Probability and Statistical Decision Theory

Volume A

Edited by

F. Konecny,

J. Mogyoródi, and

W. Wertz

PROBABILITY AND STATISTICAL DECISION THEORY

Proceedings of the 4th Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf, Austria, 4–10 September, 1983

Volume A

Edited by

F. KONECNY

University of Agriculture, Vienna, Austria

J. MOGYORÓDI

Eötvös Loránd University, Budapest, Hungary

W. WERTZ

Technical University of Vienna, Austria

D. REIDEL PUBLISHING COMPANY

A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP

DORDRECHT / BOSTON / LANCASTER / TOKYO

Library of Congress Cataloging-in-Publication Data

CIP

Pannonian Symposium on Mathematical Statistics (4th:
1983: Bad Tatzmannsdorf, Austria)
Probability and statistical decision theory.

(Proceedings of the 4th Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf, Austria, 4–10 September 1983; v. A)
Bibliography: p.
Includes index.
1. Probabilities—Congresses. 2. Statistical decision—Congresses. I. Mogyoródi, J. II. Konecny, F. III. Wertz, Wolfgang. IV. Series: Pannonian Symposium on Mathematical Statistics (4th: 1983: Bad Tatzmannsdorf, Austria). Proceedings of the 4th Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf, Austria, 4–10 September 1983; v. A.
QA276.A1P36 1983 vol. A 519.5 s [519.2] 85-14465 [QA273.A1]
ISBN 90-277-2089-4

Distributors for the U.S.A. and Canada
Kluwer Academic Publishers,
190 Old Derby Street, Hingham, MA 02043, U.S.A.

Distributors for Albania, Bulgaria, Chinese People's Republic,
Cuba, Czechoslovakia, German Democratic Republic, Hungary,
Korean People's Republic, Mongolia, Poland, Romania,
the U.S.S.R., Vietnam and Yugoslavia
Kultura Hungarian Foreign Trading Company
P.O.B. 149
H-1389 Budapest 62, Hungary

Distributors for all remaining countries
Kluwer Academic Publishers Group,
P.O.Box 322, 3300 AH Dordrecht, Holland

Joint edition published by D. Reidel Publishing Company
Dordrecht, Holland
and
Akadémiai Kiadó, Budapest, Hungary

Printed in Hungary

PREFACE

The Fourth Pannonian Symposium on Mathematical Statistics was held in Bad Tatzmannsdorf, Austria, 4-10 September, 1983. The first two Symposia were held there in 1979 and 1981, whereas the Third Symposium was staged in Visegrád, Hungary in 1982. The proceedings volumes of these conferences, published by Springer, D. Reidel, and D. Reidel & Akadémiai Kiadó, respectively, give information about the objective of the Pannonian Symposia and the topics covered.

About 130 participants from 17 countries took part in this Fourth Symposium, and 92 lectures were presented. This volume contains 24 reviewed contributions which are mainly mathematically oriented. A second group of papers dealing with problems of applied statistics, probability theory and related topics is published in a separate volume entitled "Mathematical Statistics and Applications".

The contributions dealing with probability theory concentrate on two (intersecting) main topics: on the one hand, stochastic processes, and, on the other hand, limit theorems and invariance principles: Gaussian Processes, approximation of the Wiener Process, distribution of spacings and of order statistics, limit theorems in triangular arrays. Besides, adjacent topics like erdogic theory, maximal inequalities, approximation of convolutions of distribution functions, stochastic programming etc. are dealt with. In many of these contributions stress is laid upon the weakening of the assumption of independence.

The subjects of the statistical contributions are even more homogeneous: there are decision-theoretic papers (treating admissibility, sufficiency, limits of experiments, etc.) and papers on nonparametric estimation (nonparametric estimators of regression curves, further rank-, L- and shrinkage estimators). One paper gives a comprehensive survey of two-sided parametric tests.

The reader will observe at once the close connections between many of the questions under consideration; in particular, many results of papers which can be related to probability theory have immediate applications in statistics, and some belong to the common boundary of these subjects.

