Discrete Choice Modeling

William Greene

Stern School of Business

New York University

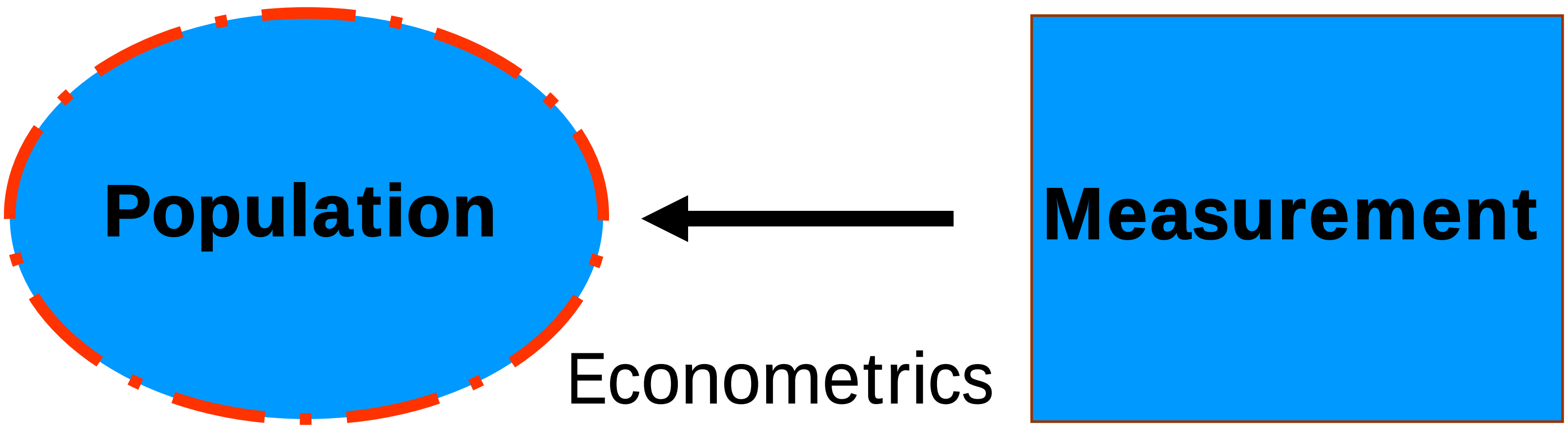
5. BAYESIAN ECONOMETRICS

Bayesian Estimation

Philosophical underpinnings: The meaning of statistical information

How to combine information contained in the sample with prior information

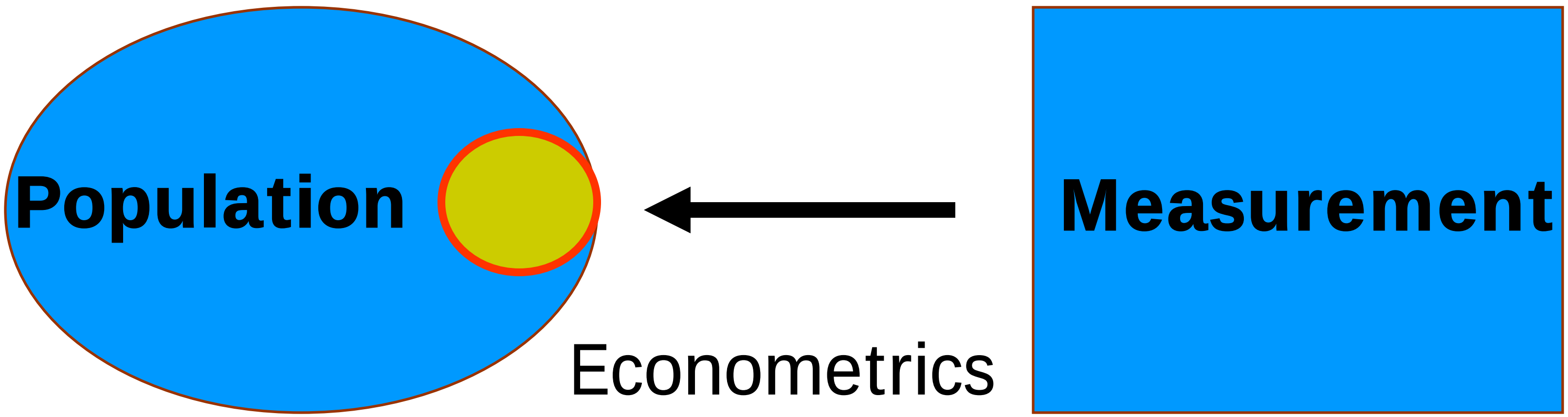
Classical Inference



Characteristics
Behavior Patterns
Choices

Imprecise inference about the entire population – sampling theory and asymptotics

Bayesian Inference



Characteristics
Behavior Patterns
Choices

Sharp, ‘exact’ inference about only the sample – the ‘posterior’ density.

Paradigms

- Classical
  - Formulate the theory
  - Gather evidence
    - Evidence consistent with theory? Theory stands and waits for more evidence to be gathered
    - Evidence conflicts with theory? Theory falls
- Bayesian
  - Formulate the theory
  - Assemble existing evidence on the theory
  - Form beliefs based on existing evidence
  - (*) Gather new evidence
  - Combine beliefs with new evidence
  - Revise beliefs regarding the theory
  - Return to (*)

On Objectivity and Subjectivity

- Objectivity and “Frequentist” methods in Econometrics – The data speak
- Subjectivity and Beliefs
  - Priors
  - Evidence
  - Posteriors
- Science and the Scientific Method

Foundational Result

- A method of using new information to update existing beliefs about probabilities of events
- **Bayes Theorem** for events. (Conceived for updating beliefs about games of chance)

$$\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \qquad = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

$$\Pr(\text{Nature} \mid \text{Evidence}) = \frac{\Pr(\text{Evidence} \mid \text{Nature})\Pr(\text{Nature})}{\Pr(\text{Evidence})}$$

Likelihoods

- (Frequentist) The likelihood is the density of the observed data conditioned on the parameters
  - Inference based on the likelihood is usually "maximum likelihood"
- (Bayesian) A function of the parameters and the data that forms the basis for inference – not a probability distribution
  - The likelihood embodies the current information about the parameters and the data

The Likelihood Principle

- The likelihood embodies ALL the current information about the parameters and the data
- Proportional likelihoods should lead to the same inferences, even given different interpretations.

