The Annals of Statistics
1975, Vol. 3, No. 2, 299-313

WEAK CONVERGENCE OF EMPIRICAL DISTRIBUTION FUNCTIONS OF RANDOM VARIABLES SUBJECT TO PERTURBATIONS AND SCALE FACTORS

BY J. S. RAO[1] AND J. SETHURAMAN[2]

Indiana University and Florida State University

The weak convergence of empirical distribution functions subject to random perturbations and scale factors to a Gaussian process is established. This result is used to study the efficiencies of tests based on spacings in goodness-of-fit problems.

1. Introduction and summary. The weak convergence of empirical distribution functions of independent identically distributed random variables is well known and has been studied by several authors, notably Doob (1949) and Donsker (1951). In an earlier paper of ours, Sethuraman and Rao (1970), while studying the asymptotic efficiencies of tests based on spacings, we found a need to study the weak convergence of empirical distribution functions of random variables subject to random perturbations and scale factors. We study this problem in this paper.

The statistical problem of testing goodness-of-fit using spacings tests is briefly as follows. Let $X_1, X_2, \cdots, X_{n-1}$ be $(n-1)$ independent random variables with a common distribution function. The goodness-of-fit problem is to test if this distribution function is equal to a specified one. A simple probability integral transformation on the random variables would permit us to equate the specified distribution function to the uniform distribution on [0, 1]. Thus, from now on, we shall assume that this reduction has been effected and under the hypothesis, the observations have a uniform distribution on [0, 1].

Let $X_1' \leqq X_2' \leqq \cdots \leqq X'_{n-1}$ be the order statistics. The sample spacings $(D_1, \cdots, D_n)$ are defined by

$$D_i = X_i' - X'_{i-1} \qquad\qquad i = 1, \cdots, n$$

where we put $X_0' = 0$, $X_n' = 1$. Clearly the carrier of the distribution function must be [0, 1] in order that this definition of the sample spacings is meaningful. Tests for goodness-of-fit problems based on the spacings have been proposed by

Received December 1971; revised March 1974.

[1] Research supported by the Air Force Office of Scientific Research, AFSC, USAF under Grant No. AFOSR71-2009.

[2] Research supported by the Office of Naval Research under Contract No. NONR 988(08), Task Order NRO42-004 and The U.S. Army Research Office (Durham) under Grant No. DA-ARO-D-31-124-72-G167. The United States Government is authorized to reporduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

AMS 1970 *classification numbers.* Primary 62E20, 62G20, 62G10; Secondary 60F05.

Key words and phrases. Weak convergence, perturbations and scale factors, empirical distribution functions, spacings, efficiencies of tests, goodness-of-fit tests.

several authors. See, for instance, Pyke (1965) or Sethuraman and Rao (1970) and the references contained therein. Important among them are tests based on

$$U(n) = \sum_{i=1}^{n} |D_i - 1/n| ,$$
$$L(n) = \sum_{i=1}^{n} \log (nD_i)/n ,$$

and the class of statistics

$$V_r(n) = \sum_{i=1}^{n} (nD_i)^r/n , \qquad r > -\tfrac{1}{2} .$$

We derive the asymptotic distributions of all these statistics under a smooth sequence of alternatives (see Section 3) using a unified approach. This allows us to compute the Pitman efficiencies of these various tests as was done in Sethuraman and Rao (1970).

The material of this paper is divided into three sections. Section 2 treats the weak convergence problem and is independent of Sethuraman and Rao (1970). Section 3 shows how one can use the weak convergence results of Section 2 in problems connected with spacings thus relating the present work to Sethuraman and Rao (1970).

2. Weak convergence of empirical distribution functions subject to perturbations and scale factors. Let $Z_1, Z_2, \cdots$ be independent and identically distributed random variables with a *strictly increasing continuous* distribution function $F(x)$ with $F(0) = 0$. Let $\{\alpha_{n1}, \alpha_{n2}, \cdots, \alpha_{nn}\}$, $\bar{Z}_n$, $n = 1, 2, \cdots$ be *positive* random variables that may depend on $Z_1, Z_2, \cdots$.

Let

(2.1) $$I(y; x) = 1 , \quad \text{if } y \leqq x , \qquad = 0 , \quad \text{if } y > x ,$$

and write

(2.2) $$F_n(x) = \sum_{i=1}^{n} I(Z_i/\alpha_{ni}; x)/n ,$$

(2.3) $$\bar{F}_n(x) = \sum_{i=1}^{n} I(Z_i/(\alpha_{ni}\bar{Z}_n); x)/n .$$

Then we say that $F_n(x)[\bar{F}_n(x)]$ is the empirical distribution function of $\{Z_1/\alpha_{n1}, \cdots, Z_n/\alpha_{nn}\}[\{Z_1/(\alpha_{n1}\bar{Z}_n), \cdots, Z_n/(\alpha_{nn}\bar{Z}_n)\}]$, that is, of $\{Z_1, \cdots, Z_n\}$ subject to perturbations $\{\alpha_{n1}, \cdots, \alpha_{nn}\}$ [perturbations $\{\alpha_{n1}, \cdots, \alpha_{nn}\}$ and scale factor $\bar{Z}_n$]. The distinction between a perturbation, α_{ni}, and a scale factor, $\bar{Z}_n$, arises from the conditions they satisfy, which are of the form

$$\alpha_{ni} = 1 + o_p(n^{-\frac{1}{2}}) ,$$
$$\bar{Z}_n = 1 + O_p(n^{-\frac{1}{2}}) .$$

(For any random variable X_n we write $X_n = o_p(n^{-\frac{1}{2}})$ if $n^{\frac{1}{2}}X_n \to 0$ in probability and we write $X_n = O_p(n^{-\frac{1}{2}})$ if for each $\varepsilon > 0$, there is a $K_\varepsilon < \infty$ such that $P\{|n^{\frac{1}{2}}X_n| > K_\varepsilon\} < \varepsilon$ for all n.) These conditions are made more precise in (2.13), (2.18) and (2.23).

Define

$$G_n(x) = \sum_{i=1}^{n} F(x\alpha_{ni})/n \,, \tag{2.4}$$

$$\eta_n(x) = n^{\frac{1}{2}}(F_n(x) - G_n(x)) \,, \tag{2.5}$$

$$\bar{\eta}_n(x) = n^{\frac{1}{2}}(\bar{F}_n(x) - G_n(x)) \,, \qquad 0 \leqq x \leqq \infty \tag{2.6}$$

where we put $\eta_n(+\infty) = \bar{\eta}_n(+\infty) = 0$.

Our main aim in this section is to prove that $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ and $\{\bar{\eta}_n(x), 0 \leqq x \leqq \infty\}$ converge weakly to Gaussian processes under suitable conditions on $F(x)$, $\{\alpha_{n1}, \cdots, \alpha_{nn}\}$ and $\bar{Z}_n$. The main results are found in Theorems 2.4, 2.5, 2.6, and 2.8 and Corollaries 2.7 and 2.9.

