SPACINGS REVISITED

RONALD PYKE
UNIVERSITY OF WASHINGTON
and
IMPERIAL COLLEGE

1. Introduction

This paper surveys some of the developments which have appeared in the literature on spacings during the five years since the presentation of two papers [16] and [17]. The first of these, by Proschan and this author, deals with the asymptotic theory of a class of tests for Increasing Failure Rate (IFR) which are based on spacings, whereas the second paper surveys the substantial literature on spacings that had appeared prior to 1965. In the present article we also set out some open problems which still remain in the asymptotic theory of tests based on spacings.

The general area of limit theorems for dependent random variables is broad and complex, with no unifying methodology. For example, problems related to rank statistics, linear combinations of order statistics and stationary sequences all require different approaches. Limit theorems for spacings represent some of the more challenging problems involving dependent variables, and the various approaches used provide interesting comparisons.

2. Basic formulations

By spacings we refer to the gaps or distances between successive points on a line. Let $\{T_n : n \geqq 0\}$ be a sequence of random variables (r.v.) for which $T_0 \leqq T_1 \leqq T_2 \cdots$. The spacings are then the differences $\{T_i - T_{i-1}\}$. There is a basic ambiguity in the theory of spacings caused by the radically different assumptions which can be placed on the T process. These differences can clearly be seen for example between the three basic models outlined below.

Model I: order statistics. For fixed n, one is given independent random variables $X_1, X_2, \cdots, X_n$, with common distribution function (d.f.) F_X. One defines $T_1 \leqq T_2 \leqq \cdots \leqq T_n$ to be the order statistics of the sample and considers the spacings $D_i = T_i - T_{i-1}$. The range for i is $2 \leqq i \leqq n$ unless the support of F_X indicates that spacings D_1 and/or D_{n+1} may be defined. The usual situation under Model I is a hypothesis testing one in which the two hypotheses are

$$H_0 : F_X \in \mathscr{F}_0, \qquad H_1 : F_X \in \mathscr{F}_1. \tag{2.1}$$

This paper was prepared with the partial support of the National Science Foundation Grant GP-23653.

Usually $\mathscr{F}_0$ consists of a single continuous d.f. which can therefore be assumed to be uniform on (0, 1).

Model II: point processes. In this model, the T process could be a general one sided point process with $T_0 = 0$, say. The most common situation is that of testing the null hypothesis that the T process is Poisson against some alternative involving non-Poissonian point processes. In this special case the spacings are taken to be $D_i = (T_i - T_{i-1})/T_{n+1}$ with $i = 1, 2, \cdots, n + 1$, where $T_0 \leqq T_1 \leqq T_2 \leqq \cdots$, are the successive points.

Model III: renewal processes. This is technically a special case of Model II. Assume that the T process is a renewal process with common d.f. F. Use the proportional spacings or interoccurrence times $D_i = (T_i - T_{i-1})/T_{n+1}$. The most common hypothesis testing problem within this context is to test the null hypothesis that F is exponential against some appropriate alternative for the common d.f. F.

Although the null hypothesis limit theory of statistics based on these spacings is the same (namely, that of uniform spacings) for all three of the specific models given above, the asymptotic theory under the alternatives is drastically different for each model. Consequently, particularly in the study of asymptotic power, asymptotic relative efficiency, and limiting distributions under contiguous alternatives must be made separately for each case and different techniques must be used.

3. Uniform spacings

The theory here is essentially complete (see [17]). Let $U_1 \leqq \cdots \leqq U_n$ be uniform (0, 1) order statistics and let $\{D_{n,i} = n(U_i - U_{i-1}): 1 \leqq i \leqq n + 1\}$ be the set of $n + 1$ weighted spacings, with $U_0 = 0$ and $U_{n+1} = 1$. Let $\{Y_i: i \geqq 1\}$ be independent exponential r.v. of mean 1. Let

$$S_n = n^{-1/2} \sum_{i=1}^{n+1} (Y_i - 1). \tag{3.1}$$

Write

$$\mathbf{D}_n = (D_{n,1}, \cdots, D_{n,n+1}), \qquad \mathbf{Y}_n = (Y_1, \cdots, Y_{n+1}), \tag{3.2}$$

