The Power of Bayes: An Old Idea Revisited

C. Shane Reese, Department of Statistics, BYU

Outline

- Historical Perspective
- Overview of Statistical Framework
- Example: Computer Model Incorporation
- Example: Supercomputer Reliability
- Concluding Ideas for Computer Science

Historical Tour

- Bayes (1702-1761) was a Nonconformist minister with an interest in mathematics
- Papers:
  Divine Providence and Government Is the Happiness of His Creatures (1731)
  An Introduction to the Doctrine of Fluxions, and a Defense of the Analyst (1736)
  Essay Towards Solving a Problem in the Doctrine of Chances (1764)
- Richard Price, Laplace (1781).
- 1950's Bayes: Savage, deFinetti, Lindley.
- 1990's Bayes: Gelfand, Berger, Smith, Berry.

Definitions

- Problem: Unknown population parameter (θ) must be estimated.

  1. Simple Example:

     θ =Probability that a selected person from the group can hit a target on the board with a ball from the back of the room.

  2. Real Example 1:

     θ = Optimal steady-state thermodynamic operating point of a fluidized bed coating process for food coating.

  3. Real Example 2:

     θ = Probability that a supercomputer constructed of 48 nodes in "parallel" can complete a 6 hour computer run with no failures.

Definitions

- Step 1 of either formulation is to pose a statistical (or probability) model for the random variable which represents the phenomenon.

  1. Simple Example:
     a reasonable choice for $f(y|\theta)$ (the sampling density or likelihood function) would be that the number of successful hits (Y) on the target would follow a *binomial* distribution with the n throws and the probability of any one throw being a "hit" called θ
  2. Real Example 1:
     Topic of next part of the talk.
  3. Real Example 2:
     Topic later in the talk.

Definitions (cont)

- Classical Data Analysis
  1. all pertinent information enters the problem through the likelihood function in the form of data $(Y_1, \ldots, Y_n)$

$$f(y_1, \ldots, y_n | \theta) = \prod_{i=1}^{n} f(y_i | \theta)$$

  2. objective in nature
  3. software packages all have this capability
  4. maximum likelihood, unbiased estimation, etc.
  5. confidence intervals, difficult interpretation

Definitions (cont)

- Bayesian Data Analysis
  1. data (enters through the likelihood function *as well as* allowance of other information

$$p(\theta|y_1,\ldots,y_n) = \text{constant} \times \prod_{i=1}^{n} f(y_i|\theta) \times \pi(\theta).$$

  2. reads: the *posterior distribution* [$p(\theta|y_1,\ldots,y_n)$] is a constant multiplied by the likelihood [$\prod_{i=1}^{n} f(y_i|\theta)$] muliplied by the *prior distribution* [$\pi(\theta)$]
  3. prior distribution: before any data collection, the view of the parameter
  4. posterior distribution: in light of the data our updated view of the parameter

Prior Distributions

1. can come from expert opinion, historical studies, previous research, or general knowledge of a situation
2. there exists a “flat prior” or “noninformative” which represents a state of ignorance.
3. Attempting to capture physical theories, expert opinion, etc. in a probability distribution is often difficult.
4. See Real Example 1 for an interesting example.

Definitions (cont)

- Bayesian Data Analysis (cont)
  1. inherently subjective (prior is controversial)
  2. few software packages have this capability
  3. result is a probability distribution
  4. *credible intervals* use the language that everyone uses anyway. (Probability that θ is in the interval is .95.
  5. see examples for demonstration.

Comparison

Bayesian	Classical
subjective	objective
little software	abundant software
direct interpretation	difficult interpretation
decision theory	hypothesis testing

Real Example 1

- Computer Experiments
- Difference From Field Trials (Experimental Data)
- Examples:
  1. Modern Industrial R&D
  2. Climate Modelling
  3. Biological/Nuclear Weapons Systems
  4. Modern Finance

Real Example 1

- Food coating example
- Observed Data: $\mathbf{X}$ = temperature of fluid, velocity of fluid, etc. Y = thermodynamic operating point
- Computer model: Computational Fluid Dynamics Model combined with other physics theories to “predict” temperature
- Initial Design: Random LHS Design.

