36-724 Spring 2006: Introduction

Brian Junker

January 16, 2006

• Approaches To Data Analysis
• Applied Bayesian methods
• Applied Computational methods
• Why Start with Bayes?
• Some “Practical” Reasons for Bayes
• Some Examples of Naturally Bayesian Situations
Approaches To Data Analysis

• Point of view: \( P(X) = \int P(X, \theta) \, d\theta \)
  – \( P(X, \theta) = P[X|\theta]f(\theta); \) \( P[X|\theta] \) is likelihood
  – \( P(X, \theta) = P[\theta|X]P(X); \) \( P[\theta|X] \) is posterior

All of our inferences about \( P[X|\theta] \) or \( P[\theta|X] \), are ultimately functions of \( X \), and inherit their stochastic properties from \( P(X) \).

• Strategies for producing inferences
  – What is a good thing to optimize?
  – What is a good heuristic?

• Discovering structure in data
  – Multiple analyses and multiple views
  – Simple plots may reveal the unexpected (plots > tables!)
Data Analysis - Roles of Models

• Some common processes
  – data collection
  – description of data set (EDA) (“exploratory”)
  – inference within the context of a family of models (estimation, prediction, decision, ...)
  – model selection (“confirmatory”)

• Interaction between exploration and modeling
  – Need for a model even in EDA
  – Need for EDA to help interpret model results
    * simulation from model
    * predictions from model
    * exploring ”response surfaces”
**Roles of Computation**

- Data reduction and summarization
  - dimension reduction
  - EDF functionals
  - graphing
  - etc.

- Computation with respect to a “general” model
  - MC hypothesis testing
  - bootstrapping/jackknife

- Computation for estimation/prediction/etc. within a parametric family

- Computation without a model
Applied Bayesian methods

- First half of the semester; Gelman text.
- Develop a “modeling toolbox” that can be used to build models for a variety of situations
- Develop a flexible set of computational tools (especially MCMC) for fitting and inference
- Explore how “substantive”/scientific information and computational feasibility influence model building
- Explore fit, sensitivity to assumptions, impact on the substantive/scientific question
Applied Computational methods

• Second half of the semester; Hastie et al. text.
• Survey a variety of approaches to discovering and modeling structure.
• Goals can be *description*, *prediction*, or just *getting a good look at the data*. Often we want to perform some sort of *dimension reduction*.
  – Select a criterion function (often specifying a *bias/variance* or *smoothness/accuracy* tradeoff), and optimize it for our goal.
  – How well do we do?
    * Statisticians usually try to specify \( \text{Data} = \text{Model} + \text{Error} \) carefully and then use the Error model to gauge uncertainty in inference.
    * Iterative sampling & simulation approaches are needed when the Error model cannot be easily specified.
• Look at links between traditional topics and modern computational ones
  – Linear models (smoothing, classification)
  – Bayes & mixture modeling (clustering)
  – Principal components (dimension reduction)
Why Start with Bayes?

- Exchangeability
- Minimize risk
- Consistent framework for incorporating substantive/scientific information and constraints in inference (subjective or not)
Exchangeability

Random variables $X_1, X_2, X_3, \ldots$ are exchangeable if any subset of $k$ of them have the same $k$-variate distribution. E.g.

- $X_1, X_2, \ldots$ all have the same univariate distribution
- $(X_1, X_2), (X_3, X_5)$, and generally $(X_j, X_k)$ all have the same bivariate distribution.
- For each $k$, all $k$-tuples $(X_{j_1}, \ldots X_{j_k})$ have the same $k$-variate distribution.

**Theorem (DeFinetti).** If the sequence $X_1, X_2, X_3, \ldots$ is exchangeable, then (two-stage experiment. . . )

$$f(x_{j_1}, \ldots x_{j_k}) = \int \prod_{m=1}^{k} f(x_{j_m} | \theta) d\pi(\theta) \quad (*)$$

Conversely if (*) holds, then $X_1, X_2, \ldots$ are exchangeable.
Risk

All inferences are, ultimately, functions of the data $X$, $\delta(X)$. In decision theory we measure the quality of an inference by

- Loss $L(\theta, \delta(X))$  E.g.: $(\theta - \delta(X))^2$, $1_{\theta \neq \delta(X)}$, etc.
- Risk $R(\theta, \delta) = E[L(\theta, \delta(X))|\theta]$
- Average Risk or Bayes Risk

$$A(\delta) = \int R(\theta, \delta)f(\theta)d\theta = \int \int L(\theta, \delta(X))P(X, \theta)dXd\theta$$

**Theorem (Bayes Rules).** Define the *posterior risk* for inference $a$, as:

$$r(X, a) = \int L(\theta, a)P(\theta|X)d\theta$$

Then the decision rule $\delta(X)$ defined by minimizing posterior risk

$$\delta(X) = \underset{a}{\text{argmin}} r(X, a)$$

also minimizes the Average (or Bayes) risk $A(\delta)$. 
Some “Practical” Reasons for Bayes

- Unified framework for dealing with parameters, latent variables and data (both *observed* data and *missing* data).
- Rich framework for thinking about error/uncertainty.
  - Allows uncertainty in “incidental” parameter estimation to “feed forward” into prediction and estimation questions of interest. Avoid (or understand) overly-optimistic estimation/prediction.
  - Characterization of uncertainty is richer (full posterior distributions, not just first and second moments) and works in small samples (non-asymptotic).
- Makes all assumptions available for study: tradeoff between data information and prior information; sensitivity analysis
- Unify estimation and model selection.
- Simple framework for learning/updating (yesterday’s posterior is today’s prior).
Some Examples of Naturally Bayesian Situations

• Exchangeability: Excess variation in simple problems (e.g. extra-binomial variation)

• Partial Exchangeability and positive association:
  – Clustered sampling in survey sampling
  – Nested observational data
  – Growth curve modeling with individual differences

• Measurement models for latent variables

• Many other problems that were not conceived of as “Bayesian” benefit from the Bayesian mathematical machinery:
  – Factor analysis, latent class, and latent variable modeling
  – Mixed effects (generalized) linear models
  – Missing data problems
  – etc.

• Many problems naturally lead to “hierarchical Bayesian modeling”.