In order to talk about weak convergence of the above stochastic processes it is necessary to introduce the space $D[0, \infty]$ of functions $p(x)$ on $[0, \infty]$ which satisfy the properties

(i) $p(x + 0)$, $p(x - 0)$ exist for each x in $(0, \infty)$ and $p(x) = p(x + 0)$,
(ii) $p(0 + 0)$ exists and is equal to $p(0)$,
(iii) $\lim_{x\to\infty} p(x)$ exists and is equal to $p(\infty)$.

A sequence $\{p_n(x)\}$ in $D[0, \infty]$ converges to $p(x)$ in $D[0, \infty]$ if there exists a sequence, $\{\lambda_n(x)\}$, of monotone one-to-one continuous maps of $[0, \infty]$ onto $[0, \infty]$ such that as $n \to \infty$

$$\sup_x |\lambda_n(x) - x| \to 0 \,,$$

and

$$\sup_x |p_n(\lambda_n(x)) - p(x)| \to 0 \,.$$

This convergence corresponds to the J_1-topology of Skorohod on the space $D[0, \infty]$ and makes it a topologically complete separable metric space. See Skorohod (1956).

A sequence of stochastic processes $\{p_n(x), x \geqq 0\}$ converges weakly to a process $\{p(x), x \geqq 0\}$ in $D[0, \infty]$ if

$$E[\mathscr{S}(p_n(\cdot))] \to E[\mathscr{S}(p(\cdot))] \tag{2.7}$$

for every functional $\mathscr{S}(\cdot)$ on $D[0, \infty]$ which is bounded and continuous in the topology just described. When this happens, we have, by the invariance principle, the very useful conclusion that the distribution of the real-valued random variable $\mathscr{S}(p_n(\cdot))$ converges weakly to the distribution of $\mathscr{S}(p(\cdot))$ for every functional $\mathscr{S}(\cdot)$ on $D[0, \infty]$ which is continuous a.e. with respect to $\{p(x), x \geqq 0\}$.

Let $\mu(x)$ be a one-to-one monotone continuous map of $[0, \infty]$ onto $[0, 1]$. For any function $q(\mu)$ in $D[0, 1]$, let $p(x) = q(\mu(x))$. This map from $D[0, 1]$ to $D[0, \infty]$ is continuous and has a continuous inverse (when $D[0, 1]$ is endowed with the J_1-topology of Skorohod) and is therefore a homeomorphism. Thus the study of convergence of probability measures on $D[0, \infty]$ can be reduced to the study of convergence of probability measures on $D[0, 1]$, which is by now classical. See for instance Skorohod (1956), Sethuraman (1965), Billingsley

(1968). The following theorem gives a well-known sufficient condition for the compactness and convergence of a sequence of stochastic processes in $D[0, 1]$; see Chentsov (1956), Sethuraman (1965) or Billingsley (1968).

THEOREM 2.1. *Let* $\{q_n(\mu), 0 \leqq \mu \leqq 1\}$, $n = 1, 2, \cdots$ *and* $\{q(\mu), 0 \leqq \mu \leqq 1\}$ *be stochastic processes with values in* $D[0, 1]$ *such that*

(i) *the marginal distributions of* $\{q_n(\mu_1), \cdots, q_n(\mu_k)\}$ *converge to that of* $\{q(\mu_1), \cdots, q(\mu_k)\}$ *weakly for every finite subset* $\{\mu_1, \cdots, \mu_k\}$ *of* $[0, 1]$ *and*

(ii) *there exists a constant C such that*

$$E\{|q_n(\mu_1) - q_n(\mu_2)|^2 |q_n(\mu_2) - q_n(\mu_3)|^2\} \leqq Ch^2 , \tag{2.8}$$

whenever $0 \leqq \mu_1 \leqq \mu_2 \leqq \mu_3 \leqq 1$ *and* $\mu_3 - \mu_2 \leqq h$, $\mu_2 - \mu_1 \leqq h$. *Then the sequence of stochastic processes* $\{q_n(\mu), 0 \leqq \mu \leqq 1\}$ *converges weakly to the process* $\{q(\mu), 0 \leqq \mu \leqq 1\}$.

Now, let $\mu(x)$ be a one-to-one monotone continuous transformation of $[0, \infty]$ onto $[0, 1]$. Let $p_n(x) = q_n(\mu(x))$ and $p(x) = q(\mu(x))$, $0 \leqq x \leqq \infty$. From our remark earlier about the homeomorphism between $D[0, 1]$ and $D[0, \infty]$ induced by μ, the sequence of stochastic processes $\{q_n(\mu), 0 \leqq \mu \leqq 1\}$ converges weakly to $\{q(\mu), 0 \leqq \mu \leqq 1\}$ if and only if the sequence of processes $\{p_n(x), 0 \leqq x \leqq \infty\}$ converges weakly to $\{p(x), 0 \leqq x \leqq \infty\}$. This provides us with a technique of investigating the convergence of our empirical distribution functions.

After this digression on the definition of $D[0, \infty]$ and weak convergence of processes on it, we return to our empirical distribution function processes. We note that for each n, $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ and $\{\bar{\eta}_n(x), 0 \leqq x \leqq \infty\}$ are in $D[0, \infty]$ and are measurable. We now state two lemmas and go on to the main theorems.

LEMMA 2.2. *For* $n = 1, 2, \cdots$, *let* $\{Y_{ni}, i = 1, \cdots, n\}$ *be independently distributed with*

$$P(Y_{ni} = 1) = p_{ni} , \qquad P(Y_{ni} = 0) = 1 - p_{ni} , \qquad i = 1, \cdots, n . \tag{2.9}$$

Let

$$Y_n = \textstyle\sum_{i=1}^n (Y_{ni} - p_{ni})/[\sum_{i=1}^n p_{ni}(1 - p_{ni})]^{\frac{1}{2}} . \tag{2.10}$$

Then as $n \to \infty$

$$P(Y_n \leqq x) \to \Phi(x)$$

for each x, *where* $\Phi(x) = \int_{-\infty}^x \exp(-t^2/2)\, dt/(2\pi)^{\frac{1}{2}}$, *if and only if*

$$S_n^2 = \textstyle\sum_{i=1}^n p_{ni}(1 - p_{ni}) \to \infty . \tag{2.11}$$

See Fisz (1963, page 207) for a proof. A sufficient condition for (2.11) to hold is that $\sum_{i=1}^n p_{ni}/n$ be bounded away from 0 and 1.

LEMMA 2.3. *Let* (Y_1, Y_2, Y_3) *be a trinomial random variable with* $P((Y_1, Y_2, Y_3) = (1, 0, 0)) = p_1$, $P((Y_1, Y_2, Y_3) = (0, 1, 0)) = p_2$, $P((Y_1, Y_2, Y_3) = (0, 0, 1)) = p_3$ *and* $p_1 + p_2 + p_3 = 1$. *Let* $Y_i^* = (Y_i - p_i)$, $i = 1, 2, 3$. *Then*

$$E(Y_i^*) = 0 , \qquad E(Y_i^{*2}) = p_i(1 - p_i) , \qquad E(Y_i^* Y_j^*) = -p_i p_j ,$$

$$E(Y_i^* Y_j^*)^2 = p_i p_j(1 - p_i)(1 - p_j) - p_i p_j(1 - 2p_i)(1 - 2p_j) , \tag{2.12}$$

$$i \neq j,\ i, j = 1, 2, 3 .$$

We are now ready to state and prove the main results on the convergence of empirical distribution functions subject to perturbations and scale factors. We begin with a degenerate case.