We wish to express our thanks to many persons who gave us indispensable aid in the edition of this volume: S. Csörgő, P. Deheuvels, M. Denker, L. Devroye, W. Eberl jr., B. Gyires, L. Györfi, J. Hurt, M. Hušková, I. Kátai, A. Kozek, W. Philipp, D. Plachky, P. Révész, P.K. Sen, F. Schipp and W. Sendler for their help in refereeing the papers and in other editorial matters, the secretarial staff of Professor I. Kátai at the University of Budapest, who did the laborious typing of the manuscript under the supervision of Mrs. Z. Andrásné Králik and, last but not least, Akadémiai Kiadó and D. Reidel Publishers for their good cooperation. (After the refereeing process and retyping, all papers were returned to the authors for correction.)

The organization of the Symposium was made possible by the help of many individuals and institutions. The organizers gratefully acknowledge the generous support given by the State Government of Burgenland, the Federal Ministry of Science and Research, the Austrian Statistical Society, the Control Data Co., Hypobank of Burgenland Co., the Volksbank Oberwart Co., the Raiffeisenbank Oberwart Co., the Kurbad Tatzmannsdorf Co. and the Local Autority of Bad Tatzmannsdorf; special thanks go to Th. Kery and Dr. R. Grohotolsky (Head and Vice-Head of Burgenland resp.), Dr. J. Karall (Member of the State Government of Burgenland). DDr. Schranz (Member of State Parliament of Burgenland) and Mag. R. Luipersbeck (Director of the Kurbad Tatzmannsdorf AG) for their help in many respects. Finally, we express

our thanks to the ladies for their splendid work in the preparation and local organization of the meeting, in particular to Ms. Ingrid Danzinger who patiently undertook most of the typing for the organization.

Wolfgang WE[illegible]

CONTENTS

Proc. of the 4th Pannonian Symp.
on Math. Stat., Bad Tatzmannsdorf, Austria 1983
F. Konecny, J. Mogyoródi, W. Wertz, eds

SPACINGS AND APPLICATIONS

PAUL DEHEUVELS

Paris

Abstract

We prove limiting theorems for the order statistics of k-spacings and survey some important results obtained in the theory of spacings with statistical applications.

0. Introduction

The aim of this paper is to survey some important aspects of the theory of spacings and its applications. We shall prove also some original results concerning the limiting properties of maximal k-spacings.

The paper is organized as follows. In section 1, we give the main definitions concerning uniform k-spacings and discuss the limiting distributions of their upper order statistics. We also state the main theorems which have been obtained concerning strong bounds. In section 2, we discuss exact and approximate finite distributions, and show their use to obtain upper and lower strong bounds for k-spacings when k is fixed. In section 3 we precise the results of section 2 by different methods. In section 4 we discuss minimal k-spacings in terms of exact and asymptotic distributions, and give strong upper and lower asymptotic bounds. We also discuss questions related to the scan statistic. In section 5, we review the main applications of spacings in statistical inference and probability theory. In section 6 we mention some extensions of the theory for general distributions and in higher dimensions. Section 7 gives a conclusion to our discussion.

Because of the great amount of material, we have made a selective choice of the topics we discuss, deliberately overlooking some important results we could not mention. It follows that this paper could hardly pretend to cover the subject.

1. Uniform spacings (I)

Let $U_1, U_2, \ldots$ be a sequence of independent and uniformly distributed random variables in $(0,1)$. Let

$$U_{0,n} = 0 < U_{1,n} < \ldots < U_{n,n} < U_{n+1,n} = 1$$

denote the order statistics of $U_1, \ldots, U_n$.

The uniform spacings of order n are defined by

$$S_{i,1}^{(n)} = U_{i,n} - U_{i-1,n}, \quad i=1,\ldots,n+1.$$

Likewise, for $1 \le k \le n+1$, the uniform k-spacings of order n are defined by

$$S_{i,k}^{(n)} = U_{i,n} - U_{i-k,n}, \quad i=k,\ldots,n+1.$$

For $1 \le k \le n+1$, let

$$M_{n-k+2,k}^{(n)} < \ldots < M_{2,k}^{(n)} < M_{1,k}^{(n)}$$

denote the order statistics of $S_{i,k}^{(n)}$, $i=k,\ldots,n+1$.

