Optimal Bayesian Analysis of A/B Tests (Randomized Controlled Trials) in Data Science at Big-Data Scale

David Draper

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

1 Mar 2018

When coping with uncertainty, You typically begin with a problem $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$, in which $\mathbb{Q}$ lists the questions of principal interest to be answered and $\mathbb{C}$ summarizes the real-world context in which those questions arise. $\mathbb{Q}$ and $\mathbb{C}$ together define $(\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})$, in which $\boldsymbol{\theta}$ (which may be infinite-dimensional) is the unknown of principal interest, $\boldsymbol{D}$ summarizes Your data resources for decreasing Your uncertainty about $\boldsymbol{\theta}$, and $\mathcal{B}$ is a finite (ideally exhaustive) set of (true/false) propositions summarizing, and all rendered true by, context $\mathbb{C}$.

The Bayesian paradigm provides one way to arrive at logically-internally-consistent inferences about $\boldsymbol{\theta}$, predictions for new data $\boldsymbol{D}^*$, and decisions under uncertainty (either through Bayesian decision theory or Bayesian game theory). In this paradigm it's necessary to build a stochastic model $\mathcal{M}$ that relates knowns ($\boldsymbol{D}$ and $\mathcal{B}$) to unknowns ($\boldsymbol{\theta}$). For inference and prediction, on which I focus in this talk, the model takes the form $\mathcal{M} = \{p(\boldsymbol{\theta} \mid \mathcal{B}), p(\boldsymbol{D} \mid \boldsymbol{\theta}, \mathcal{B})\}$, in which the (prior) distribution $p(\boldsymbol{\theta} \mid \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ *external* to Your dataset $\boldsymbol{D}$ and the (sampling) distribution $p(\boldsymbol{D} \mid \boldsymbol{\theta}, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ *internal* to $\boldsymbol{D}$, when converted to Your likelihood distribution $\ell(\boldsymbol{\theta} \mid \boldsymbol{D}, \mathcal{B}) \propto p(\boldsymbol{D} \mid \boldsymbol{\theta}, \mathcal{B})$.

The fundamental problem of applied statistics is that *the mapping from $\mathbb{P}$ to $\mathcal{M}$ is often not unique*: You have *basic uncertainty* about $\boldsymbol{\theta}$, but often You also have *model uncertainty* about how to specify Your uncertainty about $\boldsymbol{\theta}$. It turns out, however, that there are situations in which problem context $\mathbb{C}$ implies a unique choice of $p(\boldsymbol{\theta} \mid \mathcal{B})$ and/or $\ell(\boldsymbol{\theta} \mid \boldsymbol{D}, \mathcal{B})$; let's agree to say that *optimal Bayesian model specification* has occurred in such situations, leading to *optimal Bayesian analysis* via Bayes's Theorem and its corollaries.

In this talk I'll identify an important class of problems arising in the analysis of randomized controlled trials (typically called *A/B tests* in Data Science) in which optimal Bayesian analysis is made possible by the use of Bayesian non-parametric modeling, and I'll illustrate this type of analysis with an A/B test at Big-Data scale (involving about 22 million observations). Along the way I'll demonstrate that *the frequentist bootstrap is actually a Bayesian non-parametric method in disguise.*