Introduction to Bayesian Analysis in Stata

Gustavo Sánchez

StataCorp LLC

January 24, 2019
College Station, Texas
Outline

1. Bayesian analysis: Basic concepts
   - The general idea
   - The method

2. The Stata tools
   - The general command `bayesmh`
   - The `bayes` prefix
   - Postestimation commands

3. A few examples
   - Probit regression
   - Panel data random-effects Poisson model
   - Change-point model
The general idea

Frequentist

Theoretical Model
The general idea

Bayesian analysis

Outline
General idea
The method
Fundamental equation
MCMC
Stata tools
bayes: - bayesmh
Postestimation
Examples
1- Probit regression
bayestats ess
bayesgraph
bayestestmodel
2- Random-effects Poisson
bayesgraph
bayestest interval
3- Change-point model
Gibbs sampling
Summary
References
### Frequentist Analysis

- Estimates unknown fixed parameters.
- The data come from a random sample (hypothetical repeatable).
- Uses data to estimate unknown fixed parameters.
- Data expected to satisfy the assumptions for the specified model.

"Conclusions are based on the distribution of statistics derived from random samples, assuming unknown but fixed parameters."

### Bayesian Analysis

- Probability distributions for unknown random parameters.
- The data are fixed.
- Combines data with prior beliefs to get updated probability distributions for the parameters.
- Posterior distribution is used to make explicit probabilistic statements.

"Bayesian analysis answers questions based on the distribution of parameters conditional on the observed sample."
Stata’s convenient syntax: \texttt{bayes:}

\begin{verbatim}
regress y x1 x2 x3
bayes: regress y x1 x2 x3
logit y x1 x2 x3
bayes: logit y x1 x2 x3
mixed y x1 x2 x3 || region:
bayes: mixed y x1 x2 x3 || region:
\end{verbatim}
Bayesian analysis

Outline
General idea

The method
Fundamental equation
MCMC

Stata tools
bayes: - bayesmh
Postestimation

Examples
1- Probit regression
bayesstats ess bayesgraph bayestestmodel

2- Random-effects Poisson
bayesgraph bayestest interval

3- Change-point model
Gibbs sampling

Summary

References
The method

- Inverse law of probability (Bayes’ Theorem):

\[ p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{f(y;\theta)\pi(\theta)}{f(y)} \]

Where:

- \( f(y;\theta) \): probability density function for \( y \) given \( \theta \).
- \( \pi(\theta) \): prior distribution for \( \theta \).

- The marginal distribution of \( y \), \( f(y) \), does not depend on \( \theta \); then we can write the fundamental equation for Bayesian analysis:

\[ p(\theta|y) \propto L(\theta|y)\pi(\theta) \]

Where:

- \( L(\theta|y) \): likelihood function of the parameters given the data.
The method

- Inverse law of probability (Bayes’ Theorem):

\[ p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{f(y; \theta)\pi(\theta)}{f(y)} \]

Where:

- \( f(y; \theta) \): probability density function for \( y \) given \( \theta \).
- \( \pi(\theta) \): prior distribution for \( \theta \)

- The marginal distribution of \( y \), \( f(y) \), does not depend on \( \theta \); then we can write the fundamental equation for Bayesian analysis:

\[ p(\theta|y) \propto L(\theta|y)\pi(\theta) \]

Where:

- \( L(\theta|y) \): likelihood function of the parameters given the data.
The method

- Inverse law of probability (Bayes’ Theorem):

\[ p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{f(y;\theta)\pi(\theta)}{f(y)} \]

Where:

- \( f(y;\theta) \): probability density function for \( y \) given \( \theta \).
- \( \pi(\theta) \): prior distribution for \( \theta \)

- The marginal distribution of \( y \), \( f(y) \), does not depend on \( \theta \); then we can write the fundamental equation for Bayesian analysis:

\[ p(\theta|y) \propto L(\theta|y)\pi(\theta) \]

Where:

- \( L(\theta|y) \): likelihood function of the parameters given the data.
The method

- Let’s assume that both the data and the prior beliefs are normally distributed:
  
  - **The data**: \( y \sim N\left(\theta, \sigma^2_d\right) \)
  
  - **The prior**: \( \theta \sim N\left(\mu_p, \sigma^2_p\right) \)

- **Homework...**: Doing the algebra with the fundamental equation, we find that the posterior distribution would be normal with (see for example Cameron & Trivedi 2005):
  
