Statistical Science
2020, Vol. 35, No. 1, 1
https://doi.org/10.1214/20-STS351IN

Introduction to the Special Issue

I am pleased to introduce this special issue of *Statistical Science* on the general theme of Statistics arising from Science. Historically many statistical procedures arose in response to specific scientific problems. As a result of common statistical problems in different scientific contexts (and partly as a result of scientific specialization) it has become popular to explore the structures that have been developed as mathematical, computational, or even philosophical investigations, sometimes relatively remote from applications.

The goal of this issue is to provide a set of papers that describe some of the statistical challenges presented by the science of today, informed in part by traditional models, which often require additional development. Scientific problems discussed here include science in the laboratory and science in the natural world, connected to the laboratory. The subjects reflect my (sometimes distant) interests, and I cannot pretend they represent more than a small fraction of many possibilities.

The largest single group of articles comes from molecular/cellular biology, in particular the study of single molecules or single cells, which has only recently become possible as a consequence of remarkable advances in laboratory technique. The papers of Staudt *et al.* and Agarwal, Wang and Zhang show that putting single molecule/cell data into a form that allows one to ask interesting scientific questions is in itself a substantial undertaking, and they provide a glimpse of what down stream analyses may involve. The paper of Du and Kou develops a hidden Markov change-point model to analyze single molecule behavior, while Lin *et al.* apply Markov chain Monte Carlo to explore a carefully constructed parametric model to obtain a single cell analysis of the joint behavior of chromatin accessibility and gene expression. Khanh *et al.* show how two classical models of evolution can be applied to the evolution of tumors obtained from bulk data and discuss statistical analyses to connect the evolutionary models to questions of biomedical interest.

Brentnall and Cuzick and Diggle *et al.* discuss what might be called classical biostatistical problems, and show us how one is led by scientific progress to address increasingly difficult questions. Perhaps ironically, at least for Diggle *et al.* scientific progress has created a moving target that in addition to biostatistical issues has come to involve social and ethical issues as well.

Guo, Zhang and Zhang suggest an alternative to independent component analysis for focusing attention on important features of multiple time series and suggest an application to functional magnetic resonance imaging.

Stein has produced a survey of problems in climate science, which involve the analysis of increasingly large and complex sets of scientific data, and he proposes that statisticians regard part of the enterprise to be the analysis of simulations obtained from climate science's fiendlishly complex models involving uncertainties in large numbers of parameters. (With a vivid imagination I can relate the issues raised here to my summer between college and graduate school, when I worked on the design of simulations—computationally expensive at the time—for analyzing the differential equations describing the trajectory of a hand-held rocket subject to variability in a large number of parameters.)

Wang has provided us with an elegant introduction to quantum mechanics, where a focus on "spin" allows us a glimpse into quantum computation, information, and cryptography based on finite dimensional linear algebra that neatly sidesteps a discussion of the "physical" connection to classical (Hamiltonian and Lagrangian) mechanics. The foundations involve a neat relation to Markov chains. The assumption of quantum mechanics that changes of state are governed by a unitary operator is analogous to the Chapman–Kolmogorov equation, and differentiation of a unitary change of state to obtain the Schrödinger equation similarly turns the Chapman–Kolmogorov equation into a Kolmogorov differential equation, which like the Schrödinger equation is formally solved by a matrix exponential involving the Hamiltonian/Generator. (Of course, this analogy goes only so far, although it is only one more step from quantum mechanical Feynman integrals to the Feynman–Kac representation for occupation times of functionals of Markov processes.)

I want to express my thanks to the contributors to this volume for their efforts to communicate the excitement they find in research at the interface of science and statistics, which I hope and expect will continue to stimulate fundamental contributions to the development of our field.