A great number of results have been proved recently concerning the limiting behavior of $\{M_{j,k}^{(n)}, 1 \le j \le n-k+2\}$. We shall make here a brief survey of some important facts concerning these statistics.

Lemma 1.

If $Y_1, \ldots, Y_{n+1}$ are independent exponentially distributed random variables, and if $T_{n+1} = \sum_{i=1}^{n+1} Y_i$, then $\{S_{i,1}^{(n)}, 1 \le i \le n+1\}$ is distributed as $\{Y_i/T_{n+1}, 1 \le i \le n+1\}$.

Proof. See Pyke (1965), Moran (1947).

Let $Y_{1,n+1} < Y_{2,n+1} < \ldots < Y_{n+1,n+1}$ be the order statistics of $Y_1, \ldots, Y_{n+1}$. It is well known (see f.i. Galambos (1978)) that (with $P(Y_1>u) = e^{-u}$, $u>0$)

$$\lim_{n\to\infty} P(Y_{n+1,n+1} - \text{Log } n < u) = e^{-e^{-u}}, \quad -\infty<u<+\infty,$$

and likewise, for any fixed $j>1$,

$$\lim_{n\to\infty} P(Y_{n-j+1,n+1} - \text{Log } n<u) = e^{-e^{-u}} \left\{\sum_{\ell=0}^{j-1} \frac{e^{-\ell u}}{\ell!}\right\}, \quad -\infty<u<+\infty.$$

By the weak law of large numbers, we have evidently

$$\lim_{n\to\infty} n^{-1} T_{n+1} = 1 \quad \text{in probability.}$$

It follows that·

Proposition 1.

For any fixed $j\ge1$, we have

$$\lim_{n\to} P(nM_{j,1}^{(n)} - \text{Log } n<u) = e^{-e^{-u}} \left\{\sum_{\ell=0}^{j-1} \frac{e^{-\ell u}}{\ell!}\right\}.$$

Note that the limiting distribution of $M_{1,1}^{(n)}$ has been originally obtained by Levy (1939) and Stevens (1939) (see also Darling (1953) and Le Cam (1958)).

If we consider $M_{j,k}^{(n)}$ instead of $M_{j,1}^{(n)}$, similar results can be obtained, noting that by Lemma 1, $\{S_{i,k}^{(n)}, k<n<n+1\}$ is dis-

tributed as $\{Z_{i,k}/T_{n+1},\ 1\le i\le n-k+1\}$, where

$$Z_{i,k} = \sum_{\ell=i}^{i+k-1} Y_\ell,\quad 1\le i\le n-k+1,$$

follows a $\Gamma(k,1)$ Gamma distribution:

$$P(Z_{i,k} > u) = e^{-u}\{\sum_{\ell=0}^{k-1} \frac{u^\ell}{\ell!}\} .$$

For a fixed $k\ge 1$, it can be seen easily that $\{Z_{i,k}, i\ge 1\}$ defines a k-dependent stationary sequence of random variables. We need here the following lemma due to Watson (1954) (see Galambos (1978) p. 162).

Lemma 2.

Let Z_1, Z_2, ... be a k-dependent stationary sequence with common distribution $F(x)$. Assume that $a_n>0$ and b_n are sequences of numbers such that, for any x,

$$\lim_{n\to\infty} n(1-F(a_n x + b_n)) = u(x),$$

where $0<u(x)<\infty$ on an interval of positive length. Assume further that, for any $1<i\le k$,

$$\lim_{u\uparrow\omega} \frac{P(Z_1>u, Z_i>u)}{1 - F(u)} = 0,$$

where $\omega = \sup\{x, F(x)<1\}$.