Data Structure

- Field observations:

$$y_f = (y_{f_1}, \cdots, y_{f_n})'$$

where

$$y_{f_i} = \eta(x_i, \theta) + \delta(x_i) + \varepsilon_i, \quad i = 1, \cdots, n.$$

- Computer code output:

$$y_c = (y_{c_1}, \cdots, y_{c_m})'$$

where

$$y_{c_j} = \eta(x_j^*, t_j), \quad j = 1, \cdots, m.$$

- Full set of data:

$$d^T = (y_c^T, y_f^T) = (y_{c_1}, \cdots, y_{c_m}, y_{f_1}, \cdots, y_{f_n}).$$

Likelihood & Priors

1. Model:

   $\eta(\cdot,\cdot) \sim N(\mu_\eta, c_1\{(\cdot,\cdot),(\cdot,\cdot)\})$ with

   $$c_1((x^*,t),(x^{*\prime},t')) = \sigma_\eta^2 \cdot exp\left\{-\sum_{k=1}^{p}\beta_k^\eta(x_k^* - x_k^{*\prime})^2 - \sum_{k'=1}^{q}\beta_{p+k'}^\eta(t_k - t_k')^2\right\}$$

   $\delta(\cdot) \sim N(\mu_\delta, c_2(\cdot,\cdot))$ with

   $$c_2(x,x') = \sigma_\delta^2 \cdot exp\left\{-\sum_{k=1}^{p}\beta_k^\delta(x_k - x_k')^2\right\}$$

   $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$.

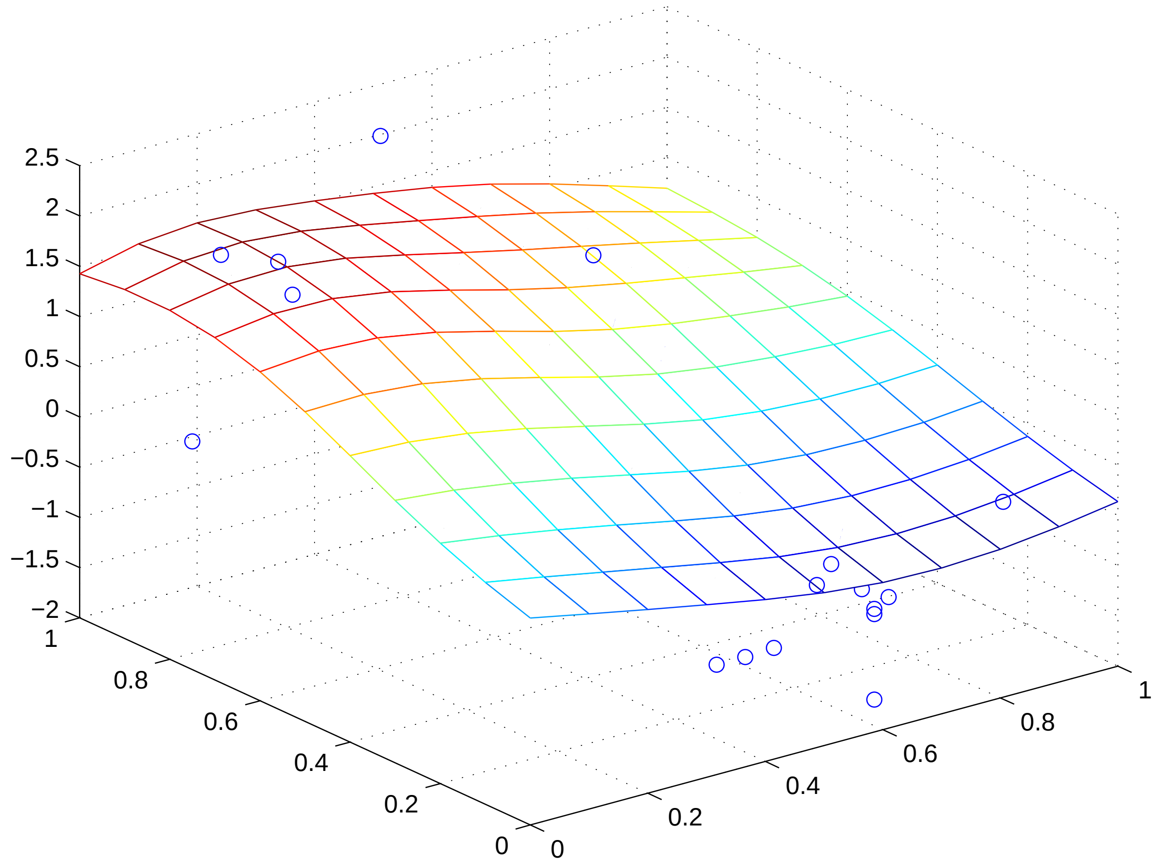
2. Parameters:

