Mathematical Statistics and Applications

Volume B

Edited by
W. Grossmann,
G. Ch. Pflug,
I. Vincze, and
W. Wertz

MATHEMATICAL STATISTICS AND APPLICATIONS

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

Volume B

Edited by

WILFRIED GROSSMANN

Institute of Statistics and Computer Science, University of Vienna, Austria

GEORG Ch. PFLUG

Institute of Mathematics, University of Gießen, FRG

ISTVÁN VINCZE

Mathematical Institute, Hungarian Academy of Sciences, Budapest, Hungary

WOLFGANG WERTZ

Institute of Statistics, 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

Pannonian Symposium on Mathematical Statistics (4th:
1983: Bad Tatzmannsdorf, Austria)
Mathematical statistics and applications.

(Proceedings of the 4th Pannonian Symposium on Mathematical Statistics, Bad Tatzmannsdorf, Austria, 4–10 September 1983; v. B)
Bibliography: p.
Includes index.
1. Mathematical statistics—Congresses. I. Grossmann, Wilfried, 1948– . II. Series: Pannonian Symposium on Mathematical Statistics (4th: 1983: Bad Tatzmannsdorf, Austria). Proceedings of the 4th Pannonian Symposium on Mathematical Statistics, Bad Tatznannsdorf, Austria, 4–10 September 1983; v. B.
QA276.A1P36 1983 vol. B 519.5 s [519.5] 85-14464
ISBN 90-277-2088-6

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 objectives 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 21 reviewed contributions which cover various aspects of the application of mathematical statistics. A second group of papers dealing with problems of probability theory and decision theory is published in a separate volume entitled "Probability and Statistical Decision Theory".

Roughly speaking, the papers can be grouped into three main categories. The first group is the application of probability theory. A special type of application is shown in the invited paper of P. Erdös, namely probabilistic methods in number theory. Further models of applied probability covered by the papers are game theory, urn models, best choice models and random graphs. The second group could be best characterized by the term mathematical statistics for models of real data. Such models are linear models, regression, discrimination, time series, analysis of censored data, goodness of fit approximation of processes. The papers show the increasing importance of new theoretical results (i.e. in nonparametric regression, robustness or limit theorems) to applied problems. The last group of papers shows various aspects of computational statistics and the application of Monte Carlo methods for numerical problems such as integration or optimization.

We wish to express our thanks to many persons who gave us indispensable aid in the preparation of this volume: J. Anděl, O. Bunke, I. Csiszár, L. Devroye, B. Gyires, J. Hurt, M. Huškova, T. Nemetz, E. Neuwirth, H. Niederreiter, D. Plachky, B. Pötscher, F. Pukelsheim, P. Révész, E. Ronchetti, H. Stetter, W. Stute, K. Urbanik, and R. Zieliński, 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 Authority of Bad Tatzmannsdorf; special thanks go to Th. Kery and Dr. R. Grohotolsky (Head and Vice-Head of Burgenland, respectively), 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 WERTZ

CONTENTS

Proc. of the 4th Pannonian Symp. on Math. Stat., Bad Tatzmannsdorf, Austria 1983
W. Grossmann, G. Pflug, I. Vincze, W. Wertz, eds

SOME APPLICATIONS OF PROBABILITY METHODS TO NUMBER THEORY

P. ERDÖS

Budapest

I wrote many papers on these subjects. In this note I will mainly state some unsolved problems which have perhaps been somewhat neglected and which perhaps cannot be solved completely by probability methods, but where slightly weaker results can be obtained by these methods. I will mainly discuss my own problems and those of my coworkers not because I consider these more important but because I hope I know more about them than the reader. Also since my memory is still good I will occasionally add some historical remarks.

First a few words about the literature, this list of references will be very incomplete.

P.D.T.A. Elliott, Probabilistic number theory, Springer Verlag, Grundlehren der Math. Wissenschaften, Vol. 239 and 240 (1980). This comprehensive handbook deals with additive and multiplicative arithmetical functions, it contains also many references to the earlier literature and many unsolved problems and interesting historical remarks.

H. Halberstam and K. F. Roth, Sequences, Springer Verlag 1982. The third chapter of this excellent book contains applications of probability methods to additive number theory.

The first survey paper on probability methods in number theory is M. Kac, Probability methods in some problems of analysis and number theory, Bull. Amer. Math. Soc. 55 (1949), 641-665. I feel that this interesting paper deserves careful study even now.