Then

$$\lim_{n\to\infty} P(\max_{1\le i\le n} Z_i < a_n x + b_n) = e^{-u(x)},\quad -\infty<x<+\infty.$$

In the case where $Z_i=Z_{i,k}$, we have:

$$1 - F(u) = e^{-u}\{\sum_{\ell=0}^{k-1} \frac{u^\ell}{\ell!}\} \sim \frac{u^{k-1}e^{-u}}{(k-1)!} \quad \text{as} \quad u\uparrow\infty$$

and

$$P(Z_1 > u, Z_i > u) = P(U + V > u, V + W > u),$$

where U, V and W are respectively Gamma $\Gamma(i-1,1)$, $\Gamma(k-i+1,1)$ and $\Gamma(i-1,1)$, independent random variables. Put $r=i-1$, $s=k-i+1$. We have:

$$P(Z_1>u, Z_i>u) = P(V>u) + \int_0^u (P(U>u-v))^2 \frac{v^{s-1}e^{-v}}{(s-1)!} dv,$$

and

$$P(U>t) = e^{-t}\{\sum_{\ell=0}^{r-1} \frac{t^\ell}{\ell!}\}, \quad P(V>u) = e^{-u}\{\sum_{\ell=0}^{s-1} \frac{t^\ell}{\ell!}\},$$

while

$$1 - F(u) = P(U+V>u) = P(V>u) + \int_0^u P(U>u-v) \frac{v^{s-1}e^{-v}}{(s-1)!} dv .$$

Since evidently as $u\to\infty$, $P(V>u)/(1-F(u))\to 0$, if we use the bound

$$\int_0^u (P(U>u-v))^2 \frac{v^{s-1}e^{-v}}{(s-1)!} dv \le P(U>a) \int_0^{u-a} P(U>u-v)\frac{v^{s-1}e^{-v}}{s-1\ !} dv + $$
$$+ P(V>u-a),$$

where $0\le a\le 1$, it suffices to take $a=\mathrm{LogLog}\ u=\mathrm{Log}_2 u$ to prove that

$$\lim_{u\uparrow\infty} \frac{P(Z_1>u, Z_i>u)}{1 - F(u)} = 0 .$$

A direct application of Lemma 2 gives

$$\lim_{n\to\infty} P(\max_{1\le i\le n-k+1} Z_{i,k} - \mathrm{Log}\ n-(k-1)\mathrm{Log}_2 n + \mathrm{Log}(k-1)!<u)=e^{-e^{-u}},$$

$-\infty<u<\infty$.

It follows from this and from Lemma 1 that:

Proposition 2.

For any fixed $j\geq 1$ and $k\geq 1$, we have, for $-\infty<u<+\infty$,

$$\lim_{n\to\infty} P(nM_{j,k}^{(n)} - \mathrm{Log}\, n - (k-1)\mathrm{Log}_2 n + \mathrm{Log}(k-1)! < u) = e^{-e^{-u}} \{\sum_{\ell=0}^{j-1} \frac{e^{-\ell u}}{\ell !}\}$$

Proof. Strictly speaking, we have given a complete proof for j=1. It happens that (see Galambos (1978) p. 203, Exercise 14) one can extend without difficulty Lemma 2 to the case of the j-th maximum.

Proposition 2 has been proved for j=1 and $k\geq 1$ by Holst (1980), who mentions without details a possible extension of his results for $j\geq 1$. A simple consequence of Proposition 2 is as follows.

Proposition 3.

For any fixed $j\geq 1$ and $k\geq 1$, we have, as $n\to\infty$,

$$\frac{nM_{j,k}^{(n)} - \mathrm{Log}\, n}{\mathrm{Log}_2 n} \to k-1 \quad \text{in probability.}$$

In a series of papers (Devroye (1981, 1982), Deheuvels (1982, 1983)), the upper and lower strong class of $M_{j,1}^{(n)}$ have been investigated. The following results have been proved to be true (in the sequel, Log_j is the j times iterated logarithm):

Proposition 4.