   (a) $\Omega = \{\mu, \theta, \phi, \sigma_\varepsilon^2\}$

   (b) $\mu = (\mu_\eta, \mu_\delta)^T$

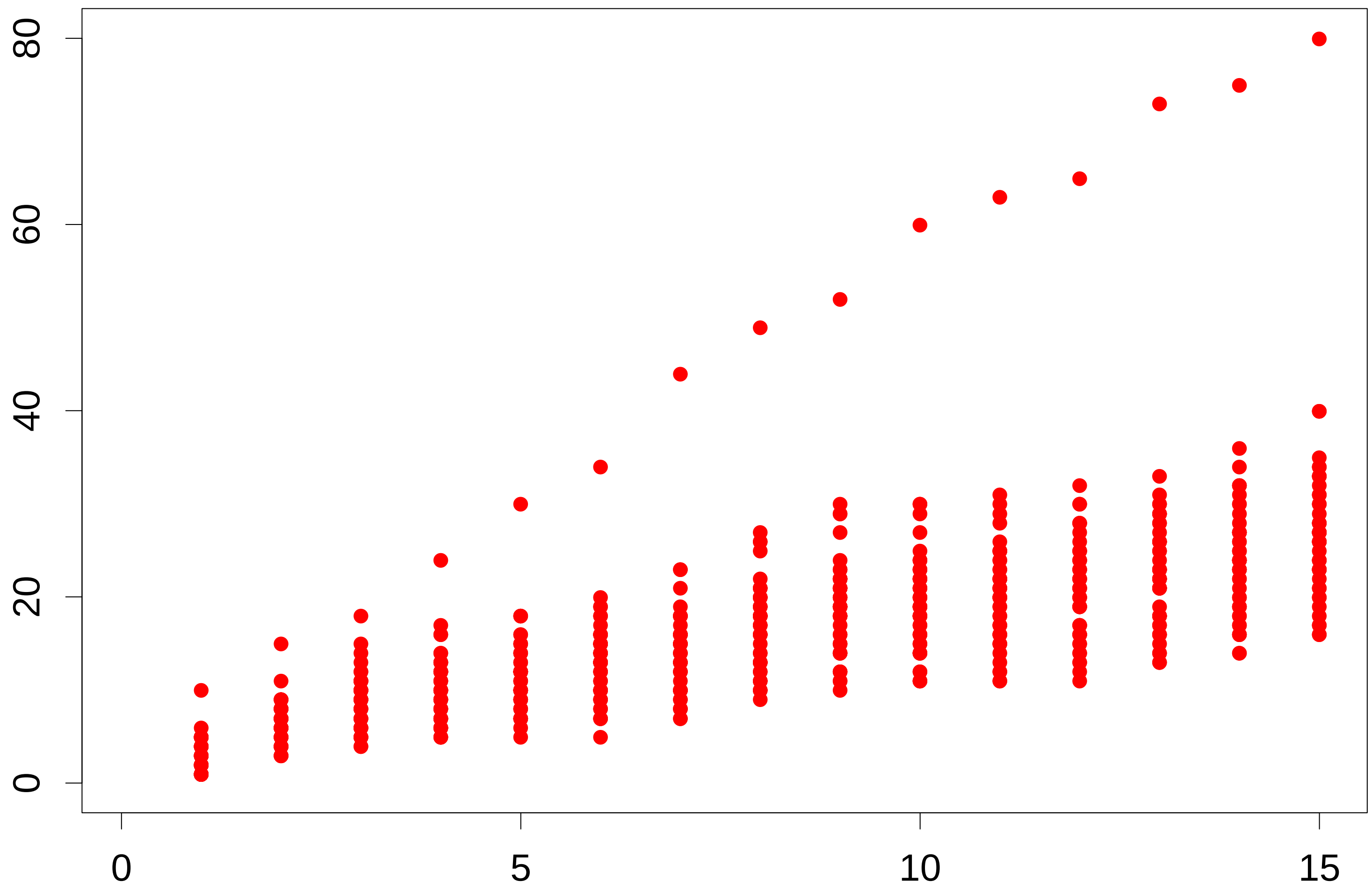
   (c) $\phi = \{\sigma_\eta^2, \sigma_\delta^2, \beta^\eta, \beta^\delta\}$