and for any Borel measurable function g on R_{n+1} let $G_n = g(\mathbf{D}_n)$ and $J_n = g(\mathbf{Y}_n)$. It is well known that $\mathscr{L}(G_n) = \mathscr{L}(J_n | S_n = 0)$. Functions of uniform spacings have been studied by several authors; the first general methodology was given by Darling [10]. The most general method available today for proving limit theorems for functions of spacings was introduced by LeCam [14], who applied classical limit theory to (J_n, S_n) and utilized the representation of spacings as exponential random variables conditioned by $S_n = 0$.

Using the same general approach, this result was generalized for a wider class of functions g by Pyke [17] in 1965 and by Wichura [25] in 1968. In the latter case, conditional distributions given $S_n = x \neq 0$ are also considered. Bickel

[2] in 1969 also generalized the theorem for the case of sums to include conditional distributions given $S_n = x \neq 0$ and obtained for this case uniformity in x.

Although the conditional construction $\mathscr{L}(D_n) = \mathscr{L}(Y_n | S_n = 0)$ provides a relatively simple collection of limit theorems for uniform spacings, there remains at least one striking open question.

Problem 1 (Uniform spacings under random sample size). Let $\{N_t: t \geqq 0\}$ be a positive integer valued process for which $N_t/t \xrightarrow{P} 1$ as $t \to \infty$. For G_n as defined above, show that $G_n \xrightarrow{L} G$ implies that $G_{N_t} \xrightarrow{L} G$ for a large class of functions g.

Random sample size versions of limit theorems have previously been obtained, for example, for sums, maxima and empirical processes of independent, identically distributed random variables; but nothing has been obtained for spacings. For example, it is unknown whether or not G_{N_t} converges in law for such a simple statistic as $G_n = n^{-1/2} \Sigma_{i=1}^{n+1} (D_{n,i} - 1)^2$.

4. Limit theory under alternatives: Model III

As indicated above, several different models lead to test statistics which under the null hypothesis are functions of uniform spacings. However, when one turns to questions of distribution theory under alternative hypotheses, one finds that each model poses distinct problems. The easiest model to work under alternatives is Model III, for in the case of a renewal process the interoccurrence times $\{T_i - T_{i-1}\}$ remain independent and identically distributed under the alternatives. Consequently most results can be derived directly from standard theory. For example, consider the problem of establishing the weak convergence of the empirical process of the spacings. Under Model III, $D_{n,i} = X_i/\bar{X}_n$ for $1 \leqq i \leqq n + 1$ where $\bar{X}_n = (X_1 + \cdots + X_{n+1})/n$ and where $\{X_i: i \geqq 1\}$ are independent with common d.f. F. Assume that F is continuous, $E(X_1) = 1$ and $\operatorname{Var}(X_1) = \sigma^2 < \infty$. Let F_n denote the empirical d.f. of $\{X_1, \cdots, X_{n+1}\}$ and let H_n denote the same of $\{D_{n,1}, \cdots, D_{n,n+1}\}$. (Excuse the unconventional but convenient use of n for $n + 1$ in the subscripts of $\bar{X}_n$, F_n, and H_n.) Then $H_n(x) = F_n(x\bar{X}_n)$ for all x. Write

$$U_n = n^{1/2}(F_n \circ F^{-1} - e), \qquad V_n = n^{1/2}(H_n \circ F^{-1} - e) \tag{4.1}$$

for the empirical processes of $\{X_i\}$ and $\{D_{n,i}\}$, respectively, defined on (0, 1), where e denotes the identity function $e(x) = x$. Elementary algebra shows that for all x

$$V_n(F(x)) = U_n(F(x\bar{X}_n)) + \left\{\frac{F(x\bar{X}_n) - F(x)}{x(\bar{X}_n - 1)}\right\} xZ_n \tag{4.2}$$

where $Z_n = n^{1/2}(\bar{X}_n - 1)$. (The second term is taken to be zero when $x = 0$.) Notice that

$$Z_n = n^{1/2}(\bar{X}_n - 1) = -n^{1/2} \int_0^\infty [F_n(x) - F(x)]\, dx = - \int_0^1 U_n\, dF^{-1}. \tag{4.3}$$

