Preface

This book is divided into two parts. The first part is devoted to some advances in testing for a stochastic ordering, and the second part is related to ANOVA procedures for nonparametric inference in experimental designs. It is worth noting that, before introducing specific arguments in the two main parts of the book, we provide an introductory first chapter on basic theory of univariate and multivariate permutation tests, with a special look at multiple-comparison and multiple testing procedures.

The concept of stochastic ordering of distributions was introduced by Lehmann (1955) and plays an important role in the theory of statistical inference. It arises in many applications in which it is believed that, given a response variable Y and an explanatory variable x, the statistical model assumes that the distribution of $Y|x$ belongs to a certain family of probability distributions that is ordered in the sense, roughly speaking, that large values of x lead to large values of the Y's.

Many types of orderings of varying degrees of strength have been defined in the literature to compare the order of magnitude of two or more distributions (see Shaked and Shanthikumar, 1994, for a review). These include likelihood ratio ordering, hazard rate ordering, and simple stochastic ordering, which are perhaps the main instances. On the one hand, these orderings make the statistical inference procedures more complicated. On the other, they contain statistical information as well, so that if properly incorporated they would be more efficient than their counterparts, wherein such constraints are ignored. These considerations emphasize the importance of statistical procedures to detect the occurrence of such orderings on the basis of random samples. Inference based on stochastic orderings for univariate distributions has been studied extensively, whereas for multivariate distributions it has received much less attention because the "curse of dimensionality" makes the statistical procedures considerably more complicated. For a review of constrained inference, we refer to the recent monograph by Silvapulle and Sen (2005).

Likelihood inference is perhaps the default methodology for many statistical problems; indeed, the overwhelming majority of work related to order-restricted problems is based on the likelihood principle. However, there are instances when one might prefer a competitive procedure. Recently there have been debates about the suitability of different test procedures: Perlman and Chaudhuri (2004a) argue in favor of likelihood ratio tests, whereas Cohen and Sackrowitz (2004) argue in favor of the so-called class of directed tests. In multidimensional problems, it is rare that a "best" inference procedure exists. However, even in such a complex setup, following Roy's union-intersection principle (Roy, 1953), it might be possible to look upon the null hypothesis as the intersection of several component hypotheses and the alternative hypothesis as the union of the same number of component alternatives, giving rise to a multiple testing problem. A classical approach is to require that the probability of rejecting one or more true null hypotheses, the familywise error rate (Hochberg and Tamhane, 1987), not exceed a given level. Generally, it is surprising that some existing procedures seem to be satisfied to stop with a global test just dealing with the acceptance or rejection of the intersection of all null hypotheses. In the form presented, it will be difficult to interpret a statistically significant finding: The statistical significance of the individual hypotheses in multiple-endpoint or multiple-comparison problems remains very important even if global tests indicate an overall effect. Indeed, most clinical trials are conducted to compare a treatment group with a control group on multiple endpoints, and the inferential goal after establishing an overall treatment effect is to identify the individual endpoints on which the treatment is better than the control. For tests of equality of means in a one-way classification, the ANOVA F test is available, but in the case of rejection of the global null hypothesis of equality of all means, one will frequently want to know more about the means than just that they are unequal.

In the majority of the situations we shall deal with, both the hypothesis and the class of alternatives may be nonparametric, and as a result it may be difficult even to construct tests that satisfactorily control the level (exactly or asymptotically). For such situations, we will consider permutation methods that achieve this goal under fairly general assumptions. Under exchangeability of the data, the empirical distribution of the values of a given statistic recomputed over transformations of the data serves as a null distribution; this leads to exact control of the level in such models. In addition, by making effective use of resampling to implicitly estimate the dependence structure of multiple test statistics, it is possible to construct valid and efficient multiple testing procedures that strongly control the familywise error rate, as in Westfall and Young (1993).