  - **The posterior**: \( \theta | y \sim N\left(\mu, \sigma^2\right) \)

Where:

\[
\begin{align*}
\mu &= \sigma^2 \left( N\bar{y} / \sigma^2_d + \mu_p / \sigma^2_p \right) \\
\sigma^2 &= \left( N / \sigma^2_d + 1 / \sigma^2_p \right)^{-1}
\end{align*}
\]
Example (Prior distributions)

Bayesian analysis

Outline

General idea

The method

Fundamental equation

MCMC

Stata tools

bayes: - bayesmh

Postestimation

Examples

1- Probit regression

bayesstats ess

bayesgraph

bayestestmodel

2- Random-effects Poisson

bayesgraph

bayestest interval

3- Change-point model

Gibbs sampling

Summary

References
Example (Posterior distributions)

Posterior density distribution according to Expert1

Posterior density distribution according to Expert2
The method

- The previous example has a closed form solution.

- What about the cases with non-closed solutions, or more complex distributions?
  - Integration is performed via simulation.
  - We need to use intensive computational simulation tools to find the posterior distribution in most cases.
  - Markov chain Monte Carlo (MCMC) methods are the current standard in most software. Stata implements two alternatives:
    - Metropolis–Hastings (MH) algorithm
    - Gibbs sampling
The method

- Links for Bayesian analysis and MCMC on our YouTube channel:
  - Introduction to Bayesian statistics, part 1: The basic concepts
    https://www.youtube.com/watch?v=0F0QoMCSKJ4&feature=youtu.be
    https://www.youtube.com/watch?v=OTO1DygELpY&feature=youtu.be
The method

- Monte Carlo Simulation
The method

- Metropolis–Hastings simulation
  - The trace plot illustrates the sequence of accepted proposal states.
The method

- We expect to obtain a stationary sequence when convergence is achieved.
The method

- An efficient MCMC should have small autocorrelation.
- We expect autocorrelation to become negligible after a few lags.
The Stata tools for Bayesian regression
The Stata tools: `bayes: bayesmh`

- **`bayes`:** Convenient syntax for Bayesian regressions
  - Estimation command defines the likelihood for the model.
  - Default priors are assumed to be "weakly informative".
  - Other model specifications are set by default depending on the model defined by the estimation command.
  - Alternative specifications may need to be evaluated.

- **`bayesmh`:** General purpose command for Bayesian analysis
  - You need to specify all the components for the Bayesian regression: likelihood, priors, hyperpriors, blocks, etc.
The Stata tools: Postestimation commands

- bayesstats ess
- bayesgraph
- bayesstats ic
- bayestest model
- bayestest interval
- bayesstats summary
Examples
Example 1: Probit regression

- Let’s look at our first example:
  - We have stats on scores, strength of schedule, and bowl game result (win/loss) for the Texas A&M University football team.
  - We fit a probit model for the probability to win the bowl game.
  - Let’s consider a couple of model specifications for a binary dependent variable, whose values depend on a linear latent variable:

\[
\begin{align*}
\text{win}_\text{bowl}^* &= \alpha_1 + \beta_{sc\text{-dif}} \times \text{score\_dif} + \beta_{sos} \times \text{sos} + \epsilon_1 \\
\text{win}_\text{bowl}^* &= \alpha_2 + \beta_{scored} \times \text{score\_avg} + \beta_{against} \times \text{against\_avg} + \epsilon_2
\end{align*}
\]

\[
\text{win}_\text{bowl} = \begin{cases} 
1 & \text{if } \text{win}_\text{bowl}^* > 0 \\
0 & \text{otherwise}
\end{cases}
\]

Where:
- \text{win\_bowl}: result in the bowl game (win/loss).
- \text{score\_dif}: Average score difference during the regular season.
- \text{sos}: Strength of schedule.
- \text{score\_avg}: Average points scored during the regular season.
- \text{against\_avg}: Average points against during the regular season.
Example 1: Probit regression

Example: Probit regression
Example 1: Probit regression

- Probit regression with the `bayes: prefix`
  ```stata
  bayes, rseed(123): probit win_bowl score_diff sos
  ```

- Equivalent model with `bayesmh`
  ```stata
  bayesmh win_bowl score_diff sos, rseed(123) ///
  likelihood(probit) ///
  prior({win_bowl:score_diff}, normal(0,10000)) ///
  prior({win_bowl:sos}, normal(0,10000)) ///
  prior({win_bowl:_cons}, normal(0,10000))
  ```
Example 1: Menu for Bayesian regression
Example 1: Menu for Bayesian regression
Example 1: Menu for Bayesian regression