For any fixed $j\geq 1$, for any $p\geq 4$, we have

$$P(nM_{j,1}^{(n)} > \mathrm{Log}\, n + \frac{1}{j}(2\mathrm{Log}_2 n + \mathrm{Log}_3 n + \ldots + \mathrm{Log}_{p-1} n + (1+\varepsilon)\mathrm{Log}_p n) \text{ i.o.}) =$$

$$P(nM_{j,1}^{(n)} < \mathrm{Log}\, n - \mathrm{Log}_3 n - \mathrm{Log}\, 2 - \varepsilon \text{ i.o.}) = 0 \text{ or } 1,$$

according as $\varepsilon>0$ or $\varepsilon<0$.

Recently, Deheuvels and Devroye (1983) have proved in the case $k \geq 1$:

<u>Proposition 5</u>.

If $k \geq 1$ is constant, then, almost surely,

$$\limsup_{n\to\infty} \frac{nM_{1,k}^{(n)} - \text{Log } n - (k-1)\text{ Log}_2 n}{\text{Log}_2 n} = 2,$$

and

$$\liminf_{n\to\infty} \frac{nM_{1,k}^{(n)} - \text{Log } n - (k-1)\text{ Log}_2 n}{\text{Log}_2 n} = 0$$

If $k=k_n \geq 1$ is nondecreasing, and if $k_n = o(\text{Log}_2 n)$ as $n\to\infty$, then

$$\lim_{n\to\infty} \frac{nM_{1,k}^{(n)} - \text{Log } n}{(k-1)\text{Log}\left(\frac{e \text{ Log } n}{k}\right)} = 1 \quad \text{a.s..}$$

Proposition 5 settles the case of k-spacings when $k=k_n=o(\text{Log } n)$. For large k's, the following result holds, due to Mason (1983):

<u>Proposition 5</u>.

Let for $1 \leq d \leq n$

$$\delta_n(d) = \max_{1\leq k\leq d} |M_{1,k}^{(n)} - \frac{k}{n}|.$$

Then

1°) If $k=k_n \geq 1$ is nondecreasing and such that $k_n=o(\text{Log } n)$, $n\to\infty$, we have

$$\lim_{n\to\infty} n\delta_n(k)/\text{Log } n = 1 \text{ a.s..}$$

2°) If $k=k_n \geq 1$ is nondecreasing and such that $k_n/\text{Log } n \to c \in (0,+\infty)$, we have

$$\lim_{n\to\infty} n\delta_n(k)/\text{Log } n = c(\alpha^+ - 1) \text{ a.s.,}$$

where α^+ is the unique root $\lambda>1$ of $\lambda+\mathrm{Log}(1/\lambda) - 1 = 1/c$.

3°) If $k=k_n\geq 1$ is nondecreasing and satisfies the so-called Csörgő-Révész-Stute conditions, i.e.

(i) $k_n/\mathrm{Log}\ n \to \infty$, $k_n/n \to 0$,

(ii) $\mathrm{Log}(n/k_n)/\mathrm{LogLog}\ n \to \infty$

we have

$$\lim_{n\to\infty} n\delta_n(k)/\{2k_n\mathrm{Log}(n/k_n)\}^{1/2} = 1 \quad \text{a.s..}$$

Proposition 5′ has obvious applications for the evaluation of the oscillation of the uniform empirical quantile process.

2. Uniform spacings. (II)

If we consider the order statistics of the uniform 1-spacings of order n:

$$M^{(n)}_{n+1,1} < \ldots < M^{(n)}_{2,1} < M^{(n)}_{1,1} ,$$

the exact distribution of $M^{(n)}_{j,1}$ is well known and given by:

Lemma 3.

For any $1\leq j\leq n+1$, we have

$$P(M^{(n)}_{j,1}<x) = \sum_{h=0}^{j-1} \binom{n+1}{h} \sum_{\ell=0}^{n+1-h} (-1)^{\ell} \binom{n+1-h}{\ell} (1-(h+\ell)x)^n_+,$$

where $a_+ = \max(a,0)$.

Proof. Lemma 3 has been proved originally by Whitworth (1887), and also by Fisher (1929, 1940), Darling (1953), Flatto and Konheim (1962), Kendall and Moran (1963) p. 31, Feller (1966), p. 28, and Holst (1980). Its proof may be derived from the fact that, if