Translations of

# MATHEMATICAL MONOGRAPHS

Volume 194

# Records: Mathematical Theory

Valery B. Nevzorov


**American Mathematical Society**
Providence, Rhode Island

# Contents

# Preface

In 2002 the mathematical theory of records will mark its 50th anniversary. The first paper on this subject by Chandler (1952) attracted the attention of many researchers and inspired many new publications. It happens sometimes that, by novelty of its subject or brilliant results, a mathematical paper arouses and maintains interest of numerous researchers to a "fashionable" problem for several years or even decades, but as the "vein of gold" is exhausted this interest drops off sharply. This was not the case for records. The number of publications on this subject has been increasing exponentially, doubling about every 10 years (around 80 in 1977, about 160 in 1987, and more than 300 by the beginning of 1998). In the Introduction, we try to explain this phenomenon. Here we only point out that numerous models of records provide a convenient object for application of various mathematical methods; on the other hand, there is a lot of numerical data on records in sports (national, Olympic, world records), hydrology (for instance, the floods in St. Petersburg have been recorded for 300 years), meteorology (the reader of course has heard on radio or TV something like "the air temperature today in our city was the lowest for the last 95 years"), etc., which motivate the mathematicians to build models compatible with the available record observations and to try to predict the future record values.

It is difficult to separate the theory of records from the theory of order statistics. Records are especially closely related to extremal order statistics. A systematic exposition of the theory of order statistics and extremes can be found in books by H. A. David "Order Statistics" (1970, 1981) and J. Galambos "The Asymptotic Theory of Extreme Order Statistics" (1978, 1987). Regarding records, a comparatively detailed review of results (without proofs) and related bibliography can be found only in the form of articles (Nevzorov (1987), Nagaraja (1988), Nevzorov and Balakrishnan (1998)). As mentioned above, the number of publications has practically doubled for the last 10 years. Hence it becomes necessary to relate the classical results for records with the latest advances. This book can be viewed as an "Introduction into the Theory of Records." We tried to present the material in the form combining the features of a textbook and a survey of literature. Thus, after learning the basic methods utilized in the theory of records, the reader will find in Appendix 1 bibliographical notes which, together with the list of references comprising about 300 papers on records and related topics, will allow him to acquire a deeper knowledge of the subject. Moreover, this book contains about 50 exercises which will allow the reader to assess the degree of his mastering the material. Hints and solutions are collected in Appendix 2.

The book is written on an intermediate level which presumes the knowledge of only standard courses of probability theory and mathematical statistics. In the first part of the book we present briefly the necessary material on order statistics

which is used in the theory of records. The reader interested in a more detailed knowledge of the theory of order statistics and its applications is referred to the book by David (1970) mentioned above and references therein.

The book can be used for preparing courses on order statistics and records, as well as for studying these areas of probability theory and mathematical statistics on one's own. A large part of the material was employed by the author in lecture courses on order statistics and records at the Faculty of Mathematics and Mechanics at the St. Petersburg State University and the Department of Statistics at Ohio State University.

I hope that the methods for analysis of records, numerous record models, and the various applications of records treated in the book will attract the attention not only of mathematicians, but also of engineers (especially those engaged in reliability of constructions), actuaries, sport statisticians, specialists in hydrology, meteorology, gerontology, and many others dealing with analysis of extremal values of various random variables and processes.

Writing the book gives me an opportunity to set out systematically the results of my research on order statistics and records. It is my pleasure to express gratitude to V. V. Petrov from whom I always received attention and support beginning with my first steps in science. I am indebted to my former students A. V. Stepanov and S. V. Malov, now scientists themselves, who have been the most scrupulous readers of my works, and with whom constant communication stimulated my research in this area. Of great importance for my scientific activity was an exchange of ideas with colleagues. I am thankful to M. Ahsanullah, N. Balakrishnan, P. Deheuvels, V. A. Egorov, G. Haiman, and H. N. Nagaraja with whom I obtained a number of joint results included in this book.

Comments useful in improving the text were made by S. M. Ananievskii, D. M. Chibisov, O. V. Rusakov, and V. V. Slavova. Advice from A. M. Trevgoda and assistance from my wife Lyudmila and my son Igor helped to speed up the process of typesetting the manuscript.

V. B. Nevzorov

St. Petersburg, April 2000

# Introduction

*Lecture* 1

One of the most popular books in the world, after the Holy Bible, is the "Guinness Book of Records" first published in September 1955. For almost half a century this book, regularly updated and republished, stimulated the appearance of many similar publications, such as the "St. Petersburg Book of Records" first published in 1995.

