNBOS, *Slippage tests*, 1966.

MCT 16 J.W. DE BAKKER, *Formal definition of programming languages with an application to the definition of ALGOL 60*, 1967.

MCT 17 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 1*, 1968.

MCT 18 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 2*, 1968.

MCT 19 J. VAN DER SLOT, *Some properties related to compactness*, 1968.

MCT 20 P.J. VAN DER HOUWEN, *Finite difference methods for solving partial differential equations*, 1968.

MCT 21 E. WATTEL, *The compactness operator in set theory and topology*, 1968.

MCT 22 T.J. DEKKER, *ALGOL 60 procedures in numerical algebra, part 1*, 1968.

MCT 23 T.J. DEKKER & W. HOFFMANN, *ALGOL 60 procedures in numerical algebra, part 2*, 1968.

MCT 24 J.W. DE BAKKER, *Recursive procedures*, 1971.

MCT 25 E.R. PAERL, *Representations of the Lorentz group and projective geometry*, 1969.

MCT 26 EUROPEAN MEETING 1968, *Selected statistical papers, part I*, 1968.

MCT 27 EUROPEAN MEETING 1968, *Selected statistical papers, part II*, 1969.

MCT 28 J. OOSTERHOFF, *Combination of one-sided statistical tests*, 1969.

MCT 29 J. VERHOEFF, *Error detecting decimal codes*, 1969.

MCT 30 H. BRANDT CORSTIUS, *Excercises in computational linguistics*, 1970.

MCT 31 W. MOLENAAR, *Approximations to the Poisson, binomial and hypergeometric distribution functions*, 1970.

MCT 32 L. DE HAAN, *On regular variation and its application to the weak convergence of sample extremes*, 1970.

MCT 33 F.W. STEUTEL, *Preservation of infinite divisibility under mixing and related topics*, 1970.

MCT 34 I. JUHASZ a.o., *Cardinal functions in topology*, 1971.

MCT 35 M.H. VAN EMDEN, *An analysis of complexity*, 1971.

MCT 36 J. GRASMAN, *On the birth of boundary layers*, 1971.

MCT 37 G.A. BLAAUW a.o., *MC-25 Informatica Symposium*, 1971.

MCT 38 W.A. VERLOREN VAN THEMAAT, *Automatic analysis of Dutch compound words*, 1971.

MCT 39 H. BAVINCK, *Jacobi series and approximation*, 1972.

MCT 40 H.C. TIJMS, *Analysis of (s,S) inventory models*, 1972.

MCT 41 A. VERBEEK, *Superextensions of topological spaces*, 1972.

MCT 42 W. VERVAAT, *Success epochs in Bernoulli trials (with applications in number theory)*, 1972.

MCT 43 F.H. RUYMGAART, *Asymptotic theory of rank tests for independence*, 1973.

MCT 44 H. BART, *Meromorphic operator valued functions*, 1973.

MCT 45 A.A. BALKEMA, *Monotone transformations and limit laws*, 1973.

MCT 46 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 1: The language*, 1973.

MCT 47 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 2: The compiler*, 1973.

MCT 48 F.E.J. KRUSEMAN ARETZ, P.J.W. TEN HAGEN & H.L. OUDSHOORN, *An ALGOL 60 compiler in ALGOL 60, Text of the MC-compiler for the EL-X8*, 1973.

MCT 49 H. KOK, *Connected orderable spaces*, 1974.

* MCT 50 A. VAN WIJNGAARDEN, B.J. MAILLOUX, J.E.L. PECK, C.H.A. KOSTER, M. SINTZOFF, C.H. LINDSEY, L.G.L.T. MEERTENS & R.G. FISKER (eds.), *Revised report on the algorithmic language ALGOL 68*.

MCT 51 A. HORDIJK, *Dynamic programming and Markov potential theory*, 1974.

MCT 52 P.C. BAAYEN (ed.), *Topological structures*, 1974.

MCT 53 M.J. FABER, *Metrizability in generalized ordered spaces*, 1974.

MCT 54 H.A. LAUWERIER, *Asymptotic analysis, part 1*, 1974.

MCT 55 M. HALL JR. & J.H. VAN LINT (eds.), *Combinatorics, part 1: Theory of designs, finite geometry and coding theory*, 1974.

MCT 56 M. HALL JR. & J.H. VAN LINT (eds.), *Combinatorics, part 2: Graph theory; foundations, partitions and combinatorial geometry*, 1974.

MCT 57 M. HALL JR. & J.H. VAN LINT (eds.), *Combinatorics, part 3: Combinatorial group theory*, 1974.

MCT 58 W. ALBERS, *Asymptotic expansions and the deficiency concept in statistics*, 1975.

A star (*) before the number means "to appear".