“Estimation”

- Assembling information
- Prior information = out of sample. Literally prior or outside information
- Sample information is embodied in the likelihood
- Result of the analysis: “Posterior belief” = blend of prior and likelihood

Bayesian Investigation

- No fixed "parameters." θ is a random variable.
- Data are realizations of random variables.
  There is a marginal distribution p(**data**)
- Parameters are part of the random state of nature, p(θ) = distribution of θ independently (prior to) the data, as understood by the analyst. (Two analysts could legitimately bring different priors to the study.)
- Investigation combines sample information with prior information.
- Outcome is a revision of the prior based on the observed information (**data**)

The Bayesian Estimator

- The posterior distribution embodies all that is "believed" about the model.
  - Posterior = f(model|data)

    = Likelihood(θ,data) * prior(θ) / P(data)
- "Estimation" amounts to examining the characteristics of the posterior distribution(s).
  - Mean, variance
  - Distribution
  - Intervals containing specified probabilities

Priors and Posteriors

- The Achilles heel of Bayesian Econometrics
- Noninformative and Informative priors for estimation of parameters
  - Noninformative (diffuse) priors: How to incorporate the total lack of prior belief in the Bayesian estimator. The estimator becomes solely a function of the likelihood
  - Informative prior: Some prior information enters the estimator. The estimator mixes the information in the likelihood with the prior information.
- Improper and Proper priors
  - $P(\theta)$ is uniform over the allowable range of θ
  - Cannot integrate to 1.0 if the range is infinite.
  - Salvation – improper, but noninformative priors will fall out of the posterior.

Symmetrical Treatment of Data and Parameters

- **Likelihood** is p(**data**|θ)
- **Prior** summarizes nonsample information about θ in p(θ)
- Joint distribution is p(**data**, θ)
- P(**data**,θ) = p(**data**|θ)p(θ)
- Use Bayes theorem to get p(θ|**data**) = **posterior distribution**

The Posterior Distribution

Sample information $L(\mathbf{data}|\theta)$

Prior information $p(\theta)$

Joint density for θ and **data** $= p(\theta,\mathbf{data}) = L(\mathbf{data}\,|\,\theta)p(\theta)$

Conditional density for θ given the data

$$\boxed{p(\theta|\mathbf{data}) = \frac{p(\theta,\mathbf{data})}{p(\mathbf{data})} = \frac{L(\mathbf{data}\,|\,\theta)p(\theta)}{\int_\theta L(\mathbf{data}\,|\,\theta)p(\theta)d\theta} = \text{posterior density}}$$

Information obtained from the investigation

$E[\theta|\text{data}]$ = posterior mean = the Bayesian "estimate"

$Var[\theta|\text{data}]$ = posterior variance used for form interval estimates

Quantiles of $\theta|$data such as median, or 2.5th and 97.5th quantiles

Priors – Where do they come from?

- What does the prior contain?
  - **Informative priors** – real prior information
  - **Noninformative priors**
- Mathematical complications
  - Diffuse
    - Uniform
    - Normal with huge variance
  - Improper priors
- **Conjugate priors**

$$\mathrm{p}(\theta|\mathbf{data}) = \frac{\mathrm{L}(\mathbf{data}\theta)\mathrm{p}(\theta)}{\int_\theta \mathrm{L}(\mathbf{data}\theta)\mathrm{p}(\theta)d\theta}$$

Application

Estimate θ, the probability that a production process will produce a defective product.

Sampling design: Choose $N = 25$ items from the production line.
D = the number of defectives.

Result of our experiment $D = 8$

Likelihood for the sample of data is $L(\theta \mid \text{data}) = \theta^{D}(1-\theta)^{25-D},\ 0 \le \theta \le 1.$

Maximum likelihood estimator of θ is $q = D/25 = 0.32$,

Asymptotic variance of the MLE is estimated by $q(1-q)/25 = 0.008704$.

Application: Posterior Density

Posterior density

$$p(\theta\,|\,\textbf{data}) = p(\theta\,|\,N,D) = \frac{\theta^{D}(1\text{-}\theta)^{N\text{-}D}p(\theta)}{\int_{\theta}\theta^{D}(1\text{-}\theta)^{N\text{-}D}p(\theta)d\theta}.$$

Noninformative prior:

All allowable values of θ are equally likely. Uniform distribution over $(0,1)$.

$$p(\theta) = 1,\ 0 \le \theta \le 1.\ \text{Prior mean} = 1/2.\ \text{Prior variance} = 1/12.$$

Posterior density

$$p(\theta\,|\,\textbf{data}) = \frac{\theta^{D}(1-\theta)^{N-D}}{\left(\dfrac{\Gamma(D+1)\Gamma(N-D+1)}{\Gamma(D+1+N-D+1)}\right)}$$

$$= \frac{\Gamma(N+2)\theta^{D}(1-\theta)^{N-D}}{\Gamma(D+1)\Gamma(N-D+1)}$$

Note: $\int_0^1 \theta^{D}(1-\theta)^{N-D}\times 1\,d\theta$ = A beta integral with a = D+1 and b = N-D+1

$$= \beta(D,N) = \frac{\Gamma(D+1)\Gamma(N-D+1)}{\Gamma(D+1+N-D+1)}$$

Posterior Moments

Posterior Density with uniform noninformative prior

$$p(\theta|N,D) = \frac{\Gamma(N+2)\theta^{D}(1-\theta)^{N-D}}{\Gamma(D+1)\Gamma(N-D+1)}$$

Posterior Mean

$$E[\theta|data] = \int_0^1 \theta \frac{\Gamma(N+2)\theta^{D}(1-\theta)^{N-D}}{\Gamma(D+1)\Gamma(N-D+1)} d\theta$$

This is a beta integral. The posterior is a beta density with

α=D+1, β=N-D+1. The mean of a beta variable = $\dfrac{\alpha}{\alpha+\beta}$

Posterior mean = $\dfrac{D+1}{N+2} = 9/27 = .3333$

Prior mean = .5000. MLE = 8/25 = .3200.

Posterior variance = $\dfrac{(D+1)/(N-D+1)}{(N+3)(N+2)^2}$ 0.007936

Prior variance = 1/12 = .08333; Variance of the MLE = .008704.

Informative prior

Beta is a common conjugate prior for a proportion or probability

$$p(\theta) = \frac{\Gamma(\alpha+\beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}, \text{ Prior mean is } E[\theta]=\frac{\alpha}{\alpha+\beta}$$

Posterior is

$$p(\theta|N,D)=\frac{\theta^{D}(1-\theta)^{N-D}\frac{\Gamma(\alpha+\beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}}{\int_0^1 \theta^{D}(1-\theta)^{N-D}\frac{\Gamma(\alpha+\beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}d\theta}$$

$$=\frac{\theta^{D+\alpha-1}(1-\theta)^{N-D+\beta-1}}{\int_0^1 \theta^{D+\alpha-1}(1-\theta)^{N-D+\beta-1}d\theta}$$

This is a beta density with parameters ($D+\alpha$, $N-D+\beta$)

The posterior mean is $E[\theta|N,D] = \frac{\alpha+D}{\alpha+\beta+N}$; $\alpha=\beta=1$ in earlier example.