"Prior" Surface with Data



Real Example 2

- Los Alamos National Laboratory Supercomputer: Blue Mountain
- Weapons development and maintenance: Stockpile Stewardship
- Problems in late 1999
- Goal: Calculate the probability of running a 6 hour job with no interruptions on a variety of architectures.


0
20
40
60
80
0
5
10
15

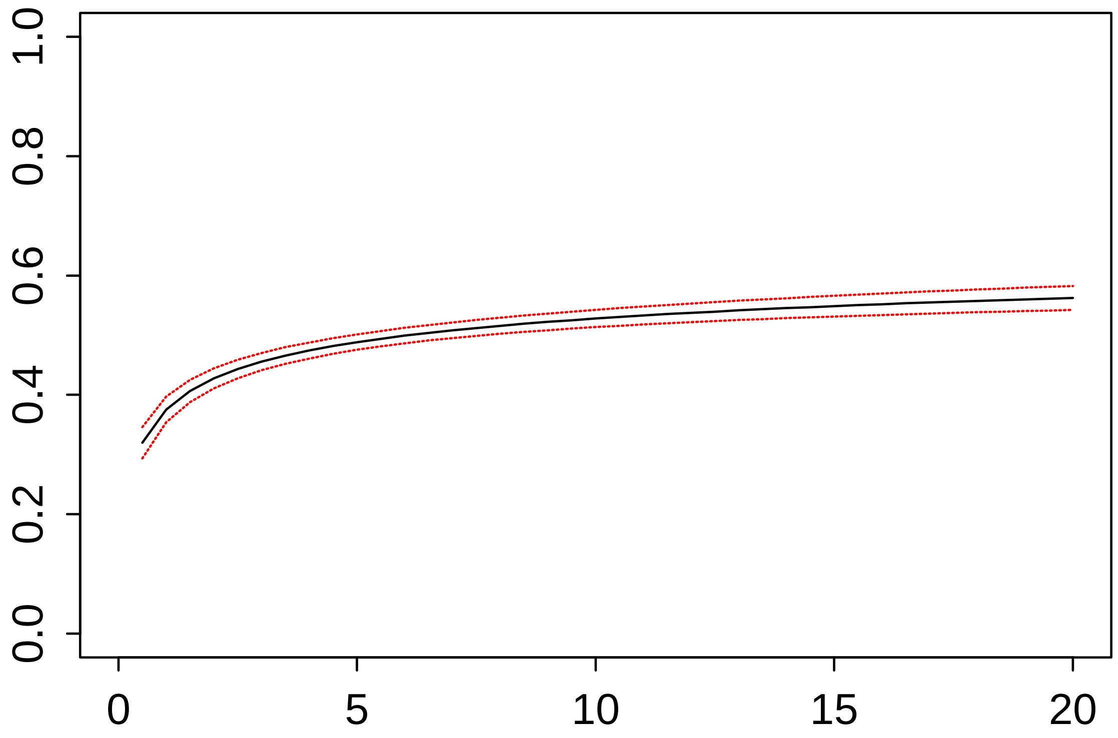
Let $\underline{x} = (x_{i1}, x_{i2}, \ldots, x_{iM})$ be the vector of failure counts for the ith SMP. Then, the sampling distribution for the failure counts for the ith SMP in time interval j is