I will give references only if they are not contained in these books.

Before I start my subject I just remark that probabilistic ideas are often useful in making plausible conjectures which cannot be attacked by our methods which are at our disposal at present. The best known such conjecture is due to Cramer: Let $p_1<p_2<\ldots$ be the sequence of consecutive primes. Then

$$(1) \qquad \overline{\lim} \frac{p_{n+1}-p_n}{(\log n)^2} = 1.$$

The Riemann hypothesis would only imply $p_{n+1}-p_n<p_n^{\frac{1}{2}+\varepsilon}$ and the currently known best inequality from below due to Rankin states that there is a $c>0$ so that for infinitely many n

$$(2) \quad p_{n+1}-p_n > \frac{c \log n \ \log\log n \ \log\log\log\log n}{(\log\log\log n)^2}$$

Rankin obtained (2) 45 years ago, no progress has been made since then except that Schönhage and Rankin himself improved the value of c. This prompted me to offer a reward of 10000 dollars for a proof that (2) holds for every c and infinitely many n. I am so sure that (2) holds that for a disproof I offer 25000 dollars. The only reason that I do not offer 10^6 dollars that is the very unlikely event that I am wrong and (2) does not hold for every ε I could not pay my debt.

Let me state two less well known conjectures of myself: Denote by $P(m)$ the greatest prime factor of m. Is it true that for every $n>n_0(\varepsilon)$ $P(n(n+1)) > (\log n)^{2-\varepsilon}$, but for infinitely many n $P(n(n+1)) < (\log n)^{2+\varepsilon}$. Also is it true that every $n > n_0(\varepsilon)$ can be written in the form $n=a+b$ $P(a\cdot b) < (\log n)^{2+\varepsilon}$? It is easy to see that the result fails if we replace $(\log n)^{2+\varepsilon}$ by $(\log n)^{2-\varepsilon}$. Very much weaker positive results have been proved by analytic methods by Balog and Sárközy. Balog and Sárközy will publish several papers on this and related subjects.

A. Balog and A. Sárközy, On sums of integers having small prime factors, I-II, Studia Sci. Math. Hung., and On sums of sequences of integers, I-III, Acta Arithmetica and Acta Math. Acad. Sci. Hung., to appear.

H. Cramer, On the order of magnitude of the difference between consecutive prime numbers, Acta Arith. 2 (1936), 23-46.

R. A. Rankin, The difference between consecutive prime numbers, J. London Math. Soc. 13 (1938), 242-247.

First a few words about additive and multiplicative number theoretic functions, this chapter will be very short in view of the book of Elliott. I proved many years ago that the density of integers n for which $\varphi(n) > \varphi(n+1)$ is $\frac{1}{2}$ and the same result holds for $\sigma(n)$ and for $d(n)$ the number of divisors of n. This later result was a conjecture of Chowla. I could never prove that the density of integers with

$P(n) > P(n+1)$ is $\frac{1}{2}$ and this conjecture is probably unattackable by methods at our disposal. Very much weaker results have been proved by Pomerance and myself.

On the other hand easy independence arguments will give that the density of integers for which $\varphi(n) > \varphi(n+1)$ and $d(n) > d(n+1)$ is $\frac{1}{4}$ since the two inequalities are asymptotically independent. Put $f(n) = \sum_{p|n} \frac{1}{\log\log p}$. With very little more trouble I can prove that the three inequalities $d(n) > d(n+1)$, $f(n) > f(n+1)$ and $\varphi(n) > \varphi(n+1)$ are asymptotically independent. On the other hand $\varphi(n) > \varphi(n+1)$ and $\sigma(n) < \sigma(n+1)$ are strongly correlated. The density of integers which satisfy both inequalities is strictly between $\frac{1}{4}$ and $\frac{1}{2}$. Finally the density of integers with $d(n) > d(n+1)$, $\omega(n) > \omega(n+1)$ is $\frac{1}{2}$ ($\omega(n)$ denotes the number of distinct prime factors of n). I am not sure if these results are in the literuture but anyone familiar with the methods of probabilistic number theory can easily supply the proofs. I just want to state one of my old problems which does not seem quite hopeless but which so far resisted all attacks. Let $f(n)$ be an additive function and assume that for a pair a, b of real numbers the density of integers n for which $a<f(n)<b$ exists and is positive. Does it then follow that $f(n)$ has a limiting distribution? It is not difficult to prove that our condition implies that the two series

