Preface

This monograph is devoted to the topic of univariate continuous distributions *on a bounded domain other than the beta distribution.* As early as 1919, E. Pairman and K. Pearson recognized the topic of continuous distribution on a limited range to be an important aspect of statistical theory and investigated estimation of their moments. However, even in the late nineties of the 20-th century only a relative few number of probabilistic models of this kind were available. Amongst them, the uniform, triangular and beta distributions are the most widely explored and used, interspersed by some "curious" distributions posed occasionally as problems or exercises. Other bounded continuous distributions were based on mathematical transformations of the normal distribution (with an unbounded domain) - the most wide-spread amongst them is still the Johnson S_B family of transformations introduced in 1949 (briefly discussed in Appendix B).

Aside from the latter system of distributions, a comparison with the multitude of existing unbounded continuous distributions developed in the 20-th century is striking and even somewhat puzzling. There are of course historical reasons for this discrepancy – the main one being that the basic origin of continuous distributions stems from the famous Pearson family (containing mostly unbounded distributions) which in turn is related to the most prominent continuous distribution – the so called Gaussian or normal (symmetric) distribution (likewise unbounded). This distribution has a multitude of names. Interested readers should consult The History of Statistics by S.M. Stigler (1986) or/and Hald (1990). Perhaps it is no coincidence that the latter distribution was used as basis for the Johnson S_B system of bounded distributions. It should be noted that the unbounded continuous distributions usually computationally have the property that estimation of their parameters does not pose serious problems since originally estimation procedures were developed with an eye on the unbounded families.

Even when dealing with a problem of a bounded nature, such as, e.g., the propensity of particular physical characteristics of the human body, unbounded distributions such as the Gaussian distribution have been used to describe the uncertainty about these characteristics. Situations are not limited to the "physical" problems. For example, in the world of finance logarithmic transforms of interest rates (typically a quantity between 0% and 100%) are combined with their one-step differences to arrive at an unbounded domain, allowing once again the use of the Gaussian distribution for describing uncertainty in these one-step differences. In cases similar to the one described above, authors usually emphasize that they are using the unbounded continuous distribution as an *approximation* to the actual bounded state of affairs. In our opinion such an approach is not quite necessary in the 21-st century when computational difficulties associated with estimation of parameters of a distribution do not pose problems any more and are easily overcome. At the very least, a more natural continuous distribution with a bounded domain should be available that describes the uncertainty of such bounded phenomena.

The only continuous bounded distribution discussed extensively in numerous textbooks on probability and statistics is the well known beta distribution or the Pearson type IV distribution, which is highly flexible and has been used – occasionally indiscriminately – for fitting of data stemming from various fields. This approach is usually empirical since the parameters of the beta distribution (in its original form) do not have a proper physical meaning and having fitted beta distributions, the fitted beta parameters themselves do not always provide a picture of the phenomenon generated by the data under scrutiny.

Moreover, like the normal distribution, the beta distribution is smooth. Whereas a "peaked" alternative for the normal distribution has been available for quite some time such as the Laplace distribution, a flexible "peaked" alternative for the beta distribution has been lacking until very recently. Smoothness of density curves may be an attractive mathematical property, but it does not always have to be dictated by the uncertainty of the phenomenon one is attempting to describe. In particular, financial data has been shown to exhibit "peaked" histograms and even displayed jump discontinuities. This fact prompted us to search for "alternative curves" on a bounded domain which:

(1) possess properties of the uncertain phenomenon one is attempting to describe,
(2) mimics the flexibility of the beta distribution and
(3) possess meaningful interpretation of parameters which permit to classify the data and draw practical conclusions.

Our initial efforts originated as early as 2000 and resulted in the construction of the *two-sided power (TSP) distributions* which cover almost all varieties of the beta distribution, enhances flexibility in the unimodal domain, possess the cumulative distribution function and quantiles in a closed form, and whose parameters provide information on the structure and properties of the distribution at hand. The distribution was independently introduced in a technical report in 1985 by B.W. Schmeiser and R. Lal.

The material available by now (2004) on bounded continuous distributions allowed us to compile a monograph consisting of 8 chapters containing predominantly novel contributions together with some neglected models. Indeed, while searching the literature, we have encountered a number of relatively obscure continuous bounded distributions, having attractive statistical properties which extend the realm of models that can be represented by bounded continuous data. A common thread amongst most of these distributions is that they can be viewed to be extensions of the triangular distribution. It would seem that these distributions were not sufficiently covered in the relevant periodical and monographic literature prior to the 21-st century. As an example, aside from the authors recent publications, no maximum likelihood procedure seems to have been available for the widely encountered three-parameter triangular distribution. It appears that only some ad-hoc estimation methods have been provided in distribution fitting software that do not contain detailed descriptions of their fitting procedures.

The initial Chapters 1 to 2 cover the triangular distribution and some of its pre 21-st century extensions, which we consider to be of substantial historical and of practical value. The text of Chapters 3 to 8 records – to a large extent – our own contributions to the post-20-th century alternatives to the beta distribution which are related to the TSP family possessing attractive properties not always shared by the beta distribution. The distributions presented in Chapters 3 to 7 are motivated by utilizing real world engineering and economic data. Some of the illustrative examples are elaborate and in Chapters 3 and 6 involve modeling of financial time series and in Chapter 7 deal with modeling of income distributions for African American, Hispanic and Caucasian (non-Hispanic) subpopulations in the U.S. revealing stochastic ordering amongst them. Examples in the remaining chapters are based on civil engineering and physics data.

We are not aiming to dethrone or demolish the time-honored beta distribution which will no doubt continue to have a secure place and plenty of meaningful applications in statistical, engineering, medical, econometrics and other applied literature, as indicated by the recently published handbook on the beta distribution by Gupta and Nadarajah (2004). Our intention is to enlarge the arsenal of univariate continuous bounded distributions and thus improve the description of those statistical models to which the newly proposed distributions may well be applicable. While in this edition we predominantly focus on the triangular distributions and its extensions, we hope to provide a more comprehensive treatment of univariate continuous distributions on a bounded domain in a future edition and emphasize the computer-oriented aspects.

We are grateful to a number of people for helping us to bring this year long project to fruition. In preparing this monograph we have benefited from the advise of Prof. T.A. Mazzuchi (The George Washington University), Prof. S. Nadarajah and as well as suggestions from the editors and referees of our own published papers on this topic. We are indebted to Dr. Simaan AbouRizk (University of Alberta) for alerting us to his dissertation as a source of the civil engineering data utilized in Chapters 1 and 4, and Dr. David Findley (U.S. Bureau of Census) for assisting us to obtain recent income distribution data for Chapter 7. Our thanks go to our student Caner Sener for careful proofreading of the manuscript and drawing our attention to a number of misprints. Needless to say, the responsibility for remaining misprints and errors is ours alone. The positive attitude and encouragement from the editors of the World Scientific Publications was very valuable. We are especially indebted to our editor Ji Zhang who generously devoted her time and have shown enthusiasm for the project. Ji Zhang is a marvelous editor with an uncanny sense on how to reshape a narrative; working with her was a delight. Finally, we want to express gratitude to our wives Greer and Rosalie for encouragement and displaying patience while we were toiling away obsessively with the manuscript and for providing a balance to our lives.

We trust that this monograph will fill a gap in the literature on statistical distributions and readers will find our efforts useful in their work. We welcome comments, suggestions and criticism from the readers.

Johan René van Dorp and Samuel Kotz
September, 2004
Washington, DC, U.S.A.