THEOREM 2.4. *Let* $\{\alpha_{ni}, i = 1, \cdots, n\}$, $n = 1, 2, \cdots$ *be non-random and let*

(2.13) $$\max_{1 \leq i \leq n} |\alpha_{ni} - 1| \to 0 \quad \text{as} \quad n \to \infty .$$

Then the sequence of processes $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ *defined in* (2.5) *converges weakly to a Gaussian process* $\{\eta(x), 0 \leqq x \leqq \infty\}$ *in* $D[0, \infty]$ *with mean function zero and covariance kernel*

(2.14) $$K(x, y) = F(x)(1 - F(y)) \quad \text{for} \quad x \leqq y .$$

PROOF. Define the processes $\{y_n(\mu), 0 \leqq \mu \leqq 1\}$, $\{z_n(\mu), 0 \leqq \mu \leqq 1\}$ and $\{y(\mu), 0 \leqq \mu \leqq 1\}$ in $D[0, 1]$ by

(2.15) $$y_n(G_n(x)) = \eta_n(x) , \qquad z_n(F(x)) = \eta_n(x) ,$$
$$y(F(x)) = \eta(x) , \qquad 0 \leqq x \leqq \infty .$$

We will prove the following in order

(i) $\{y_n(\mu), 0 \leqq \mu \leqq 1\}$ converges weakly to $\{y(\mu), 0 \leqq \mu \leqq 1\}$,
(ii) $\{z_n(\nu), 0 \leqq \nu \leqq 1\}$ converges weakly to $\{y(\nu), 0 \leqq \nu \leqq 1\}$,
(iii) $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ converges weakly to $\{\eta(x), 0 \leqq x \leqq \infty\}$.

We start with the proof of (i). Fix x. Then

(2.16) $$\eta_n(x)/[\sum_{i=1}^n F(x\alpha_{ni})(1 - F(x\alpha_{ni}))]^{\frac{1}{2}}$$
$$= \sum_{i=1}^n [I(Z_i/\alpha_{ni}; x) - F(x\alpha_{ni})]/[\sum_{i=1}^n F(x\alpha_{ni})(1 - F(x\alpha_{ni}))]^{\frac{1}{2}} .$$

From (2.13) and the fact that $F(x)$ is continuous

$$\sum_{i=1}^n F(x\alpha_{ni})(1 - F(x\alpha_{ni}))/n \to F(x)(1 - F(x))$$

as $n \to \infty$. By applying Lemma 2.2, it is easily seen that the distribution of $\eta_n(x)$ converges weakly to that of $\eta(x)$, for each x. A similar application of the multivariate extension of Lemma 2.2 shows that the finite dimensional marginal distributions of $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ converge weakly to those of $\{\eta(x), 0 \leqq x \leqq \infty\}$. In a similar fashion, it can be seen that the finite dimensional distributions of $\{y_n(\mu), 0 \leqq \mu \leqq 1\}$ converge to those of the Gaussian process $\{y(\mu), 0 \leqq \mu \leqq 1\}$. Next, let $0 \leqq \mu_1 \leqq \mu_2 \leqq \mu_3 \leqq 1$. Then

$$y_n(\mu_j) = \eta_n(x_j) = \sum_{i=1}^n \{I(Z_i/\alpha_{ni}; x_j) - F(x_j\alpha_{ni})\}/n^{\frac{1}{2}}$$

where $G_n(x_j) = \mu_j$, $j = 1, 2, 3$. Thus

$$y_n(\mu_2) - y_n(\mu_1) = \sum_{i=1}^n V_{1i}/n^{\frac{1}{2}} , \qquad y_n(\mu_3) - y_n(\mu_2) = \sum_{k=1}^n V_{2k}/n^{\frac{1}{2}} ,$$

where

$$V_{1i} = I(Z_i/\alpha_{ni}; x_2) - I(Z_i/\alpha_{ni}; x_1) - F(x_2\alpha_{ni}) + F(x_1\alpha_{ni}) ,$$
$$V_{2k} = I(Z_k/\alpha_{nk}; x_3) - I(Z_k/\alpha_{nk}; x_2) - F(x_3\alpha_{nk}) + F(x_2\alpha_{nk}) ,$$

$i, k = 1, \cdots, n$. Since $\alpha_{ni} \to 1$ uniformly in i, we may assume without loss of generality that $\frac{1}{2} \leqq \alpha_{ni} \leqq 2$ for all n and i. Thus V_{1i} and V_{2k} are Bernoulli random variables corrected for their means. Then

$$\begin{aligned} &E\{(y_n(\mu_2) - y_n(\mu_1))^2(y_n(\mu_3) - y_n(\mu_2))^2\} \\ (2.17) \qquad &= n^{-2}E\{(\textstyle\sum_{i=1}^n V_{1i})^2(\sum_{k=1}^n V_{2k})^2\} \\ &= n^{-2}\{\textstyle\sum_{i,k} E(V_{1i}^2 V_{2k}^2) + 2\sum_{i\neq i'} E(V_{1i}V_{2i}V_{1i'}V_{2i'})\}\,. \end{aligned}$$

Let

$$p_{1i} = F(x_2\alpha_{ni}) - F(x_1\alpha_{ni})\,, \qquad p_{2k} = F(x_3\alpha_{nk}) - F(x_2\alpha_{nk})\,.$$

From (2.12) in Lemma 2.3,

$$E(V_{1i}^2 V_{2k}^2) \leqq p_{1i}p_{2k}\,, \qquad E(V_{1i}V_{2i}) = -p_{1i}p_{2i}\,.$$

Substituting these in (2.17),

$$\begin{aligned} E\{(y_n(\mu_2) - y_n(\mu_1))^2(y_n(\mu_3) - y_n(\mu_2))^2\} &\leqq 3(\textstyle\sum_i p_{1i})(\sum_i p_{2i})/n^2 \\ &= 3|\mu_2 - \mu_1|\,|\mu_3 - \mu_2|\,. \end{aligned}$$

Thus from Theorem 2.1, the sequence of processes $\{y_n(\mu), 0 \leqq \mu \leqq 1\}$ converges weakly to the Gaussian process $\{y(\mu), 0 \leqq \mu \leqq 1\}$.

To prove (ii), consider the sequence of monotone continuous one-to-one maps $\lambda_n(\cdot)$ of [0, 1] onto [0, 1] given by

$$\lambda_n(\mu) = F(G_n^{-1}(\mu))\,,$$

where $G_n^{-1}(\cdot)$ is the inverse function of $G_n(\cdot)$. It is easy to see that

$$\sup_\mu |\lambda_n(\mu) - \mu| \to 0\,, \qquad z_n(\lambda_n(\mu)) = y_n(\mu)\,.$$

Hence $\{z_n(\nu), 0 \leqq \nu \leqq 1\}$ converges weakly to $\{y(\nu), 0 \leqq \nu \leqq 1\}$.