1. Make the following sequence of selection from the main menu:
   Statistics > Bayesian analysis > Regression models

2. Select "Binary outcomes"

3. Select "Probit regression"

4. Click on "Launch"

5. Specify the dependent variable (win_bowl) and the explanatory variables (score_dif sos)

6. Click on "OK"
Example 1: **bayes: prefix**

```
.bayes, rseed(123):probit win_bowl score_dif sos
```

**Burn-in ...**

**Simulation ...**

**Model summary**

**Likelihood:**
```
  win_bowl ~ probit(xb_win_bowl)
```

**Prior:**
```
  {win_bowl:score_dif sos _cons} ~ normal(0,10000)
```

(1) Parameters are elements of the linear form `xb_win_bowl`. 

---

**Outline**

- General idea
- The method
  - Fundamental equation
  - MCMC
- Stata tools
  - `bayes`: - `bayesmh`
  - Postestimation
- Examples
  1- Probit regression
     - bayestestmodel
  2- Random-effects Poisson
     - bayesgraph
     - bayestest interval
  3- Change-point model
     - Gibbs sampling
- Summary
- References
Example 1: **bayes**: prefix

`. bayes, rseed(123): probit win_bowl score_dif sos`

Bayesian probit regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 14
Acceptance rate = .2522
Efficiency: min = .06504
avg = .07364
max = .07973

Log marginal likelihood = -25.89144

<table>
<thead>
<tr>
<th>win_bowl</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>score_dif</td>
<td>.1722847</td>
<td>.1011987</td>
<td>.003668</td>
<td>.1633205</td>
<td>.0064462</td>
</tr>
<tr>
<td>sos</td>
<td>.0797042</td>
<td>.2138371</td>
<td>.007573</td>
<td>.0882321</td>
<td>-.3346481</td>
</tr>
<tr>
<td>_cons</td>
<td>-2.08378</td>
<td>1.128949</td>
<td>.044266</td>
<td>-2.033869</td>
<td>-4.501485</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Example 1: \texttt{bayesstats ess}

- Let’s evaluate the effective sample size.

```
. bayesstats ess
Efficiency summaries       MCMC sample size =   10,000

<table>
<thead>
<tr>
<th>winbowl</th>
<th>ESS</th>
<th>Corr. time</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>score_dif</td>
<td>761.28</td>
<td>13.14</td>
<td>0.0761</td>
</tr>
<tr>
<td>sos</td>
<td>797.34</td>
<td>12.54</td>
<td>0.0797</td>
</tr>
<tr>
<td>_cons</td>
<td>650.45</td>
<td>15.37</td>
<td>0.0650</td>
</tr>
</tbody>
</table>
```

- We expect to have an acceptance rate (see previous slide) that is neither too small nor too large.
- We also expect to have low correlation.
- Efficiencies over 10\% are considered good for MH. Efficiencies under 1\% would be a source of concern.
Example 1: *bayesgraph*

- We can use *bayesgraph* to look at the trace, the correlation, and the density. For example:

  . *bayesgraph diagnostic {sos}*

- The trace indicates that convergence was achieved.
- Correlation dies out after around 10 periods.
Example 1: *bayesgraph*

- We can use `bayesgraph` to look at the trace, the correlation, and the density. For example:

  ```
  . bayesgraph diagnostic {_cons}
  ```

- Correlation dies out after around 15 periods.
Example 1: `bayestest model`

- `bayestest model` is another postestimation command to compare different models.
- `bayestest model` computes the posterior probabilities for each model.
- The result indicates which model is more likely.
- It requires that the models use the same data and that they have proper posterior.
- It can be used to compare models with:
  - Different priors and/or different posterior distributions.
  - Different regression functions.
  - Different covariates.
- MCMC convergence should be verified before comparing the models.
Example 1: bayestest model

- Let’s fit two other models and compare them with the one we already fit.