$$\sum_{|f(p)|>1} \frac{1}{p}, \qquad \sum_{|f(p)|<1} \frac{f(p)^2}{p}$$

both converge, but I could not prove that

$$\sum_{|f(p)|<1} \frac{f(p)}{p}$$

also converges. If this would also be proved then by the theore of Wintner and my conjecture would be settled. Elliott gave a purely probabilistic formulation of my conjecture (see Vol. 2, 331 of Elliott's book).

An old conjecture of mine on multiplicative functions sta

Let $f(n)$ be a multiplicative function which only takes t values ± 1. I conjectured that $f(n)$ has a mean value, i.e. that

(3) $$\lim \frac{1}{x} \sum_{n=1}^{x} f(n)$$

exists, and is 0 if and only if $\sum_{f(p)=-1} \frac{1}{p} = \infty$. (3) was first proved by Wirsing and later in a more general form by Halász. Tchudakoff and I conjectured (independently) that for such a multiplicative function

(4) $$\overline{\lim} \left| \sum_{n=1}^{x} f(n) \right| = \infty.$$

In fact I have a more general conjecture: Let $f(n) = \pm 1$ be any number theoretic function (not necessarily multiplicati then to every c there is a d and m so that

(5) $$\left| \sum_{k=1}^{m} f(kd) \right| > c.$$

(5) was one of my first conjectures and is now more than 50 years old. I offer 500 dollars for a proof or disproof of (5).

Choose $f(p) = +1$ or $f(p) = -1$ with probability $\frac{1}{2}$ and for $n = \Pi\, p_i^{\alpha_i}$ put $f(n) = \prod_{p|n} f(p)^{\alpha_i}$. Wintner asked what can be said about $\sum_{n=1}^{x} f(n)$ for almost all choices of $f(p) = \pm 1$. He proved that for almost all functions $f(n)$ we have

$$(6) \qquad \sum_{n=1}^{x} f(n) < x^{\frac{1}{2} + \varepsilon} .$$

I improved (6) and showed that for almost all functions $f(n)$

$$c_1 \left(\frac{x}{\log x}\right)^{1/2} < \sum_{n=1}^{x} f(n) < x^{1/2} (\log x)^{c_2}$$

and conjectured that for almost all functions

$$(7) \qquad \overline{\lim}\, \frac{1}{x^{1/2}} \sum_{n=1}^{x} f(n) = \infty$$

but

$$(8) \qquad \lim \frac{\sum_{n=1}^{x} f(n)}{x^{1/2} (\log x)^{\varepsilon}} = 0$$

I could not even guess the analog of the law of the iterated logarithm. Halász proved (8), but only proved a slightly weaker result than (7). As far as I know nobody has a plausible guess for the true order of magnitude of $\left|\sum_{n=1}^{x} f(n)\right|$.

Many further interesting questions could be asked e.g.: What can be said about the number of 0-s of the partial sums $\sum_{n<x} f(n)$ for the random multiplicative function? By analogy with the Rademacher functions one would expect that the number of zeros is between $x^{\frac{1}{2}-\varepsilon}$ and $x^{\frac{1}{2}+\varepsilon}$,

of course more precise results would be very desirable.

G. Halász, On random multiplication functions, Publ. Math. D'Orsay 1983, Journées Arith. Coll. H. Delange 79-96.

P. Erdös and Carl Pomerance, On the largest prime factors of n and n+1, Aequationes Math. 17 (1978), 311-321.

Now I discuss some problems and results on additive number theory. These problems were first stated by Sidon more than 50 years ago, he was led to these problems by his study of lacunary trigonometric series. Let $a_1 < a_2 < \dots$ be an infinite sequence of integers, denote by $f(n)$ the number of solutions of $n=a_i+a_j$. Sidon asked me in 1932 when we first met whether there is a sequence $A = \{a_1 < a_2 < \dots\}$ for which $f(n) > 0$ for all n but for which $f(n)/n^{\varepsilon} \to 0$ for every $\varepsilon>0$ i.e. A is a basis of order 2 but $f(n)$ is small. I at first thought that the problem will not be hard and that it will be easy to construct such an A. I never succeeded in constructing such an A but about 20 years later I proved by probabilistic methods that there is a sequence A for which

$$c_1 \log n < f(n) < c_2 \log n \tag{9}$$

holds for every n. An outstanding problem here is whether there is a sequence A for which