To complete the proof of (iii) notice that the transformation $x \to F(x)$ is a monotone continuous one-to-one map of $[0, \infty]$ onto [0, 1]. Also $\eta_n(x) = z_n(F(x))$ and $\eta(x) = y(F(x))$. Thus $\{\eta_n(x), 0 \leqq x \leqq \infty\}$ converges weakly to $\{\eta(x), 0 \leqq x \leqq \infty\}$. This completes the proof of Theorem 2.4.

Now we turn to the case of random perturbations. Throughout, we shall use an asterisk as a generic symbol on the α_{ni}'s, the empirical distribution functions, etc. to denote that the α_{ni}'s involved are random variables.

THEOREM 2.5. *Let* $\{\alpha_{ni}^*, i = 1, \cdots, n\}$, $n = 1, 2, \cdots$ *be random and let*

$$(2.18) \qquad n^{\frac{1}{2}} \max_{1\leqq i\leqq n} |\alpha_{ni}^* - \alpha_{ni}| = o_p(1)\,,$$

where the α_{ni}*'s are non-random and satisfy* (2.13). *Let* $F(x)$ *have a probability density function* $f(x)$ *satisfying*

$$(2.19) \qquad \sup_x |xf(x)| \leqq C < \infty\,.$$

Let F_n^* *be given by the right-hand side of* (2.2) *with the* α_{ni}*'s replaced by* α_{ni}^*, $i = 1, \cdots, n$. *Then* $\{\eta_n^*(x) = n^{\frac{1}{2}}(F_n^*(x) - G_n(x)), 0 \leqq x \leqq \infty\}$ *converges weakly to the Gaussian process* $\{\eta(x), 0 \leqq x \leqq \infty\}$ *defined in Theorem* 2.4.

PROOF. Given $\varepsilon_1, \varepsilon_2 > 0$, there exists an n_1 such that

$$P\{n^{\frac{1}{2}} \max_i |\alpha^*_{ni} - \alpha_{ni}| > \varepsilon_1\} < \varepsilon_2 \tag{2.20}$$

for all $n \geqq n_1$. Let $\alpha_{ni1} = \alpha_{ni} - \varepsilon_1/n^{\frac{1}{2}}$ and $\alpha_{ni2} = \alpha_{ni} + \varepsilon_1/n^{\frac{1}{2}}$, $i = 1, \cdots, n$. Let

$$\eta_{nr}(x) = \sum_{i=1}^n \{I(Z_i/\alpha_{nir}; x) - F(x\alpha_{nir})\}/n^{\frac{1}{2}},$$

$r = 1, 2$. From (2.20) and the monotonicity of $I(y; x)$ in y,

$$P\{\eta_{n1}(x) + \sum_{i=1}^n [F(x\alpha_{ni1}) - F(x\alpha_{ni})]/n^{\frac{1}{2}} \leqq \eta_n^*(x) \leqq \eta_{n2}(x) + \sum_{i=1}^n [F(x\alpha_{ni2}) - F(x\alpha_{ni})]/n^{\frac{1}{2}} \text{ for all } x\} \geqq 1 - \varepsilon_2$$

for $n \geqq n_1$. Using condition (2.19) it is easy to see that

$$\lim_n \sup_x |\sum_{i=1}^n [F(x\alpha_{nir}) - F(x\alpha_{ni})]/n^{\frac{1}{2}}| \leqq 2C\varepsilon_1$$

for $r = 1, 2$. Since $\varepsilon_1, \varepsilon_2$ are arbitrary, Theorem 2.5 now follows from Theorem 2.4.

The next theorem concerns itself with non-random perturbations and a random scale factor $\bar{Z}_n$.

THEOREM 2.6. *Let* $\{\alpha_{ni}, i = 1, \cdots, n\}$, $n = 1, 2, \cdots$ *be non-random and let* (2.13) *hold. Let*

(2.21) $xf(x)$ *be continuous and tend to* 0 *as* x *tends to* ∞,

$$x^\alpha(1 - F(x)) \to 0 \quad \text{as } x \to \infty \text{ for some } \alpha > 0. \tag{2.22}$$

Let

$$\xi_n = n^{\frac{1}{2}}(\bar{Z}_n - 1) = O_p(1). \tag{2.23}$$

Then

$$\sup_{0 \leqq x \leqq \infty} |\bar{\eta}_n(x) - \eta_n(x) - xf(x)\xi_n| = o_p(1), \tag{2.24}$$

where

$$\begin{aligned}\bar{\eta}_n(x) &= \sum_{i=1}^n [I(Z_i/(\alpha_{ni}\bar{Z}_n); x) - F(x\alpha_{ni})]/n^{\frac{1}{2}} \\ &= n^{\frac{1}{2}}(F_n(x\bar{Z}_n) - G_n(x)).\end{aligned} \tag{2.25}$$

PROOF. Now,

$$\begin{aligned}\bar{\eta}_n(x) &= n^{\frac{1}{2}}(F_n(x\bar{Z}_n) - G_n(x)) \\ &= n^{\frac{1}{2}}(F_n(x) - G_n(x)) + n^{\frac{1}{2}}(\bar{Z}_n - 1) \sum_{i=1}^n x\alpha_{ni} f(x\alpha_{ni})/n \\ &\quad + n^{\frac{1}{2}}\{F_n(x\bar{Z}_n) - F_n(x) - (\bar{Z}_n - 1) \sum_{i=1}^n x\alpha_{ni} f(x\alpha_{ni})/n\} \\ &= \eta_n(x) + \xi_n \sum_{i=1}^n x\alpha_{ni} f(x\alpha_{ni})/n + R_n(x) \quad \text{say.}\end{aligned} \tag{2.26}$$

Since $\alpha_{ni} \to 1$ uniformly in i

$$\sum_{i=1}^n x\alpha_{ni} f(x\alpha_{ni})/n \to xf(x) \tag{2.27}$$

uniformly in x as $n \to \infty$. Thus in order to establish (2.24) it is enough to prove that

$$\sup_{0 \leqq x \leqq \infty} |R_n(x)| = o_p(1). \tag{2.28}$$

The proof of (2.28) is completed in (2.44). The main steps are (2.30), (2.42) and (2.43). Let

$$R_n(x, c) = \sum_{i=1}^n \{I(Z_i/(c\alpha_{ni}); x) - I(Z_i/\alpha_{ni}; x) - (c-1)x\alpha_{ni}f(x\alpha_{ni})\}/n^{\frac{1}{2}} . \tag{2.29}$$

Then

$$R_n(x, \bar{Z}_n) = R_n(x) .$$

For any given $\phi > 0$ we can find an $L < \infty$ such that

$$P\{|\bar{Z}_n - 1| \geqq L/n^{\frac{1}{2}}\} \leqq \phi \qquad \text{for all} \quad n . \tag{2.30}$$

We now obtain a bound (see (2.42)) for

$$P\{\sup_{0\leqq x\leqq T}, \sup_{|c-1|\leqq L/n^{\frac{1}{2}}} |R_n(x, c)| > \omega\}$$

and $T > 0$. This is done by a standard but unavoidably lengthy method.