$$x_{ij}|\phi,\eta \overset{indep}{\sim} \text{Poisson}\left(\left(\frac{tj}{\eta}\right)^{\phi} - \left(\frac{t(j-1)}{\eta}\right)^{\phi}\right) \text{ for } j = 1,\ldots,M,$$

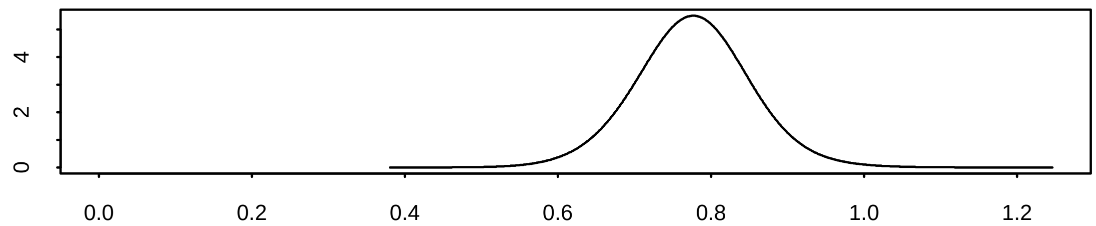
and $\underline{x}$ has probability mass function

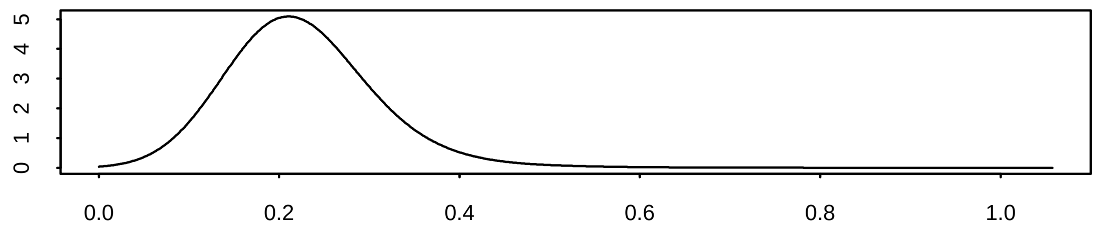
$$p(\underline{x}|\phi,\eta) = \prod_{j=1}^{M}\left[\frac{\left\{\left(\frac{tj}{\eta}\right)^{\phi} - \left(\frac{t(j-1)}{\eta}\right)^{\phi}\right\}^{x_{ij}}}{x_{ij}!}\right]\exp\left\{-\left(\frac{Mt}{\eta}\right)^{\phi}\right\}.$$

Posterior Distribution of θ



Predictive of *New* θ





Conclusions

- Science is subjective (what about the choice of a likelihood?)
- Bayes uses all available information
- Makes interpretation easier
- BAD NEWS: Computation is not generally available in standard statistical software packages.
- GOOD NEWS: They are possible (and practical) with advanced computational procedures