Mixing Prior and Sample Information

A typical result (exact for sampling from the normal distribution with known variance)

$$\text{Posterior mean} = w \times \text{Prior Mean} + (1\text{-}w) \times \text{MLE}$$

$$= w \times (\text{Prior Mean - MLE}) + \text{MLE}$$

$$w = \frac{\text{Posterior Mean - MLE}}{\text{Prior Mean - MLE}} = \frac{.3333 - .32}{.5 - .32} = .073889$$

Approximate Result

$$\text{Posterior Mean} \approx \frac{\dfrac{\text{Prior Mean}}{\text{Prior Variance}} + \dfrac{\text{MLE}}{\text{Asymptotic Variance}}}{\dfrac{1}{\text{Prior Variance}} + \dfrac{1}{\text{Asymptotic Variance}}} = \omega\text{Prior} + (1\text{-}\omega)\text{MLE}$$

$$\omega = \frac{\dfrac{1}{\text{Prior Variance}}}{\dfrac{1}{\text{Prior Variance}} + \dfrac{1}{\text{Asymptotic Variance}}} = \frac{1/(1/12)}{1/(1/12) + 1/(.008704)} = .09547$$

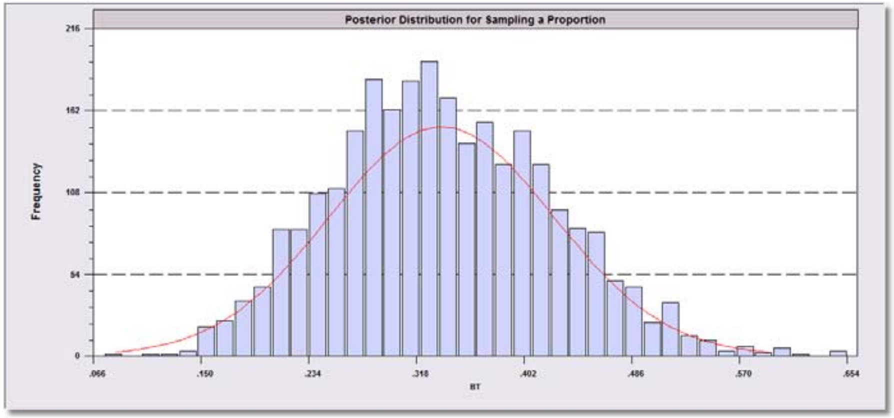
Modern Bayesian Analysis

Posterior Mean = $\int_\theta \theta p(\theta\,|\,\mathbf{data})d\theta$

Integral is often complicated, or does not exist in closed form.

Alternative strategy: Draw a random sample from the posterior distribution and examine moments, quantiles, etc.

Example: Our posterior is Beta(9,18). Based on a random sample of 5,000 draws from this population:




```
Bayesian Estimate of Distribution of θ (Posterior mean was     .333333)
Observations        =          5000     (Posterior variance was .007936)
Sample Mean         =       .334017
Sample variance     =       .007454   Standard Deviation =       .086336
Skewness            =       .248077   Kurtosis-3 (excess)=     -.161478
Minimum             =       .066214   Maximum            =       .653625
.025 Percentile     =       .177090   .975 Percentile    -       .510028
```

Bayesian Estimator

First generation: Do the integration (math)

$$E(\boldsymbol{\beta} \mid \text{data}) = \int_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta} f(\text{data} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta})}{f(\text{data})} d\boldsymbol{\beta}$$

The Linear Regression Model

Likelihood

$$L(\beta,\sigma^2|y,X)=[2\pi\sigma^2]^{-n/2}e^{-[(1/(2\sigma^2))(y-X\beta)'(y-X\beta)]}$$

Transformation using d=(N-K) and $s^2 = (1/d)(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb})$

$$\left(-\frac{1}{2\sigma^2}\right)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \left(-\frac{1}{2}ds^2\right)\left(\frac{1}{\sigma^2}\right) - \frac{1}{2}(\boldsymbol{\beta}-\mathbf{b})'\left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)(\boldsymbol{\beta}-\mathbf{b})$$

Diffuse uniform prior for β, conjugate gamma prior for σ^2

Joint Posterior

$$f(\boldsymbol{\beta},\sigma^2\mid\mathbf{y},\mathbf{X}) \propto \frac{[ds^2]^{v+2}}{\Gamma(d+2)}\left[\frac{1}{\sigma^2}\right]^{d+1} e^{-ds^2(1/\sigma^2)}[2\pi]^{-K/2}\,|\,\sigma^2(\mathbf{X}'\mathbf{X})^{-1}|^{-1/2}$$
$$\times\exp\{-(1/2)(\boldsymbol{\beta}-\mathbf{b})'[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]^{-1}(\boldsymbol{\beta}-\mathbf{b})\}$$

Marginal Posterior for β

After integrating σ^2 out of the joint posterior:

$$\boxed{f(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) \propto \frac{\dfrac{[ds^2]^{v+2}\Gamma(d+K/2)}{\Gamma(d+2)}[2\pi]^{-K/2} \mid \mathbf{X}'\mathbf{X} \mid^{-1/2}}{[ds^2 + \frac{1}{2}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b})]^{d+K/2}}.}$$

Multivariate t with mean $\mathbf{b}$ and variance matrix $\dfrac{\text{n-K}}{n-K-2}[s^2(\mathbf{X}'\mathbf{X})^{-1}]$

The Bayesian 'estimator' equals the MLE. Of course; the prior was noninformative. The only information available is in the likelihood.

Modern Bayesian Analysis

- Multiple parameter settings
- Derivation of exact form of expectations and variances for $p(\theta_1, \theta_2, \ldots, \theta_K \mid \text{data})$ is hopelessly complicated even if the density is tractable.
- Strategy: Sample joint observations $(\theta_1, \theta_2, \ldots, \theta_K)$ from the posterior population and use marginal means, variances, quantiles, etc.
- How to sample the joint observations??? (Still hopelessly complicated.)

A Practical Problem

Sampling from the joint posterior may be impossible.

E.g., linear regression.

$$f(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X}) \propto \frac{[vs^2]^{v+2}}{\Gamma(v+2)} \left[\frac{1}{\sigma^2}\right]^{v+1} e^{-vs^2(1/\sigma^2)} [2\pi]^{-K/2} \mid \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mid^{-1/2}$$

$$\times \exp(-(1/2)(\boldsymbol{\beta} - \mathbf{b})'[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]^{-1}(\boldsymbol{\beta} - \mathbf{b}))$$

What is this???

To do 'simulation based estimation' here, we need joint observations on $(\boldsymbol{\beta}, \sigma^2)$.

A Solution to the Sampling Problem

The joint posterior, $p(\boldsymbol{\beta},\sigma^2|\text{data})$ is intractable. But,

For inference about $\boldsymbol{\beta}$, a sample from the marginal posterior, $p(\boldsymbol{\beta}|\text{data})$ would suffice.

For inference about σ^2, a sample from the marginal posterior of σ^2, $p(\sigma^2|\text{data})$ would suffice.