Let ε, $\theta > 0$. We shall choose these constants later in (2.37) and (2.38). The interval $[1 - L/n^{\frac{1}{2}}, 1 + L/n^{\frac{1}{2}}]$ is covered by $L_n = [2L/\varepsilon] + 1$ small intervals of length $\varepsilon_n = \varepsilon/n^{\frac{1}{2}}$ each, the rth interval being

$$L_r^* = \{c : c_r \leqq c \leqq c_{r+1}\} , \tag{2.31}$$

where $c_r = 1 - L/n^{\frac{1}{2}} + r\varepsilon_n$. Similarly, the interval $[0, T]$ is covered by $T_n = [n^{\frac{1}{2}}T/\theta] + 1$ intervals of length $\theta_n = \theta/n^{\frac{1}{2}}$ each, the sth interval being

$$T_s^* = \{x : x_s \leqq x \leqq x_{s+1}\} , \tag{2.32}$$

where $x_s = s\theta_n$. Let

$$m_{si} = \inf_{x\in T_s^*} x\alpha_{ni}f(x\alpha_{ni}) , \qquad M_{si} = \sup_{x\in T_s^*} x\alpha_{ni}f(x\alpha_{ni}) . \tag{2.33}$$

Fix an x. Since $\alpha_{ni} \to 1$ uniformly in i we may assume without loss of generality that $\frac{1}{2} \leqq \alpha_{ni} \leqq 2$ for all n and i. For $c \in L_r^*$,

$$\begin{aligned} R_n(x, c_r) - \varepsilon \textstyle\sum_{i=1}^n x\alpha_{ni}f(x\alpha_{ni})/n &\leqq R_n(x, c) \\ &\leqq R_n(x, c_{r+1}) + \varepsilon \textstyle\sum_{i=1}^n x\alpha_{ni}f(x\alpha_{ni})/n . \end{aligned}$$

Thus

$$\begin{aligned} R_n^*(x, L) &= \sup_{|c-1|\leqq L/n^{\frac{1}{2}}} |R_n(x, c)| \\ &\leqq \max_{0\leqq r\leqq L_n} |R_n(x, c_r)| + \varepsilon \textstyle\sum_{i=1}^n x\alpha_{ni}f(x\alpha_{ni})/n . \end{aligned} \tag{2.34}$$

Next, for $x \in T_s^*$,

$$\begin{aligned} &\textstyle\sum_{i=1}^n \{I(Z_i/(c_r\alpha_{ni}); x_s) - I(Z_i/\alpha_{ni}; x_{s+1}) - (c_r - 1)\lambda_{sir}\}/n^{\frac{1}{2}} \\ &\qquad \leqq R_n(x, c_r) \\ &\qquad \leqq \textstyle\sum_{i=1}^n \{I(Z_i/(c_r\alpha_{ni}); x_{s+1}) - I(Z_i/\alpha_{ni}; x_s) - (c_r - 1)\mu_{sir}\}/n^{\frac{1}{2}} , \end{aligned}$$

where

$$\begin{aligned} \lambda_{sir} &= m_{si} , && \text{if} \quad (c_r - 1) \leqq 0 , \\ &= M_{si} , && \text{if} \quad (c_r - 1) > 0 ; \\ \mu_{sir} &= M_{si} , && \text{if} \quad (c_r - 1) \leqq 0 , \\ &= m_{si} , && \text{if} \quad (c_r - 1) > 0 . \end{aligned}$$

Thus

$$\sup_{0\le x\le T} \sup_{|c-1|\le L/n^{\frac{1}{2}}} R_n(x, c) = \sup_{0\le x\le T} |R_n{}^*(x, L)| \le \max_{0\le s\le T_n} \max_{0\le r\le L_n} R_n{}^{**}(s, r) , \tag{2.35}$$

where

$$\begin{aligned} R_n{}^{**}(s, r) &= \max \{|\textstyle\sum_{i=1}^n \{I(Z_i/(c_r\alpha_{ni}); x_s) - I(Z_i/\alpha_{ni}; x_{s+1}) - (c_r - 1)\lambda_{sir}\}/n^{\frac{1}{2}}| , \\ &\qquad |\textstyle\sum_{i=1}^n \{I(Z_i/(c_r\alpha_{ni}); x_{s+1}) - I(Z_i/\alpha_{ni}; x_s) - (c_r - 1)\mu_{sir}\}/n^{\frac{1}{2}}|\} \\ &\quad + \varepsilon \textstyle\sum_{i=1}^n M_{si}/n . \end{aligned} \tag{2.36}$$

Consider the first term of the right-hand side of the above and write an upper bound for it as

$$\begin{aligned} &|\textstyle\sum_{i=1}^n \{I(Z_i/(c_r\alpha_{ni}); x_s) - I(Z_i/\alpha_{ni}; x_{s+1}) - F(c_r x_s \alpha_{ni}) + F(x_{s+1}\alpha_{ni})\}|/n^{\frac{1}{2}} \\ &\quad + |\textstyle\sum_{i=1}^n \{F(c_r x_s \alpha_{ni}) - F(x_{s+1}\alpha_{ni}) - (c_r - 1)\lambda_{sir}\}|/n^{\frac{1}{2}} + \varepsilon \sum_{i=1}^n M_{si}/n . \end{aligned}$$

Using condition (2.21) we see that for a given $\omega > 0$ we can choose ε, θ and n_2 such that

$$|\textstyle\sum_{i=1}^n \{F(c_r x_s \alpha_{ni}) - F(x_{s+1}\alpha_{ni}) - (c_r - 1)\lambda_{sir}\}|/n^{\frac{1}{2}} + \varepsilon \sum_{i=1}^n M_{si}/n \le \omega/2 \tag{2.37}$$

for all $n \ge n_2$. Treating the second term of the right-hand side of (2.36) in a similar fashion we can choose ε, θ and n_3 such that

$$R_n{}^{**}(s, r) \le \max_u |\textstyle\sum_{i=1}^n W_{niu}|/n^{\frac{1}{2}} + \omega/2 \tag{2.38}$$

for $n \ge n_3$, where

$$W_{niu} = I(Z_i/c_r\alpha_{ni}; x_u) - I(Z_i/\alpha_{ni}; x_{u+1}) - F(x_u c_r \alpha_{ni}) + F(x_{u+1}\alpha_{ni}) .$$