- We store the results for the three models, and we use the postestimation command bayestest model to select one of them.

  quietly {
    bayes , rseed(123) saving(dif_sos,replace): ///
    probit winbowl score_dif sos
    estimates store dif_sos

    bayes , rseed(123) saving(score,replace): ///
    probit winbowl scored_avg against_avg
    estimates store scored_against

    bayes , rseed(123) saving(srs_linear,replace) ///
    prior({winbowl:srs}, normal(10,20)):
    block({winbowl:srs _cons}):
    regress winbowl srs
    estimates store srs_linear
  }

  bayestest model dif_sos scored_against srs_linear
Example 1: bayestest model

- Here is the output for bayestest model

  . quietly {
  . bayestest model dif_sos scored_against srs_linear
  }

Bayesian model tests

|                | log(ML)  | P(M)   | P(M|y)  |
|----------------|----------|--------|--------|
| dif_sos        | -25.9158 | 0.3333 | 0.3679 |
| scored_against | -26.7528 | 0.3333 | 0.1593 |
| srs_linear     | -25.6652 | 0.3333 | 0.4727 |

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.

- We could also assign different priors for the models:

  . bayestest model dif_sos scored_against srs_linear, ///
  prior(.3 .5 .2)

Bayesian model tests

|                | log(ML)  | P(M)   | P(M|y)  |
|----------------|----------|--------|--------|
| dif_sos        | -25.9158 | 0.3000 | 0.3879 |
| scored_against | -26.7528 | 0.5000 | 0.2799 |
| srs_linear     | -25.6652 | 0.2000 | 0.3322 |

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
Example 1: `bayestest model`

- Here is the output for `bayestest model`

  ```
  . quietly {
    . bayestest model dif_sos scored_against srs_linear
  }
  Bayesian model tests

  |          | log(ML) | P(M) | P(M|y) |
  |----------|---------|------|--------|
  | dif_sos  | -25.9158 | 0.3333 | 0.3679 |
  | scored_against | -26.7528 | 0.3333 | 0.1593 |
  | srs_linear | -25.6652 | 0.3333 | 0.4727 |

  Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
  ```

- We could also assign different priors for the models:

  ```
  . bayestest model dif_sos scored_against srs_linear, ///
  prior(.3 .5 .2)
  ```

  ```
  Bayesian model tests

  |          | log(ML) | P(M) | P(M|y) |
  |----------|---------|------|--------|
  | dif_sos  | -25.9158 | 0.3000 | 0.3879 |
  | scored_against | -26.7528 | 0.5000 | 0.2799 |
  | srs_linear | -25.6652 | 0.2000 | 0.3322 |

  Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
  ```
Example 2: Random-effects Poisson model
Example 2: Random-effects Poisson model

- Let’s use `bayes:` to fit a random-effects Poisson model for a count dependent variable.

\[
Pr(y_{it} = y|x_{it}, \alpha_i) = \frac{e^{-\mu_{it}} \mu_{it}^y}{y!}
\]

Where:

\[
\mu_{i,t} = \exp(x_{i,t}\beta + \alpha_i)
\]

\[
\alpha_i \sim N(0, \sigma^2_{\alpha})
\]

is the individual panel random effect.

- This is also referred to as a two-level random intercept model.
- We can also fit this model with `mepoisson` or `xtpoisson, re normal`. 
Example 2: Random-effects Poisson model

- This time we are going to work with simulated data.
- Here is the code to simulate the panel dataset:

```stata
clear
set obs 300
set seed 123

*Panel level*
generate id = _n
generate alpha = rnormal(0,.33)

*Observation level*
expand 5
bysort id:generate year = _n
xtset id year
generate x1 = rnormal()
generate x2 = runiform()
generate x3 = rnormal()

*Generate dependent variable*
generate y = rpoisson(exp(.1*x1-.1*x2+.1*x3+.75+alpha))
```
Example 2: Random-effects Poisson model

Let’s show the results with `mepoisson`:

```
. mepoisson y x1 x2 x3 || id:, nolog
Mixed-effects Poisson regression
Group variable: id
Obs per group:
   min = 5
   avg = 5.0
   max = 5
Integration method: mvaghermite
Integration pts. = 7
Log likelihood = -2646.5534
Wald chi2(3) = 68.33
Prob > chi2 = 0.0000

|    | Coef.  | Std. Err. |     z  | P>|z| | [95% Conf. Interval] |
|----|--------|-----------|--------|------|---------------------|
| y  |        |           |        |      |                     |
| x1 | 0.0806 | 0.0192    | 4.18   | 0.000| 0.0428 - 0.1184     |
| x2 | -0.113 | 0.0652   | -1.74 | 0.082| -0.241 - 0.0144    |
| x3 | 0.1285 | 0.0187   | 6.86   | 0.000| 0.0919 - 0.1653    |
| _cons | 0.7373 | 0.0416 | 17.72 | 0.000| 0.6558 - 0.8193   |
| id |        |           |        |      |                     |
| var(_cons) | 0.1087 | 0.0171 | 6.28  | 0.000| 0.0799 - 0.148    |

LR test vs. Poisson model: chibar2(01) = 116.41
Prob >= chibar2 = 0.0000
```
Example 2: Random-effects Poisson model