Since it is known that $U_n \xrightarrow{L} U_0$ where U_0 is a Brownian bridge, one may, without loss of generality for our purposes, assume that U_n converges uniformly to U_0. (See Pyke [18] and Pyke and Shorack [19].) When this is the case it follows from (4.3) that $Z_n \xrightarrow{P} Z_0$ where

$$Z_0 = -\int_0^1 U_0\, dF^{-1} \tag{4.4}$$

is a $N(0, \sigma_1^2)$ r.v. with

$$\sigma_1^2 = 2\int_0^1 \int_0^v u(1 - v)\, dF^{-1}(u)\, dF^{-1}(v) < \infty. \tag{4.5}$$

To show this, use uniform convergence to get

$$\int_\varepsilon^{1-\varepsilon} U_n\, dF^{-1} \to \int_\varepsilon^{1-\varepsilon} U_0\, dF^{-1} \tag{4.6}$$

for any $\varepsilon > 0$. Then use Chebyshev's inequality to show that the remaining integrals can be made small in probability in view of the finiteness of σ_1^2, a direct consequence itself of the finiteness of $E(X_1^2)$. This establishes the fact that $(U_n, Z_n) \xrightarrow{L} (U_0, Z_0)$ on the natural product space. It is now convenient again to use equivalent constructions as in [18] for which $Z_n \to Z_0$ and U_n converges uniformly to U_0. The expression in (4.2) then suggests that under suitable smoothing and tail conditions on $f = F'$, one should obtain that V_n converges uniformly to

$$V_0 = U_0 + (f \circ F^{-1})F^{-1} Z_0. \tag{4.7}$$

Hence V_0 is a Gaussian process of zero mean whose covariance function, by virtue of the special construction of Z_0, can be computed to be

$$E[V_0(u)V_0(v)] = u(1 - v) + h(u)h(v)\sigma^2 + h(u)g(v) + h(v)g(u) \tag{4.8}$$

for $0 \leqq u \leqq v \leqq 1$ where $h = (f \circ F^{-1})F^{-1}$ and

$$g(u) = E[Z_0 U_0(u)] = -\int_0^1 (u \wedge w - uw)\, dF^{-1}(w). \tag{4.9}$$

Sufficient conditions for this to hold are given in the following result.

THEOREM 4.1. *If* (i) *f satisfies a Lipschitz condition on bounded intervals and* (ii) $\sup_{|x|>T} f(x)x \to 0$ as $T \to \infty$, *then V_n converges in law to V_0, a Gaussian process of mean zero and covariance given by* (4.8).

PROOF. Using the special constructions described above it suffices to establish that

$$\sup_x \left| \frac{F(\bar{X}_n x) - F(x)}{(\bar{X}_n - 1)x} Z_n - f(x) Z_0 \right| |x| \to 0, \tag{4.10}$$

or equivalently that $\sup_x |f(\theta_{n,x})Z_n - f(x)Z_0||x| \to 0$, where $\theta_{n,x}$ is some value between x and $\bar{X}_n x$. Since $Z_n \to Z_0$, condition (ii) enables one to delete Z_n and

Z_0 from the expression. Then by (i), there exists for each T a constant M such that

$$\sup_{|x|<T} |f(\theta_{n,x}) - f(x)||x| \leqq M|1 - \bar{X}_n|T^2 \to 0. \tag{4.11}$$

Condition (ii) guarantees that the supremum can be made small on the complement of $(-T, T)$. *Q.E.D.*

The above result does not use the assumption that $X_i \geqq 0$ which would ordinarily be assumed within the context of Model III. Weak convergence could also be obtained relative to the stronger metrics ρ_q of [19], which are defined by $\rho_q(f, h) = \sup |f - h|/q$, for a wide class of weighting functions q. Along this line we quote the following result of Shorack [21] which involves, however, the inverse empirical (or quantile) process rather than the empirical process. The former has some implicit simplicities over the latter. Although results are given in [21] for the general case, for simplicity of notation we state the result only for the uniform case.

