Bayesian Estimation

Introduction and Priors

Tools for Macroeconomists: The essentials

Petr Sedláček

Overview for today

Introduction into Bayesian estimation

- the basic ideas
- extra information over ML: priors
- main challenge: evaluating the posterior
  - Markov Chain Monte Carlo (MCMC)
- practical issues: acceptance rate, diagnostics
- implementation in Dynare
- extensions

Calibration example

- main idea and common pitfalls

Bayesian Estimation: The Basics

Frequentist vs. Bayesian views

Frequentist view:

Frequentist vs. Bayesian views

Frequentist view:

- parameters are fixed, but unknown
- likelihood is a sampling distribution for the data
- realizations of observables $\mathcal{Y}^T$
    - just one of many possible realizations from $\mathcal{L}(\mathcal{Y}^T|\Psi)$
- inferences about Ψ
    - based on probabilities of particular $\mathcal{Y}^T$ for given Ψ

Frequentist vs. Bayesian views

Bayesian view:

- observations, not parameters, are taken as given
- Ψ are viewed as random
- inference about Ψ
  - based on probabilities of Ψ conditional on data $\mathcal{Y}^T$ $P(\Psi|\mathcal{Y}^T)$
- probabilistic view of Ψ enables incorporation of prior beliefs

Bayesian Estimation: The Basics

Bayes' Rule

Bayes' rule

Rev. T. Bayes

1701-1761

Joint density of the data and parameters is:

$$P(A, B) = P(A|B)P(B) \quad \text{or}$$
$$P(A, B) = P(B|A)P(A)$$

You can rewrite the above as

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayes' rule

Joint density of the data and parameters is:

$$
\begin{aligned}
P(A, B) =& P(A|B)P(B) \quad \text{or} \\
P(A, B) =& P(B|A)P(A)
\end{aligned}
$$

You can rewrite the above as

$$
P(\Psi|\mathcal{Y}^T) = \frac{P(\mathcal{Y}^T|\Psi)P(\Psi)}{P(\mathcal{Y}^T)}
$$

Bayes' rule

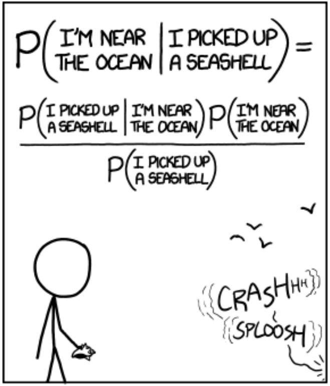
Joint density of the data and parameters is:

$$
\begin{aligned}
P(A,B) =& P(A|B)P(B) \quad \text{or} \\
P(A,B) =& P(B|A)P(A)
\end{aligned}
$$

You can rewrite the above as

$$P(\Psi|\mathcal{Y}^T) = \frac{\mathcal{L}(\mathcal{Y}^T|\Psi)P(\Psi)}{P(\mathcal{Y}^T)}$$

Bayes' rule



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Elements of Bayes' rule

- what we're interested in, posterior distribution: $P(\Psi|\mathcal{Y}^T)$
- likelihood of the data: $\mathcal{L}(\mathcal{Y}^T|\Psi)$
- our prior about the parameters: $P(\Psi)$
- probability of the data: $P(\mathcal{Y}^T)$
    - for the distribution of Ψ $P(\mathcal{Y}^T)$ is just a constant

$$P(\Psi|\mathcal{Y}^T) \propto \mathcal{L}(\mathcal{Y}^T|\Psi)P(\Psi)$$

Bayesian Estimation: The Basics

What is the Challenge?

What is the challenge?

- maximizing the posterior is typically not so problematic
- problem is that we often want to know more:
  - conditional expected values of some *function* of the posterior
  - like mean, variance, “highest-density-intervals” etc.

What is the challenge?

$$\mathbb{E}[g(\Psi)] = \frac{\int g(\Psi)P(\Psi|\mathcal{Y}^T)d\Psi}{\int P(\Psi|\mathcal{Y}^T)d\Psi}$$

- $\mathbb{E}[g(\Psi)]$ is the weighted average of $g(\Psi)$
- weights are determined by the data (likelihood) and the prior
- function $g(\Psi)$ can be all kinds of things
  - mean, variance, skewness etc.
  - marginal predictive densities: $g(\Psi)$ is an indicator function
    - equals 1 only in a certain range of the support and 0 elsewhere

What is the challenge?

What is the challenge?

