Bayesian estimation of parameters: Advantages and Practical Examples

CAS Ratemaking Seminar
Session SPE-2
March 13, 2006
Stuart Klugman, Drake University

Agenda

- **A brief history of Bayes' Theorem**
- **Why**
- **How – Theory**
- **How – Practice**
- **COTOR challenge example**

Who first proved Bayes' Theorem?

- **From "Who Discovered Bayes's Theorem? By Stephen Stigler (*American Statistician*, November 1983)**
- **The posterior odds favor Nicholas Saunderson 3:1 over Thomas Bayes.**

Who published Bayes' Theorem?

- **After his death, Bayes willed some money and his papers to Richard Price, who arranged to have the Theorem published by the Royal Society in 1764.**
- **Richard Price later become a consultant to the Equitable Life Insurance Society and published "Observations on Reversionary Payments." His nephew, William Morgan was the first actuary in both name and title (*From Actuarius to Actuary*, Robert Mitchell, SOA, 1974).**

Why Bayes? – The problem

- **Today's problem**
  - **A random sample from some probability distribution.**
  - **The name of the distribution is known, but not its parameters.**
  - **Three goals are:**
  - **Estimate the parameters and then a quantity of interest, such as a layer cost.**
  - **Place a confidence interval on the estimate.**
  - **Determine a prediction interval for the next observation.**

Confidence interval vs. prediction interval

- **A confidence interval places bounds on the expected value. It gives the accuracy of a pure premium calculation, reflecting estimation error.**
- **A prediction interval places bounds on the payment from the next policy sold. The mean is the same as the confidence interval, but process error is incorporated.**

Why Bayes? – The frequentist solution

- **Use an optimization technique to estimate the parameters.**
- **Use asymptotic theory to understand the variances.**
- **Assume a normal distribution.**
- **The approximations in the last two items may be too crude. They also are not always easy to obtain.**

Simple example

- **Twenty simulated observations from a lognormal(7,1.5) distribution.**
- **Goal is to estimate the two parameters and the mean and then obtain confidence and prediction intervals.**
- **The lognormal distribution is among the easiest to work with because the information matrix is easy to obtain.**
- **The frequentist formulas are in *Loss Models*, 2nd ed., 353-358.**

Frequentist numbers

- **The parameter estimates are μ = 7.301 and σ = 1.624.**
- **The estimate of the mean is 5,537.**
- **A 90% normal based confidence interval is 1,085 to 9,989.**
- **A 90% normal based prediction interval is -27,562 to 38,636.**
- **[For this problem we would know that many of the shortcomings could be solved by working with ln(x)].**

Why Bayes?

- **Does not deal with what might have been, only looks at what was.**
- **Always stays within the confines of the problem (the intervals will always stay positive).**
- **Requires no special maneuvers, all problems are solved the same way.**

So why not Bayes?

- **Many people are uncomfortable with a prior distribution.**
- **The mathematics of the exact solution can be challenging. For example, a direct solution of a problem with three parameters requires triple integrals of an unpleasant function.**

Solution

- **Use a prior distribution that implies as little prior knowledge as possible.**
- **Either solve the computational problem by making it discrete rather than continuous, so sums become integrals, or use special software.**
- **I will illustrate both approaches.**

Bayesian estimation

- Let θ be the vector of unknown parameters.
- Let $f(x|\theta)$ be the known distribution.
- Let $x_1,\ldots,x_n$ be *n* independent observations from that distribution.
- Let $\pi(\theta)$ be the prior distribution on the parameters.

Posterior distribution

$$\pi(\theta | x_1,\ldots,x_n) = \frac{f(x_1|\theta)\cdots f(x_n|\theta)\pi(\theta)}{\int f(x_1|\theta)\cdots f(x_n|\theta)\pi(\theta)d\theta}$$

where the integral is replaced by a sum if the parameters have a discrete distribution.

Estimating a function of the parameters

- **Suppose our goal is to estimate $g(\theta)$, a function of the parameters.**
- **We need the posterior distribution of that function. But just attach the posterior probabilities to each value. The Bayes estimate is:**

$$\widehat{g(\theta)} = \int g(\theta)\pi(\theta \mid x_1, \ldots, x_n)\,d\theta$$

- **The *p*th percentile solves (using the posterior distribution).**

$$p = \Pr[g(\theta) \le x \mid x_1, \ldots, x_n]$$

Predictive distribution

$$f(y \mid x_1, \ldots, x_n) = \int f(y \mid \theta)\pi(\theta \mid x_1, \ldots, x_n)\,d\theta$$

- **The mean is the best guess at the next observation while percentiles provide the prediction interval.**

Lognormal example-discrete approximation

- **The prior is a simple discrete uniform distribution. For μ it goes from 5.5 to 11.1 by 0.2 and for σ it goes from 1.2 to 2.7 by 0.05. There are 899 points in this prior distribution.**
- **The spreadsheet has the calculations.**
- **The table on the next slide compares the values of the two methods so far.**

Lognormal values

	Frequentist	Discrete Bayes
$\hat{\mu}$	7.301	7.301
$\hat{\sigma}$	1.624	1.777
Mean	5,537	9,932
Confidence Interval	1,085 to 9,989	2,922 to 26,937
Prediction Interval	-27,562 to 38,636	72 to 30,423