We bring out the permutation approach for models in which there is a possibly multivariate response vector $\boldsymbol{Y}$ and an ordinal explanatory variable x taking values $\{1, \ldots, k\}$, which can be thought of as several levels of a treatment. Let $\boldsymbol{Y}_i$ denote the random vector whose distribution is the conditional distribution of $\boldsymbol{Y}$ given $x = i$. We are interested in testing $\boldsymbol{Y}_1 \overset{\mathrm{d}}{=} \ldots \overset{\mathrm{d}}{=} \boldsymbol{Y}_k$ against a stochastic ordering alternative $\boldsymbol{Y}_1 \overset{\mathrm{st}}{\leq} \ldots \overset{\mathrm{st}}{\leq} \boldsymbol{Y}_k$ with at least one $\overset{\mathrm{st}}{\lneq}$.

In the statistical literature, there is relatively little on multivariate models for nonnormal response variables, such as ordinal response data. This is perhaps due to the mathematical intractability of reasonable models and to related computational problems. The aim is therefore to provide permutation methods that apply to multivariate discrete and continuous data. We deal with univariate and multivariate ordinal data in Chapters 2 and 3, respectively, and Chapter 4 contains results for multivariate continuous responses.

As previously said, the second part of the book is dedicated to nonparametric ANOVA within the permutation framework. Experimental designs are useful research tools that are applied in almost all scientific fields. In factorial experiments, processes of various natures whose behavior depends on several factors are studied. In this context, a factor is any characteristic of the experimental condition that might influence the results of the experiment. Every factor takes on different values, called levels, that can be either quantitative (dose) or qualitative (category). When several factors are observed in an experiment, every possible combination of their levels is called a treatment. The analysis of factorial designs through linear models allows us to study (and assess) the effect of the experimental factors on the response, where factors are under the control of the experimenter. They also allow for evaluating the joint effect of two or more factors (also named main factors), which are known as interaction factors. The statistical analysis is usually carried on by assuming a linear model to fit the data. Here, the model to fit the response is an additive model, where the effect of main factors and interactions are represented by unknown parameters. In addition, a stochastic error component is considered in order to represent the inner variability of the response. Usually, errors are assumed to be i.i.d. homoscedastic random variables with zero mean. This model requires some further assumptions in order to be applied. Some of them, such as independence among experimental units or the identical distribution, are reasonable and supported by experience. Other assumptions, such as normality of the experimental errors, are not always adequate. Generally it is possible to check the assumption of normality only after the analysis has been made, through diagnostic tools such as the $Q-Q$ plot (Daniel, 1959). Nevertheless, these tools are mainly descriptive; therefore the conclusions they may lead to are essentially subjective. If the normality of errors is not satisfied or cannot be justified, then the usual test statistics (such as the Student t test or the F test) are approximate. It is therefore worthwhile to reduce some assumptions, either to avoid the use of approximate tests or to extend the applicability of the methods applied.

Permutation tests represent the ideal instrument in the experimental design field since they do not require assumptions on the distribution of errors and, if normality can be assumed, they give results almost as powerful as their parametric counterpart. There are other reasons to use permutation tests; for instance, in the $I \times J$ replicated designs, even if data are normally distributed, the two-way ANOVA test statistics are positively correlated. This means that the inference on one factor may be influenced by other factors. There are other situations where parametric tests cannot be applied at all: In unreplicated full factorial designs, the number of observations equals the number of parameters to estimate in the model; therefore there are no degrees of freedom left to estimate the error variance. Permutation tests deal with the notion of exchangeability of the responses: The exchangeability is satisfied if the probability of the observed data is invariant with respect to random permutations of the indexes. The exchangeability of the responses is a sufficient condition to obtain an exact inference. In factorial design, the responses are generally not exchangeable since units assigned to different treatments have different expectations. Thus, either a restricted kind of permutation is needed or approximate solutions must be taken into account in order to obtain separate inferences on the main factor/interaction effects.

Chapter 5 is an introduction to ANOVA in a nonparametric view. Therefore, the general layout is introduced with minimal assumptions, with some particular care about the exchangeability of errors. Some of the solutions from the literature are introduced and discussed. The kinds of errors that may arise (individual and family wise errors) in such a context are introduced, and some preliminary methods to control them are suggested. The final part of the chapter leads with direct applications of the existing methods from the literature to practical examples.