THEOREM 4.2 (Shorack). *If F is an exponential* d.f. *of mean* 1,

$$W_n(t) = n^{1/2}[D_{n,i} - F^{-1}(t)] \qquad \text{for} \quad i - 1 < (n + 1)t \leqq i, \tag{4.12}$$

and q is a nonnegative function which is nondecreasing (nonincreasing) on $(0, 1/2]$ $([1/2, 1))$ *and whose reciprocal is square integrable over* $(0, 1)$, *then* $W_n \xrightarrow{L} W$ *relative to the metric* $\rho_{q/(1-e)}$.

It should be remarked that if one deletes the factor $n^{1/2}$ from (4.2), it is obvious that the Glivenko-Cantelli theorem for H_n holds whenever $\sup |x|f(x) < \infty$. This observation is due to Bickel and provides a simpler proof of the Glivenko-Cantelli theorem given in [17].

5. Limit theory under alternatives: Model I

When describing alternatives for tests based on spacings, one should keep in mind that spacings tests are only appropriate for problems whose alternative hypotheses involve the shape of the density functions. Two suitable problems of interest might be

$$H_0: F_X \text{ is uniform } (0, 1) \text{ versus } H_1 : f_X \searrow \text{ on } (0, \infty) \tag{5.1}$$

and

$$H_0: F_X \text{ is exponential, versus } H_1 : F_X \text{ is IFR (that is } f_X(1 - F_X)^{-1} \searrow) \tag{5.2}$$

Both of these alternatives restrict the shape of the density function f_X; other examples might involve the unimodality or bimodality of f_X or the monotoneity of f_X'/f_X.

For the problem in (5.2), Proschan and Pyke [16] proposed a family of tests based on the normalized spacings $\bar{D}_{n,i} = (n - i + 1)(T_i - T_{i-1})$. Under H_0 these normalized spacings are independent and identically distributed whereas under H_1 they are stochastically decreasing. The test statistics were of the form

$$G_n = \sum_{i<j} g(\bar{D}_{n,i}, \bar{D}_{n,j}) \tag{5.3}$$

for bounded nonnegative g satisfying $g(\cdot, y) \searrow$ and $g(x, \cdot) \nearrow$ for all x, y. (It is well to emphasize that only nonhomogeneous functions of spacings should be considered in Model I.) Central limit theorems for these statistics are given in [16] and [17]. Of particular interest is the Wilcoxon-like statistic, V_n = number of pairs $(\bar{D}_{n,i}, \bar{D}_{n,j})$ with $\bar{D}_{n,i} > \bar{D}_{n,j}$ for $i < j$. Similar statistics suggest themselves for testing (5.1), but in terms of the regular spacings $D_{n,i} = T_i - T_{i-1}$.

All approaches to limit theorems for alternatives under Model I can be said to depend upon Taylor's expansion methods. The dependency of the spacings causes considerable difficulties, but basically the approach is to observe that when $F_X = F$,

$$nD_{n,i} \stackrel{L}{=} n[F^{-1}(Y_i/S_{n+1}) - F^{-1}(Y_{i-1}/S_{n+1})] = Y_i/f(F^{-1}(\theta_{n,i}))\bar{X}_n \tag{5.4}$$

or

$$\bar{D}_{n,i} \stackrel{L}{=} (n - i + 1)[F^{-1} \circ H(Y_i^*) - F^{-1} \circ H(Y_{i-1}^*)] \stackrel{L}{=} Y_i r(A_{n,i}), \tag{5.5}$$

where $\{Y_i\}$ are independent exponential random variables with mean one, H is their common d.f., $\{Y_i^*\}$ denote exponential order statistics, $r = (1 - e)/f \circ F^{-1}$ is the reciprocal of the failure rate defined through F^{-1} on $(0, 1)$, and where $\{\theta_{n,i}\}$, $\{A_{n,i}\}$ are defined by appropriate Taylor expansions. Thus although the spacings are dependent, the above representations indicate their approximate independence and exponentiality. Except when working under contiguous alternatives, the errors in these approximations make a significant cumulative contribution. To illustrate this, consider somewhat simpler statistics of the form $G_n^* = \Sigma\, g(\bar{D}_{n,i}, i/n)$. In [17] it is shown that under regularity assumptions on g and f_X, $n^{1/2}(G_n^* - R_n) \xrightarrow{P} 0$ where