Let

$$p_{niu} = |F(x_u c_r \alpha_{ni}) - F(x_{u+1}\alpha_{ni})| \tag{2.39}$$

and B_{niu} be independent Bernoulli random variables such that $P(B_{niu} = 1) = p_{niu}$, $P(B_{niu} = 0) = 1 - p_{niu}$, $i = 1, \cdots, n$. Then the joint distribution of $\{W_{niu}\}$ is the same as that of $\{W^*_{niu}\}$ given by

$$\begin{aligned} W^*_{niu} &= B_{niu} - p_{niu} , && \text{if} \quad x_u c_r > x_{u+1} , \\ &= -B_{niu} + p_{niu} , && \text{if} \quad x_u c_r \le x_{u+1} . \end{aligned}$$

In either case, since the above two cases depend only on u and r and not on i,

$$\begin{aligned} &P\{|\textstyle\sum_i W_{niu}|/n^{\frac{1}{2}} > \omega/2\} \\ &\quad = P\{|\textstyle\sum_i (B_{niu} - p_{niu})| > n^{\frac{1}{2}}\omega/2\} \\ &\quad = P\{\textstyle\sum_i (B_{niu} - p_{niu}) > n^{\frac{1}{2}}\omega/2\} + P\{\sum_i (B_{niu} - p_{niu}) < -n^{\frac{1}{2}}\omega/2\} . \end{aligned}$$

For any $t > 0$,

$$\begin{aligned} P\{\textstyle\sum_i (B_{niu} - p_{niu}) > n^{\frac{1}{2}}\omega/2\} &\le \exp(-tn^{\frac{1}{2}}\omega/2) \textstyle\prod_{i=1}^n E\{\exp[t(B_{niu} - p_{niu})]\} \\ &= a(n, t, \omega) \quad \text{say.} \end{aligned}$$

Now, $E\{\exp[t(B_{niu} - p_{niu})]\} = (1 - p_{niu}(1 - e^t))\exp(-tp_{niu})$. From the definition of p_{niu} in (2.39), it follows that for each u and r, $\sup_n n^{\frac{1}{2}} p_{niu}$ is bounded by a constant independent of u and r and that $\sum_{i=1}^n p_{niu}/n^{\frac{1}{2}}$ converges to a constant which is less than C_1 (> 0). Writing out $(1/n)\log a(n, t, \omega)$ and expanding the moment generating function we see that

$$\begin{aligned}(1/n)\log a(n, t, \omega) &\leqq -\{t\omega/2 + C_1 t + C_1(1 - e^t) + O(1/n^{\frac{1}{2}})\}/n^{\frac{1}{2}} \\ &\leqq -C_2/n^{\frac{1}{2}} + O(1/n)\end{aligned}$$

for a suitable choice of t with $C_2 > 0$. For example, it is sufficient to choose $t = \log(1 + \omega/2C_1)$.
Thus

$$P\{\textstyle\sum_i (B_{niu} - p_{niu}) > n^{\frac{1}{2}}\omega/2\} \leqq \exp[-C_2 n^{\frac{1}{2}} + O(1)]\,, \tag{2.40}$$

for all u and r. We can obtain similarly that

$$P\{\textstyle\sum_i (B_{niu} - p_{niu}) < -n^{\frac{1}{2}}\omega/2\} \leqq \exp[-C_2 n^{\frac{1}{2}} + O(1)]\,. \tag{2.41}$$

Combining (2.35), (2.36), (2.37), (2.38), (2.40) and (2.41) we have

$$\begin{aligned}&P\{\sup_{0\leqq x\leqq T}|R_n(x)| > \omega,\ |\bar{Z}_n - 1| \leqq L/n^{\frac{1}{2}}\} \\ &\qquad \leqq P\{\sup_{0\leqq x\leqq T}\sup_{|c-1|\leqq L/n^{\frac{1}{2}}}|R_n(x, c)| > \omega\} \\ &\qquad \leqq 2([n^{\frac{1}{2}}T/\theta] + 1)([2L/\varepsilon] + 1)\exp[-C_2 n^{\frac{1}{2}} + O(1)]\end{aligned} \tag{2.42}$$

which can be made less than ϕ for $n \geqq n_4$.

Using the definition of $R_n(x)$ and the fact $\alpha_{ni} \geqq \frac{1}{2}$, we find that

$$\begin{aligned}&P\{\sup_{x\geqq T}|R_n(x)| > \omega,\ |\bar{Z}_n - 1| \leqq L/n^{\frac{1}{2}}\} \\ &\qquad \leqq P\{n^{\frac{1}{2}}(1 - F_n(T)) + L\sup_{x\geqq T/2} xf(x) > \omega\} \\ &\qquad = P\{n^{\frac{1}{2}}(G_n(T) - F_n(T)) > \omega - L\sup_{x\geqq T/2} xf(x) - n^{\frac{1}{2}}(1 - G_n(T))\} \\ &\qquad \leqq P\{n^{\frac{1}{2}}(G_n(T) - F_n(T)) > \omega - L\sup_{x\geqq T/2} xf(x) - n^{\frac{1}{2}}(1 - F(T/2))\}\,.\end{aligned}$$

Letting $T = 2n^{1/(2\alpha)}$, we can find n_5 such that

$$n^{\frac{1}{2}}(1 - F(T/2)) \leqq \omega/4 \qquad \text{and} \qquad L\sup_{x\geqq T/2} xf(x) \leqq \omega/4$$

for $n \geqq n_5$ in view of (2.21) and (2.22). Since $n^{\frac{1}{2}}(F_n(T) - G_n(T))$ is asymptotically normal with mean 0 and variance $F(T)(1 - F(T))$, uniformly in T,

$$\begin{aligned}&P\{\sup_{x\geqq T}|R_n(x)| > \omega,\ |\bar{Z}_n - 1| \leqq L/n^{\frac{1}{2}}\} \\ &\qquad \leqq P\{n^{\frac{1}{2}}(G_n(T) - F_n(T)) > \omega/2\} \\ &\qquad \leqq 1 - \Phi\{\omega/2[F(T)(1 - F(T))]^{\frac{1}{2}}\} + \phi \\ &\qquad \leqq 2\phi\end{aligned} \tag{2.43}$$

for $n \geqq n_6$. Combining (2.30), (2.42) and (2.43) we have

$$P\{\sup_{0\leqq x\leqq\infty}|R_n(x)| > \omega\} \leqq 4\phi \tag{2.44}$$

for large n. This completes the proof of Theorem 2.6.

COROLLARY 2.7. *Let the conditions of Theorem* 2.6 *be satisfied. For every finite collection* $\{x_1, \cdots, x_k\}$ *let the joint distribution of* $\{\eta_n(x_1), \cdots, \eta_n(x_k), \xi_n\}$ *converge to the distribution of* $\{\eta(x_1), \cdots, \eta(x_k), \xi\}$ *which is multivariate normal and where* $E(\xi) = 0$, $V(\xi) = 1$ *and* $\text{Cov}(\eta(x), \xi) = a(x)$. *Then the sequence of processes* $\{\bar{\eta}_n(x), 0 \leqq x \leqq \infty\}$ *converges weakly to a Gaussian process* $\{\bar{\eta}(x), 0 \leqq x \leqq \infty\}$ *with mean function zero and covariance kernel*

$$\bar{K}(x, y) = F(x)(1 - F(y)) + xyf(x)f(y) + xf(x)a(y) + yf(y)a(x) \quad \textit{for} \quad x \leqq y . \tag{2.45}$$

THEOREM 2.8. *Let* $\{\alpha_{ni}^*, i = 1, \cdots, n\}$, $n = 1, 2, \cdots$ *be random and satisfy* (2.18). *Let* (2.21), (2.22) *and* (2.23) *hold. Then*

$$\sup_{0 \leqq x \leqq \infty} |\bar{\eta}_n{}^*(x) - \eta_n{}^*(x) - xf(x)\xi_n| = o_p(1) \tag{2.46}$$

where $\bar{\eta}_n{}^*(x) = n^{\frac{1}{2}}(F_n{}^*(x\bar{Z}_n) - G_n(x))$.

