International Statistical Review – August 2015 issue

The *International Statistical Review* is the journal of the International Statistical Institute that serves as its conduit for disseminating accessible reviews and assessments of key statistical concepts and important emerging methodologies to members of the Institute, and more generally to all professional statisticians. The articles in the August 2015 issue of IS Review reflect this aim, and should be of interest to all readers of this newsletter. The issue starts with an article and an accompanying discussion on issues related to definition of asymmetric probability distributions determined by shape parameters. This is followed by five more articles that address a very wide range of statistical issues: the use of resampling methods in simulation studies, optimal design from a Bayesian perspective, statistical theory for extreme values, foundational issues in statistical inference and the use of data pooling for statistical efficiency. All of these articles are relevant to major areas of modern statistical development and should be relevant to the professional interests of all statisticians.

On Families of Distributions with Shape Parameters by M. C. Jones
Univariate continuous distributions are one of the fundamental components on which statistical modelling, ancient and modern, frequentist and Bayesian, multi-dimensional and complex, is based. In this article, I review and compare some of the main general techniques for providing families of typically unimodal distributions on the real line with one or two, or possibly even three, shape parameters, controlling skewness and/or tailweight, in addition to their all-important location and scale parameters. One important and useful family is comprised of the 'skew-symmetric' distributions brought to prominence by Azzalini. As these are covered in considerable detail elsewhere in the literature, I focus more on their complements and competitors. Principal among these are distributions formed by transforming random variables, by what I call 'transformation of scale'—including two-piece distributions—and by probability integral transformation of non-uniform random variables. I also treat briefly the issues of multivariate extension, of distributions on subsets of the real line and of distributions on the circle. The review and comparison is not comprehensive, necessarily being selective and therefore somewhat personal.

Assessing Variability of Complex Descriptive Statistics in Monte Carlo Studies Using Resampling Methods by Dennis D. Boos and Jason A. Osborne
Good statistical practice dictates that summaries in Monte Carlo studies should always be accompanied by standard errors. Those standard errors are easy to provide for summaries that are sample means over the replications of the Monte Carlo output: for example, bias estimates, power estimates for tests and mean squared error estimates. But often more complex summaries are of interest: medians (often displayed in boxplots), sample variances, ratios of sample variances and non-normality measures such as skewness and kurtosis. In principle, standard errors for most of these latter summaries may be derived from the Delta Method, but that extra step is often a barrier for standard errors to be provided. Here, we highlight the simplicity of using the jackknife and bootstrap to compute these standard errors, even when the summaries are somewhat complicated.

Finding Bayesian Optimal Designs for Nonlinear Models: A Semidefinite Programming-Based Approach by Belmiro P. M. Duarte and Weng Kee Wong
This paper uses semidefinite programming (SDP) to construct Bayesian optimal design for nonlinear regression models. The setup here extends the formulation of the optimal designs problem as an SDP problem from linear to nonlinear models. Gaussian quadrature formulas (GQF) are used to compute the expectation in the Bayesian design criterion, such as D-, A- or E-optimality. As an illustrative example, we demonstrate the approach using the power-logistic model and compare results in the literature. Additionally, we investigate how the optimal design is impacted by different discretising schemes for the design space, different amounts of uncertainty in the parameter values, different choices of GQF and different prior distributions for the vector of model parameters, including normal priors with and without correlated components. Further applications to find Bayesian D-optimal designs with two regressors for a logistic model and a two-variable generalised linear model with a gamma distributed response are discussed, and some limitations of our approach are noted.

Extreme Value Theory and Statistics of Univariate Extremes: A Review by M. Ivette Gomes and Armelle Guillou

Statistical issues arising in modelling univariate extremes of a random sample have been successfully used in the most diverse fields, such as biometrics, finance, insurance and risk theory. Statistics of univariate extremes (SUE), the subject to be dealt with in this review paper, has recently faced a huge development, partially because rare events can have catastrophic consequences for human activities, through their impact on the natural and constructed environments. In the last decades, there has been a shift from the area of parametric SUE, based on probabilistic asymptotic results in extreme value theory, towards semi-parametric approaches. After a brief reference to Gumbel's block methodology and more recent improvements in the parametric framework, we present an overview of the developments on the estimation of parameters of extreme events and on the testing of extreme value conditions under a semi-parametric framework. We further discuss a few challenging topics in the area of SUE.

On Some Principles of Statistical Inference by Nancy Reid and David R. Cox

Statistical theory aims to provide a foundation for studying the collection and interpretation of data, a foundation that does not depend on the particular details of the substantive field in which the data are being considered. This gives a systematic way to approach new problems, and a common language for summarising results; ideally, the foundations and common language ensure that statistical aspects of one study, or of several studies on closely related phenomena, can be broadly accessible. We discuss some principles of statistical inference, to outline how these are, or could be, used to inform the interpretation of results, and to provide a greater degree of coherence for the foundations of statistics.

On Pooling of Data and Its Relative Efficiency by Jinfeng Xu and Anthony Kuk

Pooling of data is often carried out to protect privacy or to save cost, with the claimed advantage that it does not lead to much loss of efficiency. We argue that this does not give the complete picture as the estimation of different parameters is affected to different degrees by pooling. We establish a ladder of efficiency loss for estimating the mean, variance, skewness and kurtosis, and more generally multivariate joint cumulants, in powers of the pool size. The asymptotic efficiency of the pooled data non-parametric/parametric maximum likelihood estimator relative to the corresponding unpooled data estimator is reduced by a factor equal to the pool size whenever the order of the cumulant to be estimated is increased by one. The implications of this result are demonstrated in case–control genetic association studies with interactions between genes. Our findings provide a guideline for the discriminate use of data pooling in practice and the assessment of its relative efficiency. As exact maximum likelihood estimates are difficult to obtain if the pool size is large, we address briefly how to obtain computationally efficient estimates from pooled data and suggest Gaussian estimation and non-parametric maximum likelihood as two feasible methods.