Can we deduce these? For this problem, we do have conditionals:

$$p(\boldsymbol{\beta}|\sigma^2,\text{data}) = N[\mathbf{b},\sigma^2(\mathbf{X'X})^{-1}]$$

$$p(\sigma^2|\boldsymbol{\beta},\text{data}) \propto K \times \frac{\Sigma_i(y_i - \mathbf{x}\boldsymbol{\beta})^2}{\boldsymbol{\sigma^2}} = \text{a gamma distribution}$$

Can we use this information to sample from $p(\boldsymbol{\beta}|\text{data})$ and $p(\sigma^2|\text{data})$?

Magic Tool: The Gibbs Sampler

- Problem: How to sample observations from the a population, $p(\theta_1,\theta_2,...,\theta_K|\text{data})$.
- Solution: The Gibbs Sampler.
- Target: Sample from $f(x_1, x_2)$ = joint distribution
- Joint distribution is unknown or it is not possible to sample from the joint distribution.
- Assumed: Conditional distributions $f(x_1|x_2)$ and $f(x_2|x_1)$ are both known and marginal samples can be drawn from both.
- Gibbs sampling: Obtain one draw from x_1,x_2 by many cycles between $x_1|x_2$ and $x_2|x_1$.
  - Start $x_{1,0}$ anywhere in the right range.
  - Draw $x_{2,0}$ from $x_2|x_{1,0}$.
  - Return to $x_{1,1}$ from $x_1|x_{2,0}$ and so on.
  - Several thousand cycles produces a draw
  - Repeat several thousand times to produce a sample
- Average the draws to estimate the marginal means.

Bivariate Normal Sampling

Draw a random sample from bivariate normal $\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$

(1) Direct approach: $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_r = \Gamma \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_r$ where $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ are two

independent standard normal draws (easy) and $\Gamma = \begin{pmatrix} 1 & 0 \\ \theta_1 & \theta_2 \end{pmatrix}$

such that $\Gamma\Gamma' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. $\theta_1 = \rho$, $\theta_2 = \sqrt{1-\rho^2}$.

Application: Bivariate Normal

- Obtain a bivariate normal sample (x,y) from Normal[(0,0),(1,1,ρ)]. N = 5000.
- Conditionals: x|y is N[ρy,(1- ρ^2)]
  y|x is N[ρx,(1- ρ^2)].
- Gibbs sampler: y0=0.
  - x1 = ρy0 + sqr(1- ρ^2)v where v is a N(0,1) draw
  - y1 = ρx1 + sqr(1- ρ^2)w where w is a N(0,1) draw
- Repeat cycle 60,000 times. Drop first 10,000. Retain every 10th observation of the remainder.

Gibbs Sampling for the Linear Regression Model

$$p(\boldsymbol{\beta}|\sigma^2,\text{data}) = N[\mathbf{b},\sigma^2(\mathbf{X'X})^{-1}]$$

$$p(\sigma^2|\boldsymbol{\beta},\text{data}) \propto K \times \frac{\Sigma_i(y_i - \mathbf{x}\boldsymbol{\beta})^2}{\sigma^2}$$

$$= \text{a gamma distribution}$$

Iterate back and forth between these two distributions

More General Gibbs Sampler

- Objective: Sample joint observations on $\theta_1,\theta_2,\ldots,\theta_K$. from $p(\theta_1,\theta_2,\ldots,\theta_K|\mathbf{data})$ (Let K = 3)
- Derive $p(\theta_1|\theta_2,\theta_3,\mathbf{data})$ $p(\theta_2|\theta_1,\theta_3,\mathbf{data})$ $p(\theta_3|\theta_1,\theta_2,\mathbf{data})$
- Gibbs Cycles produce joint observations

  0. Start $\theta_1,\theta_2,\theta_3$ at some reasonable values
  1. Sample a draw from $p(\theta_1|\theta_2,\theta_3,\mathbf{data})$ using the draws of θ_1,θ_2 in hand
  2. Sample a draw from $p(\theta_2|\theta_1,\theta_3,\mathbf{data})$ using the draw at step 1 for $\theta 1$
  3. Sample a draw from $p(\theta_3|\theta_1,\theta_2,\mathbf{data})$ using the draws at steps 1 and 2
  4. Return to step 1. After a burn in period (a few thousand), start collecting the draws. The set of draws ultimately gives a sample from the joint distribution.
- Order within the chain does not matter.

Using the Gibbs Sampler to Estimate a Probit Model

Probit Model: $y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$; $y = 1[y^* > 0]$; $\varepsilon \sim N[0,1]$.

Implication: $\text{Prob}[y=1|\mathbf{x},\boldsymbol{\beta}] = \Phi(\boldsymbol{\beta}'\mathbf{x})$

$\text{Prob}[y=0|\mathbf{x},\boldsymbol{\beta}] = 1 - \Phi(\boldsymbol{\beta}'\mathbf{x})$

Likelihood Function $L(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \prod_{i=1}^{N} [1 - \Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{1-y_i}[\Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{y_i}$

Uninformative prior $p(\boldsymbol{\beta}) \propto 1$

Posterior density $$p(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \frac{\left\{\prod_{i=1}^{N} [1 \quad \Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{1-y_i}[\Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{y_i}\right\}1}{\int_{\boldsymbol{\beta}} \left[\left\{\prod_{i=1}^{N} [1 \quad \Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{1-y_i}[\Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{y_i}\right\}1\right]d\boldsymbol{\beta}}$$

Posterior Mean $$\hat{\boldsymbol{\beta}} = E[\boldsymbol{\beta}|\mathbf{y},\mathbf{X}] = \frac{\int_{\boldsymbol{\beta}} \left[\boldsymbol{\beta}\left\{\prod_{i=1}^{N} [1 - \Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{1-y_i}[\Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{y_i}\right\}1\right]d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} \left[\left\{\prod_{i=1}^{N} [1 \quad \Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{1-y_i}[\Phi(\boldsymbol{\beta}'\mathbf{x}_i)]^{y_i}\right\}1\right]d\boldsymbol{\beta}}$$

Strategy: Data Augmentation

- Treat y_i^* as unknown 'parameters' with $\boldsymbol{\beta}$
- 'Estimate' $\theta = (\boldsymbol{\beta}, y_1^*, \ldots, y_N^*) = (\boldsymbol{\beta}, \mathbf{y^*})$
- Draw a sample of R observations from the joint population $(\boldsymbol{\beta}, \mathbf{y^*})$.
- Use the marginal observations on $\boldsymbol{\beta}$ to estimate the characteristics (e.g., mean) of the distribution of $\boldsymbol{\beta}|\mathbf{y},\mathbf{X}$

Gibbs Sampler Strategy

- $p(\beta|\mathbf{y}^*,(\mathbf{y},\mathbf{X}))$. If y* is known, y is known.
  $p(\beta|\mathbf{y}^*,(\mathbf{y},\mathbf{X})) = p(\beta|\mathbf{y}^*,\mathbf{X})$.
- $p(\beta|\mathbf{y}^*,\mathbf{X})$ defines a linear regression with N(0,1) normal disturbances.
- Known result for $\boldsymbol{\beta}|\mathbf{y}^*$:
  $p(\beta|\mathbf{y}^*,(\mathbf{y},\mathbf{X}), \varepsilon=N[\mathbf{0},\mathbf{I}]) = N[\mathbf{b}^*,(\mathbf{X'X})^{-1}]$
  $\mathbf{b}^* = (\mathbf{X'X})^{-1}\mathbf{X'y^*}$
- Deduce a result for $\mathbf{y}^*|\boldsymbol{\beta}$