Why are records so popular? Maybe this is because we often encounter them in everyday life, single out the records from the multitude of data, fix and memorize the record values. Of particular interest are records in sports and record values related to natural phenomena. The spectators are drawn to stadiums not only by competition excitement, but also by an opportunity to witness new record achievements which broaden the horizons of human abilities. Indeed, for the majority of people who watch sport competitions or participate in them, achievement of a new record is undoubtedly associated with progress. In order to attract potential record-breakers (and, in turn, numerous spectators and advertisers), the organizers of many athletic competitions set special prizes for breaking records, which are, as a rule, of higher value than the prizes for the winners. One can frequently see a show where daredevils risk their necks trying to break a speed record on water, ground, or in the air, or to leap a motorcycle over a record number of cars, or to achieve a record depth without using an aqualung, in order to place their names into the book of records. After Sir Edmund Hillary and Tenzing Norgay reached the summit of Mt. Everest in 1953, having thus achieved the absolute record in mountaineering for our planet, other mountaineers found the "way out" in climbing Everest or other mountains along the routes of record difficulty. People are excited by records related to geography and the various natural or social phenomena. Tourists strive to see the highest waterfall in the world, the oldest tree, or the largest cave. Cities compete with one another by erecting the highest skyscraper or building the largest stadium. Journalists will not miss a chance to report on the birthday of the oldest person in the world, to describe the life of the richest person, or to show the consequences of the most devastating hurricane in a particular region. In weather reports, newspapers publish the lowest and highest temperatures observed in the locality for the last, say, 100 years, and readers are interested to see if these local records would be broken, while old-timers tell stories to envying listeners about the extremely cold winter they survived some sixty years ago.

In English the word "record" means not only *the unsurpassed performance* in some area, but also *a report, account, chronicle, diary,* or *relics of the past.* One can see that all these meanings are logically interrelated.

Each record "achievement" — such as the oldest record in the Guinness Book of Records which is due to St. Simeon Stylites the Younger who lived for 45 years standing on top of a stone pillar near Antioch, Syria, some 14 centuries ago; the record miracle near the Sea of Galilee when five thousand men, without counting women and children, were fed with five loaves of bread and two fishes; the pandemic of plague in Europe in XIV century that claimed about 75 million lives; or the fantastic Bob Beamon's long jump of 8.90 m in 1968 — is registered in chronicles and accounts or is preserved in people's memories.

Even in mathematics, record achievements are not infrequent. It is well known, for instance, that the sequence of prime numbers is infinite and there is no largest prime number. Hence, attempts to demonstrate the largest particular prime number continue persistently. The enthusiasm of researchers combined with computer equipment progress resulted in obtaining successively in 1996–1999 the following record values of prime numbers:

$$2^{1257787}-1,\ 2^{1398269}-1,\ 2^{2976221}-1,\ 2^{3021377}-1,\ \text{and}\ \ 2^{6972593}-1.$$

The last of them, found by N. Hajratwall, G. Woltman, S. Kurovski, *et al.* on June 1, 1999, consists of 2,098,960 digits for decimal representation. Maybe by now the reader can write down a much larger prime number.

Another example refers to the probability theory. A longstanding research was devoted to evaluation of the absolute constant in the well-known Berry–Esseen inequality which sets a bound on the difference between the normal distribution and the distribution of a sum of independent identically distributed random variables with finite third moment. Esseen (1956) has shown that this constant is no less than

$$\frac{3+\sqrt{10}}{6\sqrt{2\pi}}.$$

The attempts to approach this value were made by Berry, Esseen, Bergström, Zolotarev, and others. Making substantial use of computer facilities, Shiganov (1982) obtained for this constant the value 0.7655 lowering by 0.032 the previous record 0.7975 which was due to van Beek (1972). Thus it took 10 years to beat the record.

The records as such are memorials of their time. The annals of records reflect the progress in science and technology and enable us to study and forecast the evolution of mankind on the basis of record achievements in various areas of its activity.

This motivates the necessity to construct mathematical models of records and to develop the corresponding mathematical theory. In 1952 there appeared the first paper on the mathematical theory of records by Chandler (1952). Since then there were many mathematicians who devoted their attention to this subject. All of them, working in different areas of probability theory and mathematical statistics, found interesting problems related to records which required diverse mathematical methods. Presently the author's card-file contains more than 300 papers, half of which appeared during the last 10 years. In a number of them, mathematical models for the analysis of growth of Olympic or world records were suggested (with special attention given to the mile run) and, based on this analysis, attempts were made to predict future records. While the forecast of sports achievements serves mostly for comprehension and is important primarily for the authors of record models themselves as a possibility to assess the fit of the models to real processes, a

forecast of the scale of the next record earthquake or flood would make it possible to take adequate precautions.

Statistics of the various records contains a large number of data which sometimes cover a very long time interval. Hence a statistician processing the data often deals with a censored sample containing only the record values. This happens, for example, in estimating the durability of some devices in the situation where the production process is modified after each failure. A similar situation arises in processing sports data when out of the multitude of data registered in numerous competitions of different level, only the most valuable and informative ones are preserved for history, and, of course, records are this kind of data.

It turns out, surprisingly, that the theory of records is connected with cycles of random permutations, with some algorithms of linear programming, with forming platoons (caravans) of vehicles (when moving in a single lane road for a long time, the vehicles form platoons whose leaders' speeds are a sequence of lower record values), with the optimal selection problem (the so-called secretary problem).

The theory of records relies largely on the theory of order statistics, and is especially closely connected to extreme order statistics. Hence we will discuss first the order statistics and their properties.