$$R_n = \sum [g(Y_i r(i/n), i/n) + (Y_i - 1)C(i/n)] \tag{5.6}$$

and

$$C(w) = (1 - w)^{-1} \int_w^1 (1 - u)[r'(u)/r(u)] \int_0^\infty e^{-y}(y - 1)g(yr(u), u)\, dy du. \tag{5.7}$$

This result shows how the errors cumulate to give the second summand in (5.6). It is interesting to compare the form of this result with that of Bickel and Doksum [3], derived under contiguous alternatives, which states that $n^{-1/2}(T_n - S_n) \xrightarrow{P} 0$ where $T_n = \Sigma\, h(X_i)$, $S_n = \Sigma\, a(i/n)\,(\bar{D}_{n,i} - 1)$, and h and a are functions satisfying certain regularity assumptions which are related by

$$a(u) = (1 - u)^{-1} \int_{-\log(1-u)}^\infty h'(x)e^{-x}\, dx. \tag{5.8}$$

In [3], Bickel and Doksum show that under contiguous alternatives $\{f_{\theta_n}\}$ with $\theta_n = bn^{-1/2}$ and $b \geq 0$, the test based on V_n described above is asymptotically inadmissible. They do this by showing that V_n is asymptotically equivalent to

$\Sigma\, iR_i$ (where the R_i are the ranks of the normalized spacings) which in turn is always asymptotically inferior to $-\Sigma\, i \log[1 - R_i/(n+1)]$. They show, furthermore, that among "studentized" linear functions of spacings or of their ranks there exist asymptotically most powerful ones. By analogy, the result suggests the following.

Problem 2. For the problem (5.1) of testing uniformity against monotoneity (or unimodality), do asymptotically most powerful tests exist among those based upon linear functions of (nonnormalized) spacings?

Other results involving limit theory under contiguous alternatives include the following one by Weiss [22]. Assume for X a sequence of distribution functions G_n with density functions $g_n = 1 + n^{-\delta} r$ on $(0, 1)$, where $\delta > 0$ and r is a function whose second derivative exists and is uniformly bounded in absolute value. Let $\mathbf{W} = (W_1, \cdots, W_{n+1})$ be independent exponential random variables with W_i having mean $g_n \circ G_n^{-1}(i/(n+1))$ and set $Z_i = W_i/(W_1 + \cdots + W_{n+1})$. If $\mathbf{D}_n = (D_{n,1}, \cdots, D_{n,n+1})$ are the spacings when $F_X = G_n$, then Weiss proves the following theorem.

THEOREM 5.1. (Weiss). *The ratio* $f_{\mathbf{Z}_n}(\mathbf{D}_n)/f_{\mathbf{D}_n}(\mathbf{D}_n)$ *tends to* 1 *in probability.*

It is further shown that in fact Theorem 5.1 holds when convergence in probability is replaced by convergence in 1-mean, thereby implying the contiguity of the corresponding measures.

In [23] this result is used to establish the asymptotic distribution of homogeneous functions of spacings under such contiguous alternatives. Although expressed in terms of spacings, Theorem 5.1 can equivalently be viewed as a result about the order statistics $T_1 \leqq T_2 \leqq \cdots \leqq T_n$ in view of the one to one correspondence between the two. In fact, the proof in [22] notationally works with the equivalent densities of the order statistics. Thus if $V_i = Z_1 + \cdots + Z_i$, then Theorem 5.1 is equivalent to stating that $f_{\mathbf{T}_n}(\mathbf{T}_n)/f_{\mathbf{V}_n}(\mathbf{T}_n) \xrightarrow{P} 1$.