Gibbs Sampler, Continued

- $y_i^*|\boldsymbol{\beta},\mathbf{x}_i$ is Normal$[\mathbf{x}_i'\boldsymbol{\beta},1]$
- y_i is informative about y_i^*:

  - If $y_i = 1$, then $y_i^* > 0$; $p(y_i^*|\boldsymbol{\beta},\mathbf{x}_i\ y_i = 1)$ is truncated normal: $p(y_i^*|\boldsymbol{\beta},\mathbf{x}_i\ y_i = 1) = \phi(\mathbf{x}_i'\boldsymbol{\beta})/[1-\Phi(\mathbf{x}_i'\boldsymbol{\beta})]$
    Denoted $N^+[\mathbf{x}_i'\boldsymbol{\beta},1]$

  - If $y_i = 0$, then $y_i^* \leq 0$; $p(y_i^*|\boldsymbol{\beta},\mathbf{x}_i\ y_i = 0)$ is truncated normal: $p(y_i^*|\boldsymbol{\beta},\mathbf{x}_i\ y_i = 0) = \phi(\mathbf{x}_i'\boldsymbol{\beta})/\Phi(\mathbf{x}_i'\boldsymbol{\beta})$
    Denoted $N^-[\mathbf{x}_i'\boldsymbol{\beta},1]$

Generating Random Draws from f(x)

The inverse probability method of sampling random draws:

If F(x) is the CDF of random variable x, then a random draw on x may be obtained as $F^{-1}(u)$ where u is a draw from the standard uniform (0,1).

Examples:

Exponential: $f(x)=\theta\exp(-\theta x)$;

$F(x)=1-\exp(-\theta x)$

$x = -(1/\theta)\log(1-u)$

Normal: $F(x) = \Phi(x)$; $x = \Phi^{-1}(u)$

Truncated Normal: $x=\mu_i + \Phi^{-1}[1-(1-u)*\Phi(\mu_i)]$ for y=1;

$x= \mu_i + \Phi^{-1}[u\Phi(-\mu_i)]$ for y=0.

Sampling from the Truncated Normal

The usual inverse probability transform.

Begin with a draw from U[0,1].

U_r = the draw.

To obtain a draw y_r^* from $N^+[\mu,1]$

$$y_r^* = \mu + \Phi^{-1}[1-(1-U_r)\Phi(\mu)]$$

To obtain a draw y_r^* from $N^-[\mu,1]$

$$y_r^* = \mu + \Phi^{-1}[U_r\Phi(-\mu)]$$

Sampling from the Multivariate Normal

A multivariate version of the inverse probability transform

To sample $\mathbf{x}$ from N[$\boldsymbol{\mu}$,$\boldsymbol{\Sigma}$] (K dimensional)

Let $\mathbf{L}$ be the Cholesky matrix such that $\mathbf{LL}' = \boldsymbol{\Sigma}$

Let $\mathbf{v}$ be a column of K independent random normal(0,1) draws.

Then $\boldsymbol{\mu} + \mathbf{Lv}$ is normally distributed with mean $\boldsymbol{\mu}$ and variance $\mathbf{LIL}' = \Sigma$ as needed.

Gibbs Sampler

- Preliminary:
  Obtain **X’X** then **L** such that **LL’** = (**X’X**)$^{-1}$.
- Preliminary: Choose initial value for β such as
  β_0 = **0**. Start with r = 1.
- (**y*** step) Sample N observations on **y***(r) using
  β_{r-1} , $\mathbf{x}_i$ and y_i and the transformations for the truncated normal distribution.
- (β step) Compute b*(r) = (**X’X**)$^{-1}$**X’y***(r). Draw the observation on β(r) from the normal population with mean b*(r) and variance (**X’X**)$^{-1}$.
- Cycle between the two steps 50,000 times. Discard the first 10,000 and retain every 10th observation from the retained 40,000.

Frequentist and Bayesian Results

TABLE 16.3 Probit Estimates for Doctor Visits Equation

	Maximum Likelihood		*Posterior Means and Std. Devs*	
Variable	*Estimate*	*Standard Error*	*Posterior Mean*	*Posterior S.D.*
Constant	−0.12433	0.058146	−0.12628	0.054759
Age	0.011892	0.00079568	0.011979	0.00080073
Education	−0.014966	0.0035747	−0.015142	0.0036246
Income	−0.13242	0.046552	−0.12669	0.047979
Kids	−0.15212	0.018327	−0.15149	0.018400
Married	0.073522	0.020644	0.071977	0.020852
Female	0.35591	0.016017	0.35582	0.015913

0.37 Seconds   2 Minutes

Appendix

```
Namelist; x=one,age,educ,hhninc,hhkids,married,female$
Create   ; y = doctor $

calc;k=col(x)$
Matrix   ; xx=x'x ; xxi = <xx> $
Calc     ; Rep = 500 $
Probit   ; lhs=y;rhs=x$
Calc     ; iter=0$
? Gibbs sampler
Matrix   ; beta=init(k,1,0) ; bbar=beta;bv=init(k,k,0)$
Proc     = gibbs$
Do for   ; simulate ; r =1,Rep $
Calc     ; iter=iter+1$
Create   ; mui = x'beta ; f = rnu(0,1)
         ; if(y=1)  ysg = mui + inp(1-(1-f)*phi( mui));
             (else) ysg = mui + inp(     f *phi(-mui)) $
Matrix   ; mb = xxi*x'ysg ; beta = rndm(mb,xxi) $
Matrix   ; if(iter > 100) ; bbar=bbar+beta ; bv=bv+beta*beta'$
Enddo    ; simulate $
Endproc $
Execute ; Proc = Gibbs $
Calc     ; Ri = Rep - 100 ; ri = 1/ri $
Matrix   ; bbar=ri*bbar ; bv=ri*bv-bbar*bbar' $
Matrix   ; Stat(bbar,bv,x); Stat(b,varb,x) $
```

Bayesian Model Estimation

- Specification of conditional likelihood: f(data | parameters) = L(parameters|data)
- Specification of priors: g(parameters)
- Posterior density of parameters:

$$\text{f(parameters | data)} = \frac{\text{f(data | parameters)g(parameters)}}{\text{f(data)}}$$

- Posterior mean = E[parameters|data]

The Marginal Density for the Data is Irrelevant

$$f(\boldsymbol{\beta}|\text{data}) = \frac{f(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})}{f(\text{data})} = \frac{L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})}{f(\text{data})}$$

Joint density of $\boldsymbol{\beta}$ and data is $f(\text{data},\boldsymbol{\beta}) = L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})$

Marginal density of the data is

$$f(\text{data}) = \int_{\beta} f(\text{data},\boldsymbol{\beta})d\boldsymbol{\beta} = \int_{\beta} L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}$$

$$\text{Thus, } f(\boldsymbol{\beta}|\text{data}) = \frac{L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})}{\int_{\beta} L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}}$$

$$\text{Posterior Mean} = \int_{\boldsymbol{\beta}} p(\boldsymbol{\beta}|\text{data})d\boldsymbol{\beta} = \frac{\int_{\boldsymbol{\beta}} \boldsymbol{\beta}\, L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}}{\int_{\boldsymbol{\beta}} L(\text{data}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}}$$

Requires specification of the likeihood and the prior.