Special/Simple case:

- we are able to draw Ψ from $P(\Psi|\mathcal{Y}^T)$
  - can evaluate integral via Monte Carlo integration
- you won't be lucky enough to experience this case

Our situation:

- we can calculate $P(\Psi|\mathcal{Y}^T)$, but we cannot *draw* from it

What do you mean "cannot draw from it"?!

```
function R = randn(s, m, n, varargin)
%RANDN Pseudorandom numbers from a standard normal distribution.
%   R = RANDN(S,N) returns an N-by-N matrix containing pseudorandom values drawn
%   from the standard normal distribution.  RANDN draws those values from the
%   random stream S.  RANDN(S,M,N) or RANDN(S,[M,N]) returns an M-by-N matrix.
%   RANDN(S,M,N,P,...) or RANDN(S,[M,N,P,...]) returns an M-by-N-by-P-by-...
%   array.  RANDN(S) returns a scalar.  RANDN(S,SIZE(A)) returns an array the
%   same size as A.
%
%   Note: The size inputs M, N, P, ... should be nonnegative integers.
%   Negative integers are treated as 0.
%
%   R = RANDN(..., 'double') or R = RANDN(..., 'single') returns an array of
%   normal values of the specified class.
%
%   The sequence of numbers produced by RANDN is determined by the internal
%   state of the random stream S.  RANDN uses one or more uniform values from S
%   to generate each normal value.  Control S using its RESET method and its
%   properties.
%
%   See also RANDN, RANDSTREAM, RANDSTREAM/RAND, RANDSTREAM/RANDI.

%   Copyright 2008-2015 The MathWorks, Inc.

if nargin < 2
   R = builtin('_RandStream_randn',s);
elseif nargin < 3
   R = builtin('_RandStream_randn',s,m);
elseif nargin < 4
   R = builtin('_RandStream_randn',s,m,n);
else
   R = builtin('_RandStream_randn',s,m,n,varargin{1:end});
end
```

What do you mean “cannot draw from it”?!

```
%BUILTIN  Execute built-in function from overloaded method.
%   BUILTIN is used in methods that overload built-in functions to execute
%   the original built-in function. If F is the name of a built-in function,
%   specified as a character vector or string scalar, then BUILTIN(F,x1,...,xn)
%   evaluates that function at the given arguments.
%
%   BUILTIN(...) is the same as FEVAL(...) except that it will call the
%   original built-in version of the function even if an overloaded one
%   exists (for this to work, you must never overload BUILTIN).
%
%   [y1,..,yn] = BUILTIN(F,x1,...,xn) returns multiple output arguments.
%
%   See also FEVAL.

%   Copyright 1984-2017 The MathWorks, Inc.
%   Built-in function.
```

What do you mean “cannot draw from it”?!

3

Link

Answer by MathWorks Support Team STAFF on 4 Oct 2017

Accepted Answer

A built-in function is part of the MATLAB executable. MATLAB does not implement these functions in the MATLAB language. Although most built-in functions have a .m file associated with them, this file only supplies documentation for the function. The source code detailing the implementation of built-in functions is confidential and proprietary information of MathWorks, Inc. We cannot send the source code to customers. However, some details about the algorithms used for many of these functions, including EIG and FFT, are available in the help documentation. Please type:

What is the challenge?

Special/Simple case:

- we are able to draw Ψ from $P(\Psi|\mathcal{Y}^T)$
  - can evaluate integral via Monte Carlo integration
- you won't be lucky enough to experience this case

Our situation:

- we can calculate $P(\Psi|\mathcal{Y}^T)$, but we cannot *draw* from it
  - numerical integration
  - Markov Chain Monte Carlo (MCMC) integration

Priors

Idea of priors

- summarize prior information
  - previous studies
  - data not used in estimation
  - pre-sample data
  - other countries etc.
- more on prior selection in “extensions”

Prior predictive analysis

- always check whether priors “make sense”
- use the prior as the posterior
  - steady state?
  - impulse response functions?
- more details in “extensions”
  - do certain combinations of priors make sense?

Priors

What prior distribution to pick?

- there are established choices - depend on parameters
  - beta, support $\in [0,1]$: e.g. persistence
  - (inverted-) gamma, support $\in (0,\infty)$: e.g. volatility
  - normal, support $\in (-\infty,\infty)$: where we don't really know
- uniform often chosen as "uninformative" prior
  - not so simple

Priors

Some Terminology

Some terminology

1749-1827

Some terminology

- Jeffreys prior
  - non-informative prior
- improper vs. proper priors
  - improper prior is non-integrable (integral is ∞)
  - important to have proper distributions for model comparison

Some terminology

- (natural) conjugate priors
  - family of prior distributions
  - after multiplication with the likelihood
  - produce a posterior of the same family
- Minnesota (Litterman) prior
  - used in VARs for distribution of lags

Taking Stock

Taking Stock

Bayesian estimation

- centered around Bayes' rule
- key difficult is evaluating the posterior distribution
- priors summarize addition, prior, information (on top of the data)


That's all Folks!