Continuous priors

- **With continuous priors the sums become integrals.**
- **For a long time there was no easy way to do the calculations.**
- **Now there is Markov Chain Monte Carlo.**
- **The essence is on the next slide.**

MCMC

- **The goal is to simulate observations from the posterior distribution.**
- **These simulated values become the posterior distribution.**
- **The simulation is accomplished by using a sequence of conditional distributions. That is, simulate a value of one unknown parameter by conditioning not only on the data but also on the other parameters.**

WinBUGS

- **This is a free program that performs MCMC analysis.**
- **You can write code or have code generated from a graphic representation of the model.**
- **For the lognormal model I have selected priors with huge variances. They are normal(0, $1{,}000^2$) for μ and gamma(0.001, 1,000) for $1/\sigma^2$.**
- **WinBUGS can also generate the predictive distribution and the posterior distribution of functions of the parameters.**

WinBUGS code

```
model;
{ mu ~ dnorm( 0.0,1.0E-6) <sets prior on mu>
   sigma ~ dgamma(0.001,0.001) <sets prior on the
   reciprocal of the variance>
   for( i in 1 : 20 ) {x[i] ~ dnorm(mu,sigma)}
   for( i in 1 : 20 ) {x[i] <- log(y[i])} <The
   observations are from the lognormal distribution>
s <- 1/sqrt(sigma) <defines s as the std dev>
m <- exp(mu+s*s/2) <defines the mean>
p ~ dnorm(mu,sigma) <these two set the predictive
   value>
ep <- exp(p)}
```

WinBUGS results

	Frequentist	Discrete Bayes	WinBUGS
$\hat{\mu}$	7.301	7.301	7.301
$\hat{\sigma}$	1.624	1.777	1.735
Mean	5,537	9,932	10,330
Confidence Interval	1,085 to 9,989	2,922 to 26,937	2,806 to 24,330
Prediction Interval	-27,562 to 38,636	72 to 30,423	77 to 28,230

COTOR Challenge

- Round 3 offered the following problem.
- Data from a heavy-tailed distribution has been collected over 7 years; 70 observations each year.
- The distribution type does not change over time, nor do non-scale parameters.
- The scale parameter changes according to inflation.
- Determine point estimates and CI and PIs for 500x500 in year 8.

Disclaimer

- **I picked the model, a 70-30 mixture of Pareto and Exponential along with a random process to generate inflation rates.**
- **The 490 observations were simulated from that model.**
- **My analysis relies somewhat on this inside knowledge.**

A strategy

- **Use traditional frequentist techniques to pick the winning model. This tends to flow better than Bayesian approaches. My choice is the Schwarz Bayesian Criterion.**
- **Use a Bayesian analysis to get the requested estimates.**
- **This approach in one form or another was adopted by many of the COTOR participants.**

Model selection

Name	lnL	Params	SBC	rank
logn-geom	-5533.06	3	-5542.35	4
Pareto-quad	-5527.71	4	-5540.1	2
Pareto-geom	-5527.84	3	-5537.13	1
logn-exp-geom	-5527.05	5	-5542.53	5
Pareto-exp-quad	-5526.02	6	-5544.6	6
Pareto-exp-geom	-5525.99	5	-5541.48	3

Geom = geometric trend for inflation

Quad = quadratic trend for inflation

Winning model

- **A single Pareto distribution.**
- **The parameters are 1.071 and 6,232.**
- **The inflation rate is 0.1697 and thus the scale parameter in year *i* is 6232exp(.1679**i*).**
- **A 20x20x20 discrete Bayes analysis produced 1.071, 6,427, and 0.1652.**
- **A WinBugs analysis with vague priors produced 1.069, 6,228, and 0.1722.**

Layer costs

- For a Pareto distribution, the layer cost is

$$\frac{\theta}{\alpha-1}\left[\left(\frac{\theta}{\theta+500,000}\right)^{\alpha-1}-\left(\frac{\theta}{\theta+1,000,000}\right)^{\alpha-1}\right]$$

- The (discrete) posterior mean is 12,970 and the 5th and 95th percentiles are 8,234 and 18,846. This forms the confidence interval for the expected cost of the layer.
- The true answer is 12,735.

Further results

- The predictive distribution has cumulative probability 0.9624 at 500,000, so the prediction interval is 0 to 0. The 97.5th percentile is at 248,100.
- From WinBUGS the expected layer cost is 13,380 with a confidence interval of 8,510 to 19,260. The prediction interval is also 0 to 0 with a97.5th percentile at 272,800.

References

- **WinBUGS (free) is available at http://www.mrc-bsu.cam.ac.uk/bugs/**
- **A good introduction to MCMC and WinBUGS is *Actuarial Modeling with MCMC and BUGS*, D. Scollnik, 2001, *NAAJ*, 96-125.**
- **http://www.math.ucalgary.ca/~scollnik/abcd/ has additional worked examples with BUGS code that follow up the ideas in his paper.**
- **The WinBUGS files and EXCEL sheets for my examples are available by request to me at stuart.klugman@drake.edu.**