Bayesian Estimators

- Bayesian “Random Parameters” vs. Classical Randomly Distributed Parameters
- Models of Individual Heterogeneity
  - Sample Proportion
  - Linear Regression
  - Binary Choice
  - Random Effects: Consumer Brand Choice
  - Fixed Effects: Hospital Costs

A Random Effects Approach

- Allenby and Rossi, “Marketing Models of Consumer Heterogeneity”
  - Discrete Choice Model – Brand Choice
  - Hierarchical Bayes
  - Multinomial Probit
- Panel Data: Purchases of 4 brands of ketchup

Structure

Conditional data generation mechanism

$y_{it,j}^* = \boldsymbol{\beta}_i' x_{it,j} + \varepsilon_{it,j} =$ utility for consumer i, choice t, brand j.

$Y_{it,j} = 1[y_{it,j}^* = \text{maximum utility among the J choices}]$

$x_{it,j} = (\text{constant, log price, "availability," "featured"})$

$\varepsilon_{it,j} \sim N[0, \lambda_j], \lambda_1 = 1$

Implies a J outcome multinomial probit model.

Priors

Prior Densities

$$\boldsymbol{\beta}_i \sim N\left[\overline{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}\right],$$

$$\text{Implies } \boldsymbol{\beta}_i = \overline{\boldsymbol{\beta}} + \mathbf{w}_i, \mathbf{w}_i \sim N[\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}}]$$

$$\lambda_j \sim \text{Inverse Gamma}[v, s_j]$$

$$\text{(looks like chi-squared)}, v = 3, s_j = 1$$

Priors over model parameters

$$\overline{\boldsymbol{\beta}} \sim N\left[\overline{\overline{\boldsymbol{\beta}}}, aV_{\beta}\right], \overline{\overline{\boldsymbol{\beta}}} = \mathbf{0}$$

$$\mathbf{V}_{\beta}^{-1} \sim \text{Wishart}[v_0, \mathbf{V}_0], v_0 = 8, \mathbf{V}_0 = 8\mathbf{I}$$

Bayesian Estimator

- Joint Posterior = $E[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N,\bar{\boldsymbol{\beta}},\boldsymbol{V}_{\boldsymbol{\beta}} | {}_1 \quad]_J$ **data**
- Integral does not exist in closed form.
- Estimate by random samples from the joint posterior.
- Full joint posterior is not known, so not possible to sample from the joint posterior.
- Gibbs sampler is used to sample from posterior

Gibbs Cycles for the MNP Model

Marginal posterior for the individual parameters

(Known and can be sampled)

$$\boldsymbol{\beta}_i \mid \bar{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}, \boldsymbol{\lambda}, \text{data}$$

Marginal posterior for the common parameters

(Each known and each can be sampled)

$$\bar{\boldsymbol{\beta}} \mid \mathbf{V}_{\boldsymbol{\beta}}, \boldsymbol{\lambda}, \text{data}$$

$$\mathbf{V}_{\boldsymbol{\beta}} \mid \bar{\boldsymbol{\lambda}}, \quad , \text{data}$$

$$\boldsymbol{\lambda} \mid \bar{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}, \text{data}$$

Results

- Individual parameter vectors and disturbance variances
- Individual estimates of choice probabilities
- The same as the “random parameters probit model” with slightly different weights.
- Allenby and Rossi call the classical method an “approximate Bayesian” approach.
  - (Greene calls the Bayesian estimator an “approximate random parameters model”)
  - Who’s right?
    - Bayesian layers on implausible uninformative priors and calls the maximum likelihood results “exact” Bayesian estimators.
    - Classical is strongly parametric and a slave to the distributional assumptions.
    - Bayesian is even more strongly parametric than classical.
    - Neither is right – Both are right.

A Comparison of Maximum Simulated Likelihood and Hierarchical Bayes

- Ken Train: "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit"
- Mixed Logit

$$U(i,t,j) = \boldsymbol{\beta}_i'\mathbf{x}(i,t,j) + \varepsilon(i,t,j),$$

$$i = 1,...,N \text{ individuals,}$$

$$t = 1,...,T_i \text{ choice situations}$$

$$j = 1,...,J \text{ alternatives (may also vary)}$$

Stochastic Structure – Conditional Likelihood

$$\text{Prob}(i,j,t) = \frac{\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j,t})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j,t})}$$

$$\text{Likelihood for individual } i \; = \; \prod_{t=1}^{T}\frac{\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j^*,t})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}_i'\mathbf{x}_{i,j,t})}$$

j^* = indicator for the specific choice made by i at time t.

Note individual specific parameter vector $\boldsymbol{\beta}_i$.

Classical Approach

$$\boldsymbol{\beta}_i \sim \mathrm{N}[\mathbf{b},\boldsymbol{\Omega}];\ \text{write } \boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$$

$$\boldsymbol{\beta}_i = \mathbf{b} + \mathbf{w}_i$$

$$= \mathbf{b} + \boldsymbol{\Gamma}\mathbf{v}_i \ \text{ where } \boldsymbol{\Gamma} = \mathit{diag}(\gamma_j^{1/2}) \ \text{(uncorrelated)}$$

$$\text{Log-likelihood} = \sum_{i=1}^{N} \log \int_{\mathbf{w}} \prod_{t=1}^{T} \frac{\exp[(\mathbf{b}+\mathbf{w}_i)'\mathbf{x}_{i,j*,t}]}{\sum_{j=1}^{\mathbf{J}} \exp[(\mathbf{b}+\mathbf{w}_i)_i'\mathbf{x}_{i,j,t}]} d\mathbf{w}_i$$

Maximize over $\mathbf{b},\boldsymbol{\Gamma}$ using maximum simulated likelihood

(random parameters model)

Mixed Model Estimation

- MLWin: Multilevel modeling for Windows
  - http://multilevel.ioe.ac.uk/index.html
  - Uses mostly Bayesian, MCMC methods
  - "Markov Chain Monte Carlo (MCMC) methods allow Bayesian models to be fitted, where prior distributions for the model parameters are specified. By default *MLwin* sets diffuse priors which can be used to approximate maximum likelihood estimation." (From their website.)

Bayesian Approach – Gibbs Sampling and Metropolis-Hastings

$$\text{Posterior} = \prod_{i=1}^{N} L(\text{data} \mid \boldsymbol{\beta}_i, \boldsymbol{\Omega}) \times \text{priors}$$

$$\text{Prior} = \text{Product of 3 independent priors for}$$

$$(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N, \gamma_1,\ldots,\gamma_N, \mathbf{b})$$

$$= N(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N \mid \mathbf{b}, \boldsymbol{\Omega}) \text{ (normal)}$$

$$\times \text{InverseGamma}(\gamma_1,\ldots,\gamma_K \mid \text{parameters})$$

$$\times g(\mathbf{b} \mid \text{assumed parameters}) \text{ (Normal with large variance)}$$

Gibbs Sampling from Posteriors: b

$$p(\mathbf{b} \mid \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, \boldsymbol{\Omega}) = \text{Normal}[\bar{\boldsymbol{\beta}}, (1/N)\boldsymbol{\Omega}]$$

$$\bar{\boldsymbol{\beta}} = (1/N)\sum_{i=1}^{N}\boldsymbol{\beta}_i$$

Easy to sample from Normal with known mean and variance by transforming a set of draws from standard normal.

Gibbs Sampling from Posteriors: $\boldsymbol{\Omega}$

$$p(\gamma_k \mid \mathbf{b}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N) \sim \text{Inverse Gamma}[1+N, 1+N\bar{V}_k]$$

$$\bar{V}_k = (1/N)\sum_{i=1}^{N} (\beta_{k,i} - b_k)^2 \text{ for each } k = 1,\ldots,K$$

Draw from inverse gamma for each k:

Draw $R = 1+N$ draws from $N[0,1] = h_{r,k}$,

then the draw is $\dfrac{(1+N\bar{V}_k)}{\sum_{r=1}^{R} h_{r,k}^2}$

Gibbs Sampling from Posteriors: β_i

$$p(\boldsymbol{\beta}_i \mid \mathbf{b}, \boldsymbol{\Omega}) = M \times L(\text{data} \mid \boldsymbol{\beta}_i) \times g(\boldsymbol{\beta}_i \mid \mathbf{b}, \boldsymbol{\Omega})$$

M = a constant, L = likelihood, g = prior

This is the definition of the posterior.

Not clear how to sample.

Use Metropolis - Hastings algorithm.

Metropolis – Hastings Method

Define :

$\boldsymbol{\beta}_{i,0}$ = an 'old' draw (vector)

$\boldsymbol{\beta}_{i,1}$ = the 'new' draw (vector)

$\mathbf{d}_r = \sigma\boldsymbol{\Gamma}\mathbf{v}_r$,

σ = a constant (see below)

$\boldsymbol{\Gamma}$ = the diagonal matrix of standard deviations

$\mathbf{v}_r$ = a vector of K draws from standard normal

Metropolis Hastings: A Draw of β_i

$$\text{Trial value}: \tilde{\boldsymbol{\beta}}_{i,1} = \boldsymbol{\beta}_{i,0} + \mathbf{d}_r$$

$$R = \frac{\text{Posterior}(\tilde{\boldsymbol{\beta}}_{i,1})}{\text{Posterior}(\boldsymbol{\beta}_{i,0})} \text{(Ms cancel)}$$

$$U = \text{a random draw from } U(0,1)$$

$$\text{If } U < R, \text{ use } \tilde{\boldsymbol{\beta}}_{i,1}, \text{ else keep } \boldsymbol{\beta}_{i,0}$$

$$\text{During Gibbs iterations, draw } \boldsymbol{\beta}_{i,1}$$

$$\sigma \text{ controls acceptance rate. Try for .4.}$$

Application: Energy Suppliers

- N=361 individuals, 2 to 12 hypothetical suppliers
- X=
  - (1) fixed rates,
  - (2) contract length,
  - (3) local (0,1),
  - (4) well known company (0,1),
  - (5) offer TOD rates (0,1),
  - (6) offer seasonal rates]

Estimates: Mean of Individual β_i

	MSL Estimate (Asymptotic S.E.)	Bayes Posterior Mean (Posterior Std.Dev.)
Price	-1.04 (0.396)	-1.04 (0.0374)
Contract	-0.208 (0.0240)	-0.194 (0.0224)
Local	2.40 (0.127)	2.41 (0.140)
Well Known	1.74 (0.0927)	1.71 (0.100)
TOD	-9.94 (0.337)	-10.0 (0.315)
Seasonal	-10.2 (0.333)	-10.2 (0.310)

Nonlinear Models and Simulation

- Bayesian inference over parameters in a nonlinear model:
  - 1. Parameterize the model
  - 2. Form the likelihood conditioned on the parameters
  - 3. Develop the priors – joint prior for all model parameters
  - 4. Posterior is proportional to likelihood times prior. (Usually requires conjugate priors to be tractable.)
  - 5. Draw observations from the posterior to study its characteristics.

Simulation Based Inference

Form the likelihood L($\boldsymbol{\theta}$,data)

Form the prior p($\boldsymbol{\theta}$)

Form the posterior $K \times p(\boldsymbol{\theta})L(\boldsymbol{\theta},\text{data})$ where K is a constant that makes the whole thing integrate to 1.

Posterior mean = $\int_{\boldsymbol{\theta}} \boldsymbol{\theta}\ K \times p(\boldsymbol{\theta})L(\boldsymbol{\theta},\text{data})d\boldsymbol{\theta}$

Estimate the posterior mean by $\hat{E}(\boldsymbol{\theta}\,|\,\textbf{data}) = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\theta}_{\mathbf{r}}^{\mathbf{S}}$

by simulating draws from the posterior.

Large Sample Properties of Posteriors

- Under a uniform prior, the posterior is proportional to the likelihood function
  - Bayesian ‘estimator’ is the mean of the posterior
  - MLE equals the mode of the likelihood
  - In large samples, the likelihood becomes approximately normal – the mean equals the mode
  - Thus, in large samples, the posterior mean will be approximately equal to the MLE.

Conclusions

- Bayesian vs. Classical Estimation
  - In principle, some differences in interpretation
  - As practiced, just two different algorithms
  - The religious debate is a red herring
- Gibbs Sampler. A major technological advance
  - Useful tool for both classical and Bayesian
  - New Bayesian applications appear daily

Applications of the Paradigm

- Classical econometricians doggedly cling to their theories even when the evidence conflicts with them – that is what specification searches are all about.
- Bayesian econometricians NEVER incorporate prior evidence in their estimators – priors are always studiously noninformative. (Informative priors taint the analysis.) As practiced, Bayesian analysis is not Bayesian.

Methodological Issues

- Priors: Schizophrenia
  - Uninformative are disingenuous (and not Bayesian)
  - Informative are not objective
- Using existing information? Received studies generally do not do this.
- Bernstein von Mises theorem and likelihood estimation.
  - In large samples, the likelihood dominates
  - The posterior mean will be the same as the MLE

Standard Criticisms

- Of the Classical Approach
  - Computationally difficult (ML vs. MCMC)
  - No attention is paid to household level parameters.
  - There is no natural estimator of individual or household level parameters
  - Responses: None are true. See, e.g., Train (2003, ch. 10)

- Of Classical Inference in this Setting
  - Asymptotics are “only approximate” and rely on “imaginary samples.” Bayesian procedures are “exact.”
  - Response: The inexactness results from acknowledging that we try to extend these results outside the sample. The Bayesian results are “exact” but have no generality and are useless except for this sample, these data and this prior. (Or are they? Trying to extend them outside the sample is a distinctly classical exercise.)

Modeling Issues

- As N → ∞, the likelihood dominates and the prior disappears ➔ Bayesian and Classical MLE converge. (Needs the mode of the posterior to converge to the mean.)
- Priors
  - Diffuse → large variances imply little prior information. (NONINFORMATIVE)
  - INFORMATIVE priors – finite variances that appear in the posterior. “Taints” any final results.

Reconciliation: Bernstein-Von Mises Theorem

- The posterior distribution converges to normal with covariance matrix equal to 1/N times the information matrix (same as classical MLE). (The distribution that is converging is the posterior, not the sampling distribution of the estimator of the posterior mean.)
- The posterior mean (empirical) converges to the mode of the likelihood function. Same as the MLE. A proper prior disappears asymptotically.
- Asymptotic sampling distribution of the posterior mean is the same as that of the MLE.

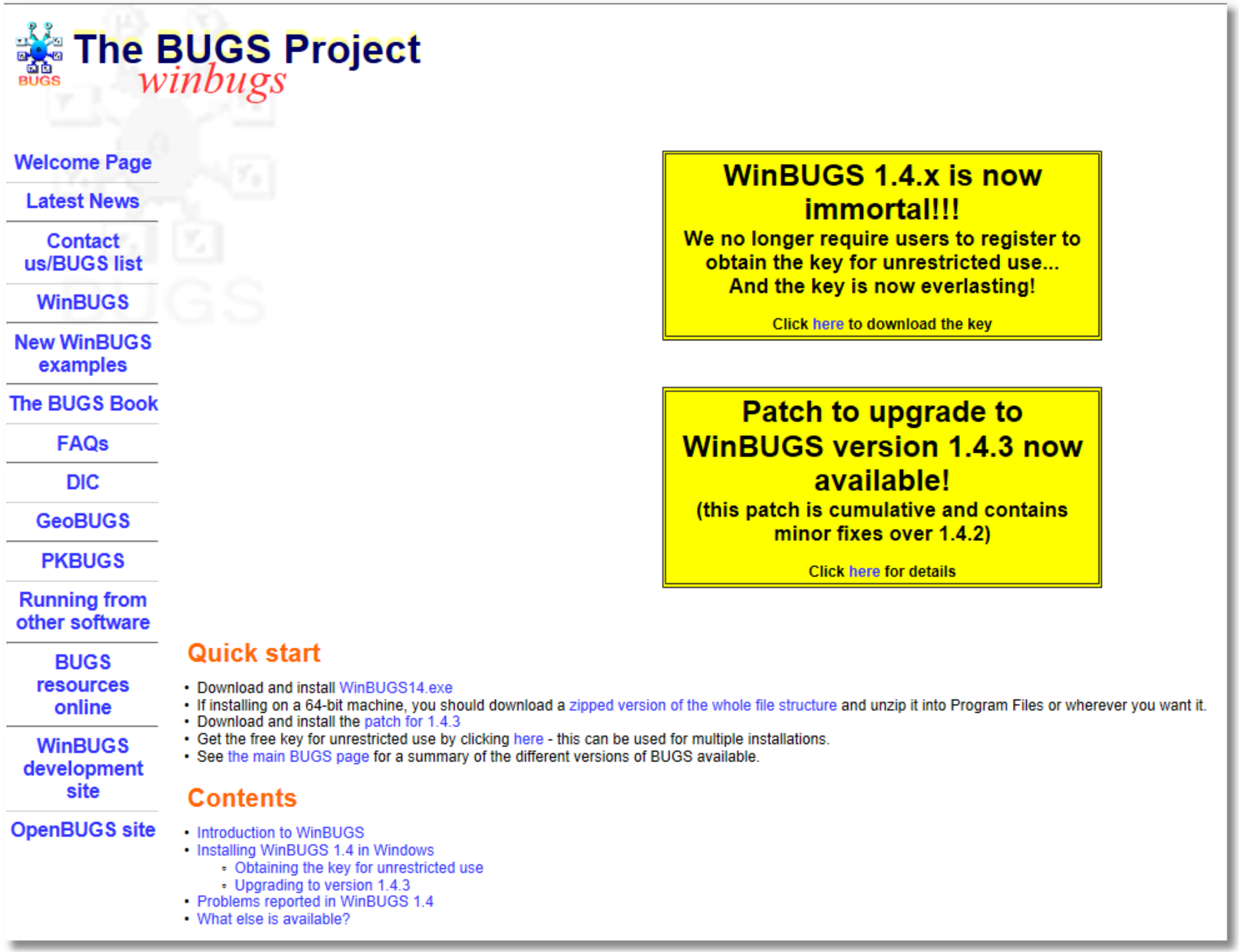
Sources

- Lancaster, T.: An Introduction to Modern Bayesian Econometrics, Blackwell, 2004
- Koop, G.: Bayesian Econometrics, Wiley, 2003
- ... “Bayesian Methods,” “Bayesian Data Analysis,” ... (many books in statistics)
- Papers in Marketing: Allenby, Ginter, Lenk, Kamakura,...
- Papers in Statistics: Sid Chib,...
- Books and Papers in Econometrics: Arnold Zellner, Gary Koop, Mark Steel, Dale Poirier,...

Software

- Stata, Limdep, SAS, etc.
- R, Matlab, Gauss
- WinBUGS
  - **B**ayesian inference **U**sing **G**ibbs **S**ampling

http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml



BUGS

The BUGS Project

winbugs

- Welcome Page
- Latest News
- Contact us/BUGS list
- WinBUGS
- New WinBUGS examples
- The BUGS Book
- FAQs
- DIC
- GeoBUGS
- PKBUGS
- Running from other software
- BUGS resources online
- WinBUGS development site
- OpenBUGS site

WinBUGS 1.4.x is now immortal!!!

We no longer require users to register to obtain the key for unrestricted use... And the key is now everlasting!

Click here to download the key

Patch to upgrade to WinBUGS version 1.4.3 now available!

(this patch is cumulative and contains minor fixes over 1.4.2)

Click here for details

Quick start

- Download and install WinBUGS14.exe
- If installing on a 64-bit machine, you should download a zipped version of the whole file structure and unzip it into Program Files or wherever you want it.
- Download and install the patch for 1.4.3
- Get the free key for unrestricted use by clicking here - this can be used for multiple installations.
- See the main BUGS page for a summary of the different versions of BUGS available.

Contents