In Chapter 6 a nonparametric solution to test for effects in replicated designs is introduced. This part is dedicated to extending the solution proposed by Pesarin (2001) and Salmaso (2003) for a 2×2 balanced replicated factorial design with n units per treatment. Since the responses are not exchangeable, the solution is based on a particular kind of permutations, named synchronized permutations. In particular, by exchanging units within the same level of a factor and by assuming the standard side conditions on the constraints, it is possible to obtain a test statistic for main factors and interactions that only depends on the effects under testing and on a combination of exchangeable errors. The proposed tests are uncorrelated with each other, and they are shown to be almost as powerful as the two-way ANOVA test statistics when errors are normally distributed. After introducing the test statistics, two algorithms are proposed to obtain Monte Carlo synchronized permutations. If we desire a post hoc comparison, simultaneous confidence intervals on all pair wise comparisons can be obtained by similarly applying synchronized permutations. The tests proposed are then compared with the classical parametric analysis.

Chapter 7 is devoted to the problem of the unreplicated full factorial design analysis. Again, the problem of exchangeability of the responses arises and, given the peculiarity of the problem, it does not seem possible to obtain exact permutation tests for all factors unless testing for the global null hypothesis that there are no treatment effects. The paired permutation test introduced by Pesarin and Salmaso (2002) is exact, but it is only applicable to the first M largest effects. A further approximate solution is then proposed. Such a solution is based on the decomposition of the total response variance under the full model and under some restricted models that are obtained in accordance with the null hypothesis under testing. The test statistic is a ratio of uncorrelated random variables, that allows us to evaluate the increase of explained variance in the full model due to the main effect under testing. The proposed test statistic allows the individual error rate to be controlled under the effect sparsity assumption. It does not control the experimental error rate, and its power is a decreasing function of the number of active effects and their sizes (the bigger the size of one effect, the bigger the noncentrality parameter in the denominator of the test statistic). To allow of control the experiment-wise error rate and in order to gain power, another version of the statistical procedure is introduced, a step-up procedure based on the comparison among noncentrality parameters of the estimates of factor effects. This test needs a calibration, which requires the central limit theorem, in order to control the experiment-wise error rate. The calibration can be obtained by either providing some critical p-values for each step of the procedure in accordance with a Bonferroni (or Bonferroni-Holm) correction or by obtaining a single critical p-value based on the distribution of the minP from simulated data under the global null hypothesis. This test is shown to be very powerful, as it can detect active factors even when there is no effect sparsity assumption (except on the smallest estimated effect, which cannot be tested). Note that a similar calibration can be provided in order to control the individual error rate at level α by choosing the critical α-quantile from the simulated null distribution of the sequential p-values. A power comparison with Loughin and Noble's test (1997) and an application from Montgomery (1991) are finally reported and discussed.

Each chapter of the book contains `R` code to develop the proposed theory. All `R` codes and related functions are available online at www.gest.unipd.it/∼salmaso/web/springerbook. This Website will be maintained and updated by the authors, also providing errata and corrigenda of the code and possible mistakes in the book.

The authors wish to thank John Kimmel of Springer-Verlag and the referees for their valuable comments and publishing suggestions. In addition, they would like to acknowledge the University of Padova and the Italian Ministry for University and Scientific and Technological Research (MIUR - PRIN 2006) for providing the financial support for the necessary research and developing part of the `R` codes.

Dario Basso, Fortunato Pesarin, Luigi Salmaso, and Aldo Solari
Padova, December 2008

corresponding author: Luigi Salmaso, email: salmaso@gest.unipd.it
book Website: www.gest.unipd.it/˜salmaso/web/springerbook.htm

This work has been supported by the Italian Ministry of University and Research (PRIN 2006133284_003) and by the University of Padova (CPDA088513). Research projects' coordinator: Prof. Luigi Salmaso.

Permutation Tests for Stochastic Ordering and ANOVA
Theory and Applications with R
Basso, D.; Pesarin, F.; Salmaso, L.; Solari, A.
2009, XIV, 218 p., Softcover
ISBN: 978-0-387-85955-2