In a recent paper, Rao and Sethuraman [20], motivated by problems involving circular data, also develop limit theorems for spacings under contiguous Model I alternatives. They use the representation (5.4) and postulate, as in Weiss [22], a sequence of distribution functions G_n with densities $g_n = 1 + n^{-\delta} r$, $\delta \geqq 1/4$, where r possesses a uniformly continuous derivative r'. The approach of Rao and Sethuraman is to study the empirical processes of the spacings, and they begin by studying these in the approximate situation in which $\theta_{n,i}$ in representation (5.4) is a constant. Specifically, they study the empirical process of $\{Y_i/c_{n,i}: 1 \leqq i \leqq n\}$ where $\{Y_i\}$ are independent exponential random variables of mean one, and where the constants are determined by $c_{n,i} = c_n(i/n)$ where

$$c_n(u) = 1 + A(u)n^{-\delta} + R_n(u), \qquad \delta \geqq 1/4. \tag{5.9}$$

In this representation, the function A satisfies certain smoothness conditions and R_n is uniformly $o(n^{-1/2})$.

If H denotes the exponential d.f. of mean one, set

$$U_i(y) = I_{[Y_i \leqq y]} - H(y); \qquad W_n(y) = n^{-1/2} \sum_{i=1}^{n} U_i(y c_{n,i}) \tag{5.10}$$

for $y \geqq 0$. Then W_n is the empirical process of $\{Y_i/c_{n,i}\}$ and may be written as

$$W_n(y) = n^{-1/2} \sum_{i=1}^{n} U_i(y) + n^{-1/2} \sum_{i=1}^{n} [U_i(yc_{n,i}) - U_i(y)] = V_n(y) + Q_n(y), \qquad y \geqq 0. \tag{5.11}$$

Here, V_n is the usual empirical process for an exponential sample whose weak convergence is known. The remainder term Q_n is bounded in absolute value by the modulus of continuity of the exponential empirical process over intervals of width

$$\sup \{|1 - c_{n,i}|: 1 \leqq i \leqq n\} \leqq n^{-\delta} \sup \{|A(u)|: 0 \leqq u \leqq 1\}. \tag{5.12}$$

However, only for $\delta > 1/2$ is the bound sufficient to enable one to deduce $Q_n \xrightarrow{P} 0$ directly from the weak convergence of V_n. To establish the limiting behavior of Q_n for general δ one must make better use of the fact that Q_n is a sum of n independent processes and apply the standard methods of weak convergence (see [4]). In passing, observe that for any $\delta > 0$, $Q_n(y) \xrightarrow{P} 0$ for each *fixed* y. To verify this, compute

$$\operatorname{Var} Q_n(y) \leqq n^{-1} \sum_{i=1}^{n} |e^{-y} - e^{-c_{n,i}y}| = n^{-1} \sum_{i=1}^{n} u|1 - u^{c_{n,i}-1}| \tag{5.13}$$

for $u = e^{-y}$. Since $u|1 - u^b| \leqq |1 - (1 + b)^{-1}|(1 + b)^{-1/b} \leqq |1 - (1 + b)^{-1}|$ for $0 \leqq u \leqq 1$ and real b, it follows that

$$\operatorname{Var} Q_n(y) \leqq n^{-1} \sum_{i=1}^{n} |1 - c_{n,i}{}^{-1}| \tag{5.14}$$

which converges to zero under even weaker assumptions than (5.9).

The study of empirical processes of "perturbed" random variables also has applications in the theory of rank statistics when the underlying independent distribution functions are not identical.

Problem 3. Let $\{F_{n,i}: 1 \leqq i \leqq n, n \geqq 1\}$ be a triangular array of distribution functions and let $\{X_i: i \geqq 1\}$ be a sequence of independent uniform (0, 1) random variables. Find necessary and sufficient conditions on $\{F_{n,i}\}$ to ensure the weak convergence of the empirical processes

$$b_n \sum_{i=1}^{n} \{I_{[X_i \leqq F_{n,i}(y)]} - F_{n,i}(y)\}, \qquad -\infty < y < \infty, \tag{5.15}$$

for suitable constants $\{b_n\}$.

6. Limit theory under alternatives: Model II

This is a situation of considerable importance and the reader is referred to Cox and Lewis [9] for a review of the pertinent literature. Let it suffice here to mention the following general question, suggested by the results of Weiss [22], Bickel and Doksum [3], and Rao and Sethuraman [20] described above.

Problem 4. Let $\{\lambda_\theta : \theta \geqq 0\}$ be a family of intensity functions for nonhomogeneous Poisson processes in which $\lambda_0 \equiv 1$. Find sufficient conditions of these functions to ensure that a sequence $\{P_{\theta_n}\}$ of corresponding measures is contiguous with P_0, the measure for the Poisson process of rate 1.

7. Other results

In this section, capsule descriptions of other recent references on spacings are given.

In [5] and [6], Blumenthal studies the limit theory of statistics of the form $\Sigma_{i=1}^n \{D_i^X/D_i^Y\}^r$ where $\{D_i^X\}$ and $\{D_i^Y\}$ are the spacings of two independent samples, with common d.f. F_X and F_Y, respectively, and where r is a constant. The proofs are difficult, indicating again the particular complexities associated with spacings and the need for more general methods. A study of the limiting behavior of statistics of the form $\Sigma_i\, g(D_i^X, D_i^Y)$ suggests itself. In [7], Blumenthal establishes a strong limit theorem in connection with the two sample spacings problem. Note the open problems stated on p. 112 of [7].

In [13], Kale derives some of the standard spacings' statistics as functions of the data that minimize certain distances between the empirical and null hypothesis d.f. In particular the statistics $\Sigma\, D_{n,i}^{-1}$, $\Sigma\, (D_{n,i} - 1/n)^2$, $\Sigma \log D_{n,i}$, $\min D_{n,i}$ and $\max D_{n,i}$ can be obtained in this way. See also [11] and [12].

The asymptotic distribution of the k smallest spacings, for fixed k, is obtained by Weiss [24] for the case of a continuous density f_X over (0, 1) which is bounded away from zero.

In unpublished papers, Blumenthal [8] derives the asymptotic normality of $\Sigma \log D_{n,i}$ for Model I alternatives satisfying restrictive smoothness conditions, while Shorack [21] derives the weak convergence of the empirical processes of spacings relative to a wide class of metrics and applies his results to tests based on linear combinations of functions of ordered spacings.

8. Miscellaneous problems

Problem 5. A study of rates of convergence for limiting distributions of spacings would be of interest. No results are presently available. In the case of uniform spacings presumably classical theory could be applied by making use of LeCam's approach [14].

Problem 6. Show that

$$\sum_{i,j} \int\int |f_{D_{n,i},D_{n,j}}(x, y) - f_{D_{n,i}}(x) f_{D_{n,j}}(y)|\,dxdy = O(n) \tag{8.1}$$

for a large class of underlying densities f_X. This would give an applicable measure of the rate of asymptotic independence of spacings; (see [17], p. 434).

Problem 7. If, as in [17], p. 420, one defines the empirical distribution functions of the spacings $D_{n,i}$ and $\bar{D}_{n,i}$, respectively, by

$$L_n(x) = n^{-1} \sum_i I[D_{n,i} \leqq x], \qquad M_n(x) = n^{-1} \sum_i I[\bar{D}_{n,i} \leqq x] \tag{8.2}$$

and the associated processes by

$$(8.3) \qquad Y_n(x) = n^{1/2}[L_n(x) - L(x)], \qquad Z_n(x) = n^{1/2}[M_n(x) - M(x)]$$

where

$$(8.4) \qquad L(x) = 1 - \int_{-\infty}^{\infty} f(y)e^{-xf(y)}\,dy, \qquad M(x) = 1 - \int_0^1 e^{-x/r(u)}\,du.$$

can one verify the weak convergence of these processes? In [17] the convergence of the finite dimensional distribution functions is established rather straightforwardly, but the weak convergence has not been shown. Such a result could possibly have application to the rank statistics of Bickel and Doksum [3].