PROOF. This theorem follows from Theorem 2.6 in the same way that Theorem 2.5 follows from Theorem 2.4.

COROLLARY 2.9. *Under the condition on the convergence of the distributions of* $(\eta_n(x_1), \cdots, \eta_n(x_k), \xi_n)$ *stated in Corollary* 2.7 *and under the conditions of Theorem* 2.8, *the sequence of stochastic processes* $\{\bar{\eta}_n{}^*(x), 0 \leqq x \leqq \infty\}$ *converges weakly to the Gaussian process* $\{\bar{\eta}(x), 0 \leqq x \leqq \infty\}$ *defined in Corollary* 2.7.

3. Application to tests based on spacings. The theorems of the earlier section were motivated by the following important application to goodness-of-fit tests based on spacings and the derivation of their Pitman's asymptotic relative efficiencies (ARE's) which will be considered now.

For computing the ARE's, it is enough to obtain the limiting distributions under a sequence of alternatives which converge to the hypothesis. Hence we will specify the alternative hypothesis by a distribution function $A_n(x)$ depending on n converging to the uniform distribution, which corresponds to the null hypothesis. Under the alternative hypothesis, we specify the distribution function to be given by

$$A_n(x) = x + L_n(x)/n^\delta , \qquad 0 \leqq x \leqq 1 \tag{3.1}$$

where $L_n(0) = L_n(1) = 0$ and $\delta \geqq \frac{1}{4}$. We further assume that $L_n(x)$ is twice differentiable on [0, 1] and there is a function $L(x)$ which is twice continuously differentiable and such that

$$L(0) = L(1) = 0 , \qquad n^{\delta^*} \sup_{0 \leqq x \leqq 1} |L_n''(x) - l'(x)| = o(1) \tag{3.2}$$

where $l(x)$ and $l'(x)$ are the first and second derivatives of $L(x)$ and $\delta^* = \max(0, \frac{1}{2} - \delta)$. This sequence of alternatives is smooth in a certain sense and has been considered before. For instance, see Cibisov (1961). Notice that for such smooth alternatives the following also hold:

$$n^{\delta^*} \sup_{0 \leqq x \leqq 1} |L_n(x) - L(x)| = o(1) , \qquad n^{\delta^*} \sup_{0 \leqq x \leqq 1} |L_n'(x) - l(x)| = o(1) . \tag{3.3}$$

Let $X_1, \cdots, X_{n-1}$ be $(n-1)$ independent random variables from $A_n(x)$, $n = 2, 3, \cdots$ with the corresponding ordered observations $X_1' \leqq \cdots \leqq X'_{n-1}$. We should add a suffix n to these observations to show their dependence on n, but we shall not do so to simplify the notation. Let $a_n(p)$ and $A_n^{-1}(p)$ denote the density and the inverse function corresponding to $A_n(p)$. Let

$$k_n(p) = a_n(A_n^{-1}(p)) = \left[\frac{dA_n^{-1}(p)}{dp}\right]^{-1}. \tag{3.4}$$

In view of (3.1), (3.2) and (3.3), it can be verified that

$$A_n^{-1}(p) = p - L(p)/n^\delta + o(1/n^\delta) \tag{3.5}$$

$$k_n(p) = 1 + l(p)/n^\delta - L(p)l'(p)/n^{2\delta} + o(1/n^{\delta+\delta^*}) \tag{3.6}$$

where $o(\cdot)$ is uniform in p. Our aim now is to relate the sample spacings from $A_n(x)$ to randomly perturbed and randomly scaled exponential random variables via the spacings from a uniform distribution on [0, 1].

Let $U_1, \cdots, U_{n-1}$ be $(n-1)$ independently and identically distributed random variables with uniform distribution on [0, 1]. These are then arranged in increasing order as $U_1' \leqq \cdots \leqq U'_{n-1}$ and the uniform spacings are defined as

$$T_i = U_i' - U'_{i-1}, \qquad i = 1, \cdots, n, \tag{3.7}$$

where again we put $U_0' = 0$, $U_n' = 1$.

For two random variables X and Y, we write $X \sim Y$ to mean that X and Y are distributionally equivalent, that is, the distributions of X and Y are identical. We then have

$$\{X_i', i = 0, \cdots, n\} \sim \{A_n^{-1}(U_i'), i = 0, \cdots, n\},$$

and thus

$$\begin{aligned}\{D_i, i = 1, \cdots, n\} &\sim \{A_n^{-1}(U_i') - A_n^{-1}(U'_{i-1}), i = 1, \cdots, n\}\\ &= \{T_i/k_n(\tilde{U}_i), i = 1, \cdots, n, \text{ where } U'_{i-1} \leqq \tilde{U}_i \leqq U_i'\}\\ &= \{T_i/\alpha^*_{ni}, i = 1, \cdots, n\},\end{aligned} \tag{3.8}$$

where

$$\alpha^*_{ni} = 1 + l(\tilde{U}_i)/n^\delta - L(\tilde{U}_i)l'(\tilde{U}_i)/n^{2\delta} + R_{ni}, \tag{3.9}$$

with

$$\sup_i n^{\frac{1}{2}}|R_{ni}| \to 0$$

almost everywhere in view of (3.6). Also from the existence of the limiting distribution of the Kolmogorov-Smirnov statistic, we have

$$\sup_i n^{\frac{1}{2}}|U_i' - i/n| = O_p(1),$$

so that

$$\begin{aligned}\sup_i n^{\delta^*}|l(\tilde{U}_i) - l(i/n)| &= o_p(1),\\ \sup_i |L(\tilde{U}_i)l'(\tilde{U}_i) - L(i/n)l'(i/n)| &= o_p(1).\end{aligned} \tag{3.10}$$

Let the non-random α_{ni} corresponding to α^*_{ni} of (3.9) be defined as

$$\alpha_{ni} = 1 + l(i/n)/n^{\delta} - L(i/n)l'(i/n)/n^{2\delta} . \tag{3.11}$$

Then condition (2.18) of Theorem 2.5 is satisfied, as well as condition (2.13) of Theorem 2.4.