(10) $$f(n) = (1 + o(1)) \log n$$

I expect that such a sequence A does not exist. As a first step one should prove that for every sequence A

(11) $$\overline{\lim} \, |f(n) - \log n| = \infty$$

(11) will perhaps not be hard to prove. I offer 500 dollars for a proof or disproof of (10). If there is no A satisfying (10) then one could ask: Put

$$\overline{\lim}\, f(n)/\log n = C_1, \ \underline{\lim}\, f(n)/\log n = C_2$$

Is there an $\varepsilon>0$ so that for every A, $C_1/C_2 > 1+\varepsilon$?

This question just occured to me while I write these lines and I hope it will not turn out to be trivial. Let $g(n)$ be a monotonic function which tends to infinity arbitrarily slowly. It is easy to prove by the probability method that there is a sequence A for which

$$f(n)/g(n) \log n \to 1.$$

An old conjecture of Turán and myself states that if

(12) $f(n) > 0$ for all n then $\overline{\lim}\, f(n) = \infty$

I offer 500 dollars for a proof or disproof of (12).

Perhaps $f(n) > 0$ for all $n > n_0$ already implies that there is an absolute constant c so that $f(n) > c \log n$. If this conjecture is true one can again ask: Is it true that in fact $c > c_0$?

Another possible strengthening of my conjecture with Turán would be: Assure $a_k < c\, k^2$. Is it then true that $\overline{\lim}\, f(n) = \infty$ and perhaps even $f(n) > c \log n$ for infinitely many n?

Sidon calls an infinite sequence A a B_2 sequence if the integers $a_i + a_j$ are all distinct. When we first met Sidon asked for a B_2 sequence for which a_n increases as slowly as possible. The greedy algorithm easily gives that there is a B_2 sequence for which $a_k < c\, k^3$ and we both conjectured that in fact there is a B_2 sequence for which $a_k < k^{2+\varepsilon}$. Rényi and I proved by probabilistic methods that there is a sequence A with $a_k < k^{2+\varepsilon}$ and $f(n) < C_\varepsilon$ for all n. For nearly 50 years we could not prove that there is a B_2 sequence for which $a_k = o(k^3)$. A few years ago Ajtai, Komlós and Szemerédi proved by an ingenious combination of combinatorial and probabilistic methods that there is a B_2 sequence for which

$$a_k < c\, k^3 / \log k. \tag{13}$$

(13) at the moment seems to be the natural boundary of their method.

Let me state an interesting problem in connection with the greedy algorithm. We construct a B_2 sequence as follows:

Assume $a_1<a_2<\dots<a_{k-1}$ has already been constructed. Then a_k is the smallest integer for which $\{a_1,\dots,a_k\}$ is a B_2 sequence. It is easy to see that $a_k<c\,k^3$ and Chowla and Mian carried out extensive calculations on the basis of which they suggest that $a_k \sim k^{2+c}$. It is not even known if

$$(14)\quad a_k/k^2 \to \infty \qquad \text{and} \qquad a_k/k^3 \to 0.$$

I am not an expert on algorithms but find (14) a fascinating conjecture and offer 250 dollars for a proof or disproof.

I proved that if A is a B_2 sequence then

$$(15)\qquad \overline{\lim}\ \frac{a_k}{k^2 \log k} \geq c > 0$$

for a certain $c>0$. Is (15) best possible? I have no information and offer 500 dollars for a proof or disproof. It is best possible in the following much weaker sense. I proved (15) by showing that if $a_k < c_1 k^2 \log k$ for all $k > k_0$, c_1 sufficiently small, then the number of solutions $g(x)$ of $0 < a_j - a_i < x$ satisfies $g(x) > x$. Thus our sequence cannot be a B_2 sequence. It is not hard to show that (15) is best possible in this case i.e. if $a_k = C\,k^2 \log k$, C sufficiently large, then $g(x)<x$. Krickeberg and I proved that there is a B_2 sequence for which

$$(16)\qquad \underline{\lim}\ a_k/k^2 < C$$

for some C. The best result in (16) is due to Krickeberg: