An Introduction to Contemporary Bayesian Thinking and Methods

First of Two Mini-Courses
Department of Economics
Indiana University

John Geweke

University of Iowa

May 14, 2007
Outline

1. Motivation
2. Observables, unobservables and objects of interest
3. Conditioning and updating
4. Simulators
5. Modeling
6. Decision-making
Motivation: Decision making examples

- Development and introduction of a new drug in the United States
  - *Food and Drug Administration*: Approve the drug?
  - Pharmaceutical company: Proceed to the next stage?

- Promotion of a product
  - *Executives*: What is the optimal price and promotion policy?
  - Middle management / Research: What is the relation of price and promotion to market share?

- Mergers and acquisitions
  - *Regulatory authorities*: Should the merger or acquisition be permitted?
  - Economists and lawyers: How should evidence on price impact be synthesized and presented?
Student-teacher ratio in a public school district

- *State education policy-makers*: Should funding policy be linked to test scores? If so, how?
- *Federal education policy-makers*: Impact of “No child left behind”
- *District superintendent*: How should the trade-off between costs and test scores be resolved?
- Will a lower student-teacher ratio raise test scores? By how much?

Value at risk for financial institutions

- *Institutions* must report: Decrease in value such that a greater decrease has probability 5%
  - Return at risk example $P(\text{Return} \leq x\%) = 5$
  - “$x\%$” varies from -0.75% to -3% depending ...
Common characteristics of decision making examples

- Decision must be made using less-than-perfect information.
- Decision either must be made at a specified time, or cannot be postponed indefinitely.
- Elements of decision making – both information and consequences – are largely quantitative and involve uncertainty.
- Multiple sources of information bear on the decision.
- Both investigators and clients are key players.
Effective communication between investigators and clients

- Make all assumptions explicit.
- Explicitly quantify all of the essentials, including the assumptions.
- Synthesize, or provide the means to synthesize, different approaches and models.
- Represent uncertainty in ways that will be useful to the client.
Observables, unobservables and objects of interest: Models

Definition: A *model* is a simplified description of reality that is at least potentially useful in decision making.

*All models are false*, but

- Some are useful;
- Some are better than others;
- With inspiration and perspiration they can be improved.

Well-known example:

- Issac Newton was “wrong.”
- Albert Einstein was “right.”
- If you’re going to the moon and back, Newton is very useful.
Scientific models

- Reduce certain aspects of reality to a few quantitative concepts that are unobservable.
- Use the quantitative concepts to organize observables in a way that is useful in decision making.

“Given the values of the unobservables, the observables will behave in the following way.”
Some notation:

- $A$ (for “assumptions”) is a model.
- $\theta_A$ is a vector of unobservables:
  - parameters;
  - latent variables.
- $y$ is a vector of observables:
  - $ex \ ante \ y$ is random;
  - $ex \ post \ y = y^0$, “data.”

“Given the values of the unobservables, the observables will behave in the following way.”

$means$: a specification of

$$p(y \mid \theta_A, A)$$
The formal problem

- We have
  \[ p(y \mid \theta_A, A) \]  
  and we must get \( p(y \mid A) \).

- We know that
  \[ p(y \mid A) = \int p(\theta_A, y \mid A) \, d\theta_A, \]  
  Therefore, what is missing is
  \[ p(\theta_A \mid A) = p(\theta_A, y \mid A) / p(y \mid \theta_A, A). \]

- You cannot get from the model (1) to the prediction (2) without the additional information (3).
Objects of interest

Examples:

- What is the probability the average test score exceeds 680 when the student-teacher ratio is 20?
- What is 5% quantile of return on tomorrow’s S&P 500 index?
- What is the probability the price of Gatorade will rise more than 5% if Pepsico buys Quaker Oats?
Vector of interest

Vector of interest: \( \omega \in \Omega \subseteq \mathbb{R}^q \)

Examples:

- **What is the probability the average test score exceeds 680 when the student-teacher ratio is 20?**
  - \( \omega = \text{Average test score} \)

- **What is the 5% quantile of return on tomorrow’s S&P 500 index?**
  - \( \omega = y_{T+1}, \text{tomorrow’s return on S&P 500 index} \)

- **What is the probability the price of Gatorade will rise more than 5% if Pepsico buys Quaker Oats?**
  - \( \omega = \text{Price of Gatorade following the acquisition} \)
The decision at hand depends on the distribution of $\omega$. The model must specify

$$p(\omega \mid y, \theta_A, A),$$

as must any competing model.

Simple examples:
- Average test score
- Return on the S&P 500 index

Complex example:
- Price of Gatorade following Pepsi-Quaker Oats merger
A complete model has three components:

1. $p(\theta_A | A)$,
2. $p(y | \theta_A, A)$,
3. $p(\omega | y, \theta_A, A)$.

From these three components

$$p(\theta_A, y, \omega | A) = p(\theta_A | A) p(y | \theta_A, A) p(\omega | y, \theta_A, A).$$
Conditioning and updating

- Conditional on the model $A$,

$$p(\theta_A, y, \omega \mid A) = p(\theta_A \mid A) p(y \mid \theta_A, A) p(\omega \mid y, \theta_A, A).$$

- The relevant probability density for the decision at hand is

$$p(\omega \mid y^o, A).$$

- This is the single most important principle in Bayesian inference in support of decision making.
Steps in obtaining $p(\omega \mid y^o, A)$

- **Prior density:**
  
  $$p(\theta_A \mid A)$$

- **Observables density:**
  
  $$p(y \mid \theta_A, A)$$

- **Posterior density of $\theta_A$:**

  $$p(\theta_A \mid y^o, A) = \frac{p(\theta_A, y^o \mid A)}{p(y^o \mid A)} = \frac{p(\theta_A \mid A) p(y^o \mid \theta_A, A)}{p(y^o \mid A)}$$

  $$\propto p(\theta_A \mid A) p(y^o \mid \theta_A, A).$$

- Then

  $$p(\omega \mid y^o, A) = \int_{\Theta_A} p(\theta_A \mid y^o, A) p(\omega \mid y^o, \theta_A, A).$$
Bayesian updating

- Let

\[ \mathbf{Y}_t' = (y'_1, \ldots, y'_t); \ t = 0, \ldots, T, \ \mathbf{Y}_0 = \{\emptyset\}. \]

- Recursive model representation

\[
p (\mathbf{Y}_T \mid \theta_A, A) = \prod_{t=1}^{T} p (y_t \mid \mathbf{Y}_{t-1}, \theta_A, A)
\]

- When we have time \( t \) information but not time \( t + 1 \) information then

\[
p (\theta_A \mid \mathbf{Y}_t^o, A) \propto p (\theta_A \mid A) p (\mathbf{Y}_t^o \mid \theta_A, A)
\]

\[
= p (\theta_A \mid A) \prod_{s=1}^{t} p (y_s^o \mid \mathbf{Y}_{s-1}^o, \theta_A, A).
\]
An Introduction to Contemporary Bayesian Thinking and Methods

Conditioning and updating

Bayesian updating

\[
p(\theta_A \mid Y_t^o, A) \propto p(\theta_A \mid A) \prod_{s=1}^{t} p(y_s^o \mid Y_{s-1}^o, \theta_A, A).
\]

- When time \(t + 1\) information \(y_{t+1}^o\) becomes available then

\[
p(\theta_A \mid Y_{t+1}^o, A) \propto p(\theta_A \mid A) \prod_{s=1}^{t+1} p(y_s^o \mid Y_{s-1}^o, \theta_A, A)
\]

\[
\propto p(\theta_A \mid Y_t^o, A) p(y_{t+1}^o \mid Y_t^o, \theta_A, A).
\]

- In this process the prior density can be regarded as

\[
p(\theta_A \mid Y_t^o, A),
\]

- the conditional density of observables can be regarded as

\[
p(y_{t+1} \mid Y_t^o, \theta_A, A),
\]

- and

\[
p(\theta_A \mid Y_{t+1}^o, A)
\]

is the posterior density.
Density relevant for decision making:

\[ p(\text{“Relevant but uncertain quantities”} \mid \text{“Known data and explicit assumptions”}) \]

which is

\[ p(\omega \mid y^o, A) = \int_{\Omega_A} p(\theta_A \mid y^o, A) p(\omega \mid \theta_A, y^o, A) \, d\theta_A \]

Just ask President Harry Truman ...

More generally, Bayesian econometrics and statistics

- Always conditions on known data and explicit assumptions
- Is completely integrated with the theory of economic behavior under uncertainty

... More to come on both.
Simulators: Potential of simulation

\[ p(\omega \mid y^o, A) = \int_{\Theta_A} p(\theta_A \mid y^o, A) p(\omega \mid \theta_A, y^o, A) \, d\theta_A \]

- If we can simulate \( \theta_A^{(m)} \sim p(\theta_A \mid y^o, A) \)
- followed by \( \omega^{(m)} \sim p(\omega \mid \theta_A^{(m)}, y^o, A) \) for \( m = 1, 2, \ldots, M \)
- then it follows that \( \left\{ \theta_A^{(m)}, \omega^{(m)} \right\} \sim p(\theta_A, \omega \mid y^o, A) \)
- and so \( \omega^{(m)} \sim p(\omega \mid y^o, A) \).
Simulation in the value at risk (VaR) example

Financial manager must assess VaR of an asset 5 days hence

\[ \omega = p_{T+5} = p_T^o \exp \left( \sum_{s=1}^{5} y_{T+s} \right) \]

Data: \( y^o = (y^o_1, ..., y^o_T)' \)
Assumptions: Model A with unobservable parameter vector \( \theta_A \)
Decision: \( c \) with the property

\[ \int_{-\infty}^{p_T-c} p(\omega \mid y^o, A) \, d\omega = 0.05. \]
Solving the VaR decision problem with simulation

- \( \theta_A^{(m)} \sim p(\theta_A | y^o, A) \)
  - This is a hard problem – It involves backward simulation.
  - Recent books by Koop, Lancaster, Geweke
  - Some concrete applications in Part 2

- \( \omega^{(m)} \sim p(\omega | \theta_A, y^o, A) \)
  - This is a relatively easy problem – It involves forward simulation.
Why is forward simulation relatively easy?

\[ \omega = p_{T+5} = p_T^o \exp \left( \sum_{s=1}^{5} y_{T+s} \right) \]

The model specifies

\[ p (y_t \mid Y_{t-1}, \theta_A, A) \]

and simulating

\[ y_t \sim p (y \mid Y_{t-1}, \theta_A, A) \quad (t = 1, 2, \ldots) \]

is usually straightforward.
A concrete example – GARCH(1,1):
\[ y_t = \alpha + \epsilon_t, \quad \text{var} (\epsilon_t) = h_t, \quad h_t = \beta_0 + \beta_1 \epsilon_{t-1}^2 + \gamma_1 h_{t-1} \]

- The posterior simulator gives us
  \[ \theta^{(m)} = \left( \alpha^{(m)}, \beta_0^{(m)}, \beta_1^{(m)}, \gamma_1^{(m)} \right)' \quad (m = 1, 2, \ldots, M) \]
- Steps in simulating \( \omega^{(m)} \)

1. Define \( y_T^{(m)} = y_T^o \)
2. For \( s = 1, \ldots, 5: \)
   1. \( \epsilon_{T+s-1}^{(m)} = y_{T+s-1}^{(m)} - \alpha^{(m)} \)
   2. \( h_{T+s}^{(m)} = \beta_0^{(m)} + \beta_1 \epsilon_{T+s-1}^{(m)}^2 + \gamma_1^{(m)} h_{T+s-1}^{(m)} \)
   3. \( y_{T+s}^{(m)} \sim N \left( \alpha^{(m)}, h_{T+s}^{(m)} \right) \)
3. \( \omega^m = p_T^o \exp \left( \sum_{s=1}^{5} y_{T+s}^{(m)} \right) \)
4. Sort \( \{ \omega^1, \ldots, \omega^{(M)} \} \) and choose the appropriate order statistic to approximate \( c \quad \int_{-\infty}^{c} p \left( \omega \mid Y_T^o, A \right) = .05. \)
Modeling

How good are the assumptions

\[ p(\theta_A | A), \]

\[ p(y | \theta_A, A), \]

\[ p(\omega | y, \theta_A, A)? \]

Two approaches:
- Model comparison
- Prior predictive analysis
Model comparison

- Competing complete models \( A = \{A_1, A_2, \ldots, A_J\} \), specifying
  \[
p \left( \theta_{A_j} \mid A_j \right), \ p \left( y \mid \theta_{A_j}, A_j \right), \ p \left( \omega \mid y, \theta_{A_j}, A_j \right) \quad (j = 1, \ldots, J)
\]
- To this we add the prior probability of each model
  \[
p \left( A_j \right) = P \left( A_j \mid A \right), \text{ with } \sum_{j=1}^{J} p \left( A_j \right) = 1.
\]
- For each model \( j \) we have already derived
  \[
p \left( \omega \mid y^o, A_j \right).
\]
- Then
  \[
p \left( \omega \mid y^o, A \right) = \sum_{j=1}^{J} p \left( \omega \mid y^o, A_j \right) p \left( A_j \mid y^o, A \right),
\]
  sometimes called model averaging.
\[ p(\omega \mid y^o, A) = \sum_{j=1}^{J} p(\omega \mid y^o, A_j) p(A_j \mid y^o, A) \]

- The “weights” are

\[ p(A_j \mid y^o, A) = \frac{p(A_j) p(y^o \mid A_j)}{p(y^o \mid A)} = \frac{p(A_j) p(y^o \mid A_j)}{\sum_{j=1}^{J} p(A_j) p(y^o \mid A_j)} \]

- For this to be operational we must know \( p(y^o \mid A_j) \)

\[ p(y^o \mid A_j) = \int_{\Theta_{A_j}} p(\theta_{A_j}, y^o \mid A_j) \, d\theta_{A_j} \]

\[ = \int_{\Theta_{A_j}} p(\theta_{A_j} \mid A_j) p(y^o \mid \theta_{A_j}, A_j) \, d\theta_{A_j}, \]

- the *marginal likelihood* of model \( A_j \) \((j = 1, \ldots, J)\).
A useful decomposition

From the relationship

\[ p(A_j | y^o, A) = \frac{p(A_j) p(y^o | A_j)}{p(y^o | A)} = \frac{p(A_j) p(y^o | A_j)}{\sum_{j=1}^J p(A_j) p(y^o | A_j)} \]

- we see that for any pair of models \((A_i, A_j)\),

\[ \frac{p(A_i | y^o)}{p(A_j | y^o)} = \frac{p(A_i)}{p(A_j)} \cdot \frac{p(y^o | A_i)}{p(y^o | A_j)}. \]

- Posterior odds ratio = Prior Odds ratio \(\times\) Bayes factor

- The marginal likelihood

\[ p(y^o | A_j) = \int_{\Theta_{A_j}} p(\theta_{A_j} | A_j) p(y^o | \theta_{A_j}, A_j) \, d\theta_{A_j} \]

contains the evidence in the data about model \(A_j\).
Prior predictive analysis

What do the first two components of a complete model

\[ p(\theta_{A_j} | A_j), \]
\[ p(y | \theta_{A_j}, A_j) \]

say about any specified function of observables \( g(y_1, \ldots, y_T) \)?

We can find out:

- For \( m = 1, \ldots, M \):
  1. \( \theta^{(m)}_A \sim p(\theta_A | A) \),
  2. \( y^{(m)} \sim p(y | \theta^{(m)}_A, A) \),
  3. \( g^{(m)} = g(y^{(m)}) \).
In our VaR example:

$$g(y) = \begin{cases} 
1 & \text{if } \min_{t=1,\ldots,T} y_t \leq -0.20 \\
0 & \text{if } \min_{t=1,\ldots,T} y_t > -0.20 
\end{cases}$$

Then $M^{-1} \sum_{m=1}^{M} g\left(y^{(m)}\right)$ approximates

$$P\left(\min_{t=1,\ldots,T} y_t \leq -0.20 \mid A\right).$$
Decision making

<table>
<thead>
<tr>
<th>Economics</th>
<th>Decision Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>State of nature $x$</td>
<td>Vector of interest $\omega$</td>
</tr>
<tr>
<td>Decision $d$</td>
<td>Action $a$</td>
</tr>
<tr>
<td>Utility function $U(d, x)$</td>
<td>Loss function $L(a, \omega)$</td>
</tr>
<tr>
<td>$\hat{d} = \arg\max E[U(d, x)]$</td>
<td>$\hat{a} = \arg\min E[L(a, \omega)]$</td>
</tr>
</tbody>
</table>

$\hat{a} = \arg\min E[L(a, \omega) | \text{What?}]$
\[
\hat{a} = \arg \min_a E \left[ L(a, \omega) \mid \text{What?} \right]
\]

- The proper conditioning is the relevant available information and the working assumptions:

\[
E \left[ L(a, \omega) \mid y^o, A \right] = \int_\Omega L(a, \omega) p(\omega \mid y^o, A) \, d\omega
= \int_\Omega \int_{\Theta_A} L(a, \omega) p(\theta_A \mid y^o, A) p(\omega \mid \theta_A, y^o, A) \, d\theta_A \, d\omega.
\]

- Simulation can help us here:

\[
\omega^{(m)} \sim p(\omega \mid y^o, A) \quad (m = 1, \ldots, M)
\implies M^{-1} \sum_{m=1}^M L\left(a, \omega^{(m)}\right) \xrightarrow{a.s.} E \left[ L(a, \omega) \mid y^o, A \right].
\]
An Introduction to Contemporary Bayesian Thinking and Methods

Decision making

Solving decision problems with simulation

\[
M^{-1} \sum_{m=1}^{M} L \left( a, \omega^{(m)} \right) \overset{a.s.}{\rightarrow} E \left[ L (a, \omega) \mid y^o, A \right]
\]

- **Binary decisions**: If \( a = 0 \) or \( 1 \), then simulate

  \( \omega^{(m)} \sim p (\omega \mid y^o, A) \) until it is clear whether or not

  \[
  E \left[ L (a = 1, \omega) - L (a = 0, \omega) \mid y^o, A \right]
  \]

  is positive or negative.

- **Continuous decisions**: If \( a \) is continuous and \( L \) is differentiable, then make use of

  \[
  M^{-1} \sum_{m=1}^{M} \frac{\partial L \left( a, \omega^{(m)} \right)}{\partial a} \overset{a.s.}{\rightarrow} \frac{\partial E \left[ L (a, \omega) \mid y^o, A \right]}{\partial a}
  \]

  in the usual way to find

  \[
  \hat{a}^{(M)} = \arg \min_a \left[ M^{-1} \sum_{m=1}^{M} L \left( a, \omega^{(m)} \right) \right].
  \]
The pitch

- Recent developments in applied Bayesian econometrics
  - Late 1990’s: Marketing
  - Early 00’s: Macroeconomics (in particular DSGE’s)
  - My prediction for late 00’s: Microeconomics, in particular choice modeling and IO

- Strategic thinking for graduate students at the dissertation stage
  - Contemporary Bayesian econometrics makes you very useful in all areas of economics
  - Recently / currently academic departments are research institutions are awakening to the need for this skill.
  - The demand will be filled mainly from new production, not out of inventory.
The next steps ...

- See that it works
- Learn how to do it
- Really understand why it works.

Some recent textbooks for economists