Now let $Z_1, Z_2, \cdots$ be independent identically distributed exponential random variables with density e^{-z}, $z \geqq 0$. Let $Z_n^* = (Z_1 + \cdots + Z_n)$ and $\bar{Z}_n = Z_n^*/n$. Then it is well known that

$$\{T_i, i = 1, \cdots, n\} \sim \{Z_i/Z_n^*, i = 1, \cdots, n\} .$$

Thus (3.8) may be rewritten as

$$\{D_i, i = 1, \cdots, n\} \sim \{Z_i/\alpha^*_{ni} Z_n^*, i = 1, \cdots, n\} . \tag{3.12}$$

Under uniform distribution, $E(D_i) = 1/n$ for all i. We shall therefore call $\{nD_i, i = 1, \cdots, n\}$ 'normalized' spacings. From (3.12), we have the following distributional equivalence between the normalized spacings and exponential random variables

$$\{nD_i, i = 1, \cdots, n\} \sim \{Z_i/\alpha^*_{ni} \bar{Z}_n, i = 1, \cdots, n\} . \tag{3.13}$$

The empirical distribution function $H_n(x)$ of the normalized spacings is of central interest in this context and is defined as

$$H_n(x) = \textstyle\sum_{i=1}^n I(nD_i; x)/n , \qquad x \geqq 0 . \tag{3.14}$$

Using the equivalence (3.13), we note that

$$\begin{aligned} \{H_n(x), x \geqq 0\} &\sim \{\textstyle\sum_{i=1}^n I(Z_i/\alpha^*_{ni} \bar{Z}_n; x)/n, x \geqq 0\} \\ &= \{\bar{F}_n^*(x), x \geqq 0\} \end{aligned} \tag{3.15}$$

where $\bar{F}_n^*(\cdot)$ is as defined in (2.3) with random perturbations. In our terminology $\bar{F}_n^*(\cdot)$ denotes the empirical distribution function of the exponential random variables $Z_1, Z_2, \cdots$ perturbed by the random factors α^*_{ni} and randomly scaled by $\bar{Z}_n$. The equivalence in (3.15) says that the distributions of the stochastic processes $\{H_n(x), x \geqq 0\}$ and $\{\bar{F}_n^*(x), x \geqq 0\}$ coincide in $D[0, \infty]$.

Thus the problem of finding the asymptotic distribution of $\{H_n(x), x \geqq 0\}$ is reduced to that of $\{\bar{F}_n^*(x), x \geqq 0\}$, which has been dealt with in Section 2. It is easily checked that all the required conditions on the perturbation factors α^*_{ni}, on the distribution function $F(x)$ (which is the exponential distribution here), as well as the condition (2.23) on $\xi_n = n^{\frac{1}{2}}(\bar{Z}_n - 1)$ are satisfied. Further, when the non-random α_{ni} have the structure given in (3.11), using a second order Taylor expansion, integrating by parts to evaluate $\int L(p)l'(p)\, dp$ and the fact that $\int l(p)\, dp = 0$, $G_n(x)$ of Section 2 can be expressed in the form

$$\begin{aligned} G_n(x) &= (1 - e^{-x}) , && \text{if } \delta > \tfrac{1}{4} \\ &= (1 - e^{-x}) + (\textstyle\int_0^1 l^2(p)\, dp) e^{-x}(x - x^2/2)/n^{\frac{1}{2}} , && \text{if } \delta = \tfrac{1}{4} , \end{aligned} \tag{3.16}$$

ignoring terms which are of smaller order than $n^{-\frac{1}{2}}$ uniformly in x. Thus from Corollary 2.7 we have the following useful theorem.

THEOREM 3.1. *Under the alternatives* (3.1), *the sequence of stochastic processes* $\{\zeta_n(x) = n^{\frac{1}{2}}(H_n(x) - G_n(x)), x \geqq 0\}$ *converges weakly to the Gaussian process* $\{\zeta(x), x \geqq 0\}$ *in* $D[0, \infty]$ *with mean function zero and covariance kernel*

$$K(x, y) = e^{-y}(1 - e^{-x} - xye^{-x}), \qquad 0 \leqq x \leqq y \leqq \infty . \tag{3.17}$$

This theorem on the empirical distribution function of the normalized spacings forms the basic result for deriving the asymptotic distributions of test statistics based symmetrically on spacings. From the invariance principle, we have

THEOREM 3.2. *Let* $g(\cdot)$ *be a real-valued measurable function on* $D[0, \infty]$ *which is* a.e. *continuous with respect to the probability measure induced by the Gaussian process* $\{\zeta(x), x \geqq 0\}$ *of the previous theorem. Then the distribution of the real-valued random variable* $g(\zeta_n)$ *converges weakly to the distribution of* $g(\zeta)$ *as* $n \to \infty$.

Using these results, the limiting distributions of the symmetric spacings test statistics

$$\begin{aligned} V_r(n) &= \textstyle\sum_{i=1}^{n} (nD_i)^r/n\,, \quad r > -\tfrac{1}{2}\,, \\ U(n) &= \textstyle\sum_{i=1}^{n} |nD_i - 1|/n\,, \\ L(n) &= \textstyle\sum_{i=1}^{n} \log (nD_i)/n \end{aligned} \tag{3.18}$$

have been obtained under the alternatives (3.1) and their ARE's compared in Sethuraman and Rao (1970).

One interesting fact that emerges from this analysis is that the symmetric spacings tests cannot discriminate alternatives $A_n(x)$ if $\delta > \frac{1}{4}$ so that comparison of the ARE's must be made for a sequence of alternatives converging to the hypothesis at the rate of $n^{-\frac{1}{4}}$. It has also been demonstrated in Sethuraman and Rao (1970) that among a wide class of such tests, the test statistic

$$V_2(n) = \textstyle\sum_{i=1}^{n} (nD_i)^2/n$$

due to Greenwood (1946) has the maximum efficiency.

Some further results on the weak convergence of empirical distribution function of the so called 'modified spacings' (see Sethuraman and Rao (1970)) have also been obtained by the authors and will appear in a future paper.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[2] CHENTSOV, N. N. (1956). Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the heuristic approach to the Kolmogorov-Smirnov tests. *Theor. Probability Appl.* **1** 140-144.

[3] CIBISOV, D. M. (1961). On the tests of fit based on sample spacings. *Theor. Probability Appl.* **6** 325-329.

[4] DONSKER, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277-281.

[5] DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistic.* **20** 393–403.

[6] FISZ, M. (1963). *Probability Theory and Mathematical Statistics*, (3rd ed.). Wiley, New York.

[7] GREENWOOD, MAJOR (1946). The statistical study of infectious diseases. *J. Roy. Statist. Soc.* **109** 85–103.

[8] PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B* **27** 395–449.

[9] SETHURAMAN, J. (1965). Limit theorems for stochastic processes. Technical Report No. 10, Stanford Univ.

[10] SETHURAMAN, J. and RAO, J. S. (1970). Pitman efficiencies of tests based on spacings, *Nonparametric Techniques in Statistical Inference*, (M. L. Puri ed.). Cambridge Univ. Press. 405–416.

[11] SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Prob. Appl.* **1** 261–290.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47401

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48103