◇ ◇ ◇ ◇ ◇

ADDENDUM

The discussion in Section 5 of [20] was based on a preprint of the same title (Tech. Rpt. No. Math-Stat/11/69-Mar. 1969, Indian Statist. Instit.). Reference [20] is given since the results there are more general and accessible than those of the preprint.

REFERENCES

[1] L. Abramson, "The distribution of the smallest sample spacing," (abstract), *Ann. Math. Statist.*, Vol. 37 (1966), p. 1421.

[2] P. J. Bickel, "Tests for monotone failure rate. II," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1250–1260.

[3] P. J. Bickel and K. A. Doksum, "Tests for monotone failure rate based on normalized spacings," *Ann. Math. Statist.*, Vol. 40 (1969), pp. 1216–1235.

[4] P. Billingsley, *Convergence of Probability Measures*, New York, Wiley, 1968.

[5] S. Blumenthal, "Contributions to sample spacings theory, I: Limit distributions of sums of ratios of spacings," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 904–924.

[6] ———, "Contributions to sample spacings theory. II: Tests of the parametric goodness of fit and two-sample problems," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 925–939.

[7] ———, "Limit theorems for functions of shortest two-sample spacings and a related test," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 108–116.

[8] ———, "Logarithms of sample spacings," *SIAM J. Appl. Math.*, Vol. 16 (1968), pp. 1184–1191.

[9] D. R. Cox and P. A. W. Lewis, *Statistical Analysis of Series of Events*, London, Methuen, 1966.

[10] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Statist.*, Vol. 28 (1953), pp. 823–838.

[11] J. R. Gebert and B. K. Kale, "Goodness of fit tests based on discriminatory information," *Statistiche Hefte*, Vol. 10 (1969), pp. 197–200.

[12] V. P. Godambe and B. K. Kale, "A test of goodness of fit," *Statistiche Hefte*, Vol. 8 (1967), pp. 165–172.

[13] B. K. Kale, "Unified derivation of tests of goodness of fit based on spacings," *Sankhyā*, *Ser. A*, Vol. 31 (1969), pp. 43–48.

[14] L. LeCam, "Un théorème sur la division d'un intervalle par des points pris au hasard," *Publ. Inst. Statist. Univ. Paris*, Vol. 7 (1958), pp. 7–16.

[15] J. I. Naus, "Some probabilities, expectations and variances for the size of largest clusters and smallest intervals," *J. Amer. Statist. Assoc.*, Vol. 61 (1966), pp. 1191–1199.

[16] F. Proschan and R. Pyke, "Tests for monotone failure rate," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 3, pp. 293–312.

[17] R. Pyke, "Spacings," *J. Roy. Statist. Soc., Ser. B*, Vol. 7 (1965) pp. 395–449.

[18] ———, "Applications of almost surely convergent constructions of weakly convergent processes," *International Symposium on Probability and Information Theory, Lecture Notes on Mathematics*, 89 (edited by M. Behara, K. Krickeberg, and J. Wolfowitz), Berlin, Heidelberg, New York, Springer-Verlag, 1969.

[19] R. Pyke and G. R. Shorack, "Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 755–771.

[20] J. S. Rao and J. Sethuraman, "Pitman efficiencies of tests based on spacings," *Proceedings of the First International Symposium on Non-Parametric Techniques in Statistical Inference*, New York, Cambridge University Press, 1970, pp. 405–415.

[21] G. R. Shorack, "Convergence of spacings' processes with applications," to be published.

[22] L. Weiss, "On asymptotic sampling theory for distributions approaching the uniform distribution," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 4 (1965), pp. 217–221.

[23] ———, "Limiting distributions of homogeneous functions of sample spacings for distributions approaching the uniform distribution," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 10 (1968), pp. 193–197.

[24] ———, "The joint asymptotic distribution of the k-smallest sample spacings," *J. Appl. Probability*, Vol. 6 (1969), pp. 442–448.

[25] M. Wichura, "On the weak convergence of non-Borel probabilities on a metric space," Ph.D. thesis, Columbia University, 1968.

[26] S. Blumenthal, "Tests of fit based on partial sums of the ordered spacings," *Ann. Inst. Statist. Math.*, Vol. 22 (1970), pp. 261–276.