UCONN
UNIVERSITY OF CONNECTICUT

Bayesian Computation

CAS Centennial Celebration and Annual Meeting
New York, NY
November 10, 2014

Brian M. Hartman, PhD ASA
Assistant Professor of Actuarial Science
University of Connecticut

CAS Antitrust Notice

UCONN

General differences between Bayesian and Frequntist statistics

- Frequentist
  - Parameters are fixed but unknown
  - Probability based on repeated samples (sampling distribution)
  - Uses asymptotic approximations
- Bayesian
  - Parameters are random variables
  - Subjective prior combined with data
  - No asymptotic approximations

UCONN

Confidence Intervals

- What you have to say: “We are 95% confident that the population mean is between X and Y” (frequentist)
- What you want to say: “There is a 95% probability that the population mean is between X and Y” (Bayesian)

UCONN

Hypothesis Testing

- P-value: "Given the null hypothesis is true, the p-value is the probability that you obtain a sample statistic as extreme or more extreme than the observed statistic." (frequentist)
- What you want to find: "What is the probability that my null hypothesis is true?" (Bayesian)

UCONN

Any other reasons I should use Bayesian methods?

- Ability to incorporate expert opinion and prior knowledge in a structured way.
- Ability to easily find any quantities of interest (e.g. 95% interval for the mean or variance of a gamma distributed loss)
- Easier to set up and estimate complicated models

UCONN

Why doesn't everyone learn Bayesian statistics first?

- Thomas Bayes died in 1761
- Bayes' Theorem: $\Pr(\theta|Y) = \frac{\Pr(Y|\theta)\Pr(\theta)}{\Pr(Y)}$
- Seems simple enough, but the difficulty lies in the denominator, $\Pr(Y)$
- $P(Y) = \int_{\Theta} \Pr(Y|\theta)\Pr(\theta)\,d\theta$
- There are many examples where this integral can be solved analytically

UCONN

Why doesn't everyone learn Bayesian statistics first?

- But that is only using one parameter, with two

$$\int_{\Theta_1}\int_{\Theta_2} \Pr(Y|\theta_1,\theta_2)\Pr(\theta_1,\theta_2)\,d\theta_2\,d\theta_1$$

- Or more generally . . .

$$\int_{\Theta_1}\cdots\int_{\Theta_k} \Pr(Y|\theta_1,\dots,\theta_k)\Pr(\theta_1,\dots,\theta_k)\,d\theta_k\cdots d\theta_1$$

- No matter how much you love math, you are not going to solve this analytically (outside of a few restrictive examples).

UCONN

Why doesn't everyone learn Bayesian statistics first?

- To estimate $\Pr(Y)$ numerically requires:
  - Methodology (MCMC, much work in late 1980s through 1990s)
  - Computing power
- That is why frequentist statistics were preferred in essentially every practical application before 1990.
- That is a lot of history to fight against.

UCONN

Bayesian Computing

- Now there are various pieces of software which make writing and fitting Bayesian models much simpler.
  - WinBUGS
  - JAGS
  - STAN
- Go ahead and try it. I think you will like it.

UCONN

What about the subjectivity?

- "Bayesian methods are not scientific because of the subjective prior. Frequentist methods remove that bias."
- Let me respond to that assertion with an example
  - Suppose in 12 independent tosses of a coin, I observe 9 heads.
  - I wish to test the following hypotheses
    - $H_o: \theta = 0.5$
    - $H_a: \theta > 0.5$
  - θ is the true probability of a head.

Source: Lindley, D. V. and Phillips, L. D. (1976) Inference for a Bernoulli Process (a Bayesian view). *Amer. Statist.*, ***30***, *112-119*

UCONN

What about the subjectivity?

Knowing only that information, there are two possible sampling distributions.

1. Binomial, $n = 12$ fixed beforehand

$$L_1(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{12}{9}\theta^9(1-\theta)^3$$

2. Negative binomial, flip until third tail

$$L_2(\theta) = \binom{r+x-1}{x}\theta^x(1-\theta)^r = \binom{11}{9}\theta^9(1-\theta)^3$$

Source: Lindley, D. V. and Phillips, L. D. (1976) Inference for a Bernoulli Process (a Bayesian view). *Amer. Statist.*, ***30***, *112-119*

UCONN

What about the subjectivity?

And our sample will give two different p-values

1. Binomial

$$\alpha_1 = \Pr_{\theta=0.5}(X \geq 9) = \sum_{j=9}^{12} \binom{12}{j} \theta^j (1-\theta)^{12-j} = 0.075 > 0.05$$

2. Negative binomial

$$\alpha_1 = \Pr_{\theta=0.5}(X \geq 9) = \sum_{j=9}^{\infty} \binom{2+j}{j} \theta^j (1-\theta)^3 = 0.0325 < 0.05$$

Source: Lindley, D. V. and Phillips, L. D. (1976) Inference for a Bernoulli Process (a Bayesian view). *Amer. Statist.*, ***30**, 112-119*

UCONN

What about the subjectivity?

- Only the results should be relevant, not how the experiment is monitored
- This goes back to the definition of the p-value, "observations more extreme" are unobserved.
- Not only is there subjectivity in frequentist statistics as well, it gives inferential weight to unobserved samples

Source: Lindley, D. V. and Phillips, L. D. (1976) Inference for a Bernoulli Process (a Bayesian view). *Amer. Statist.*, ***30**, 112-119*

UCONN

Should we always use Bayesian Methods?

- I want to say "YES!!"
- But no, Bayesian methods
  – are more computationally intensive (sometimes impossibly)
  – require tests for prior sensitivity/robustness
  – Difficult to confirm Markov Chain convergence

UCONN

What else is in the chapter?

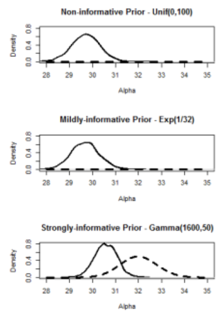
- Basic Computational Methods
  – Gibbs
  – Metropolis-Hastings
  – Convergence metrics
- Prior Distributions
  – Prior elicitation
  – Noninformative priors
  – Prior sensitivity
- R and WinBUGS code for examples
- Many references for how to use Bayesian methods in a wide variety of actuarial applications

Prior Sensitivity Example

Auto Claim Severity Data

$y \sim \text{Gamma}(\alpha, \beta)$
$\beta \sim \text{Unif}(0, 1000)$

Prior distributions:
$\alpha \sim \text{Unif}(0, 1000)$
$\alpha \sim \text{Exp}(1/32)$
$\alpha \sim \text{Gamma}(1600, 50)$



UCONN

UCONN
UNIVERSITY OF CONNECTICUT

Bayesian Computation

CAS Centennial Celebration and Annual Meeting
New York, NY
November 10, 2014

Brian M. Hartman, PhD ASA
Assistant Professor of Actuarial Science
University of Connecticut