- We now fit the model with `bayes`:
  
  \[
  \text{bayes, nodots rseed(123): ///}
  \text{mepoisson y x1 x2 x3 || id:}
  \]

- Equivalent model with `bayesmh`
  
  \[
  \text{bayesmh y x1 x2 x3, rseed(123) ///}
  \text{likelihood(poisson) reffects(id) ///}
  \text{prior({y:x1 x2 x3 _cons}, normal(0,10000)) ///}
  \text{prior({y:i.id}, normal(0,{sigma2})) ///}
  \text{prior({sigma2}, igamma(.01,.01)) ///}
  \text{block({sigma2}) nodots}
  \]
Example 2: Random-effects Poisson model

. bayes, nodots rseed(123) : ///
  > mepoisson y x1 x2 x3 || id:

Burn-in . . .
Simulation . . .
Multilevel structure

id
  {U0}: random intercepts

Model summary

Likelihood:
  y ~ mepoisson(xb_y)

Priors:
  {y:x1 x2 x3 _cons} ~ normal(0,10000)
    {U0} ~ normal(0,{U0:sigma2})

Hyperprior:
  {U0:sigma2} ~ igamma(.01,.01)

(1) Parameters are elements of the linear form xb_y.
Example 2: Random-effects Poisson model

```stata
.bayes, nodots rseed(123) : ///
>    mepoisson y x1 x2 x3 || id:
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1</td>
<td>.0810731</td>
<td>.0192223</td>
<td>.000803</td>
<td>.0805926</td>
<td>.0448467 .1195346</td>
</tr>
<tr>
<td>x2</td>
<td>-.1137537</td>
<td>.0648044</td>
<td>.003071</td>
<td>-.1128703</td>
<td>-.2428485 .0164924</td>
</tr>
<tr>
<td>x3</td>
<td>.1296011</td>
<td>.0183267</td>
<td>.00082</td>
<td>.1294387</td>
<td>.0931207 .167355</td>
</tr>
<tr>
<td>_cons</td>
<td>.7368688</td>
<td>.0427745</td>
<td>.002624</td>
<td>.7378466</td>
<td>.6528039 .8186462</td>
</tr>
<tr>
<td>id</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U0:sigma2</td>
<td>.1099352</td>
<td>.0177164</td>
<td>.001096</td>
<td>.1093387</td>
<td>.0765145 .1469857</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Example 2: Random-effects Poisson model

.bayesstats ess
Efficiency summaries       MCMC sample size =    10,000

<table>
<thead>
<tr>
<th></th>
<th>ESS</th>
<th>Corr. time</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1</td>
<td>572.89</td>
<td>17.46</td>
<td>0.0573</td>
</tr>
<tr>
<td>x2</td>
<td>445.22</td>
<td>22.46</td>
<td>0.0445</td>
</tr>
<tr>
<td>x3</td>
<td>499.81</td>
<td>20.01</td>
<td>0.0500</td>
</tr>
<tr>
<td>_cons</td>
<td>265.72</td>
<td>37.63</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

id
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>U0:sigma2</td>
<td>261.41</td>
<td>38.25</td>
<td>0.0261</td>
</tr>
</tbody>
</table>
Example 2: `bayesgraph diagnostic`

- We can look at the diagnostic graph for a couple of variables:

  `. bayesgraph diagnostic {y:x1}`

- The trace seems to indicate convergence.
- Autocorrelation becomes negligible after about 15 periods.
- Densities are similar for first and second halves of the MCMC sample.
**Example 2: bayesgraph diagnostic**

- We now look at the diagnostic graphs for \{U0: sigma2\}

```
.bayesgraph diagnostic {U0: sigma2}
```

- The trace seems to indicate convergence.
- Autocorrelation is slightly high, but decays steadily.
- Densities are similar for first and second halves of the MCMC sample.
Example 2: bayestest interval

- We can perform interval testing with the postestimation command bayestest interval.
- It estimates the probability that a model parameter lies in a particular interval.
- For continuous parameters, the hypothesis is formulated in terms of intervals.
- We can perform point hypothesis testing only for parameters with discrete posterior distributions.
- bayestest interval estimates the posterior distribution for a null hypothesis about intervals for one or more parameters.
- bayestest interval reports the estimated posterior mean probability for Ho.

bayestest interval ({y:x1},lower(.08) upper(.12)) ///
({y:x2},lower(-.12) upper(-.09))
**Example 2:** `bayestest interval`

- We can, for example, perform separate tests for different parameters:

  ```stata
  . bayestest interval ({y:x1},lower(.08) upper(.12)) ///
    > ({y:x2},lower(-.12) upper(-.09))
  ```

  Interval tests MCMC sample size = 10,000

<table>
<thead>
<tr>
<th>prob1</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.4909</td>
<td>0.49994</td>
<td>.0199632</td>
</tr>
<tr>
<td>prob2</td>
<td>.1926</td>
<td>0.39436</td>
<td>.0145117</td>
</tr>
</tbody>
</table>

  - If we draw $\theta_1$ from the specified prior and we use the data to update the knowledge about $\theta_1$, then there is a 49% chance that $\theta_1$ belongs to the interval (.08,.12).

- We can also perform a joint test:

  ```stata
  . bayestest interval (({y:x1},lower(.08) upper(.12)) ///
    > ({y:x2},lower(-.12) upper(-.09)),joint)
  ```

  Interval tests MCMC sample size = 10,000

<p>| prob1, .08 &lt; {y:x1} &lt; .12, -.12 &lt; {y:x2} &lt; -.09 |</p>
<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.0885</td>
<td>0.28403</td>
</tr>
</tbody>
</table>
Example 2: \texttt{bayestest interval}

- We can, for example, perform separate tests for different parameters:

\begin{verbatim}
. bayestest interval ([y:x1], lower(.08) upper(.12)) ///
> ([y:x2], lower(-.12) upper(-.09))
\end{verbatim}

\begin{verbatim}
Interval tests MCMC sample size = 10,000
prob1 : .08 < \{y:x1\} < .12
prob2 : -.12 < \{y:x2\} < -.09
\end{verbatim}

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.4909</td>
<td>0.49994</td>
<td>.0199632</td>
</tr>
<tr>
<td>prob2</td>
<td>.1926</td>
<td>0.39436</td>
<td>.0145117</td>
</tr>
</tbody>
</table>

- If we draw $\theta_1$ from the specified prior and we use the data to update the knowledge about $\theta_1$, then there is a 49% chance that $\theta_1$ belongs to the interval (.08,.12).

- We can also perform a joint test:

\begin{verbatim}
. bayestest interval (([y:x1], lower(.08) upper(.12)) ///
> ([y:x2], lower(-.12) upper(-.09), joint)
\end{verbatim}

\begin{verbatim}
Interval tests MCMC sample size = 10,000
prob1 : .08 < \{y:x1\} < .12, -.12 < \{y:x2\} < -.09
\end{verbatim}

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.0885</td>
<td>0.28403</td>
<td>.0098171</td>
</tr>
</tbody>
</table>
Example 2: Show random effects

```
. bayes, show({U0[1/6]}) noheader
```

<table>
<thead>
<tr>
<th>U0[id]</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[95% Cred. Int.]</td>
</tr>
<tr>
<td>1</td>
<td>.1005875</td>
<td>.2248611</td>
<td>.005989</td>
<td>.1137852</td>
<td>-.3503203 .5382369</td>
</tr>
<tr>
<td>2</td>
<td>-.1376598</td>
<td>.2372418</td>
<td>.006347</td>
<td>-.1312831</td>
<td>-.6391449 .3238192</td>
</tr>
<tr>
<td>3</td>
<td>.1669656</td>
<td>.2171576</td>
<td>.006349</td>
<td>.1645487</td>
<td>-.2620912 .5840191</td>
</tr>
<tr>
<td>4</td>
<td>.1415134</td>
<td>.2192747</td>
<td>.006385</td>
<td>.1401843</td>
<td>-.3075952 .5717826</td>
</tr>
<tr>
<td>5</td>
<td>-.0802774</td>
<td>.2361239</td>
<td>.007224</td>
<td>-.0747518</td>
<td>-.5665242 .3531596</td>
</tr>
<tr>
<td>6</td>
<td>.1128583</td>
<td>.2338012</td>
<td>.006719</td>
<td>.1093227</td>
<td>-.3585934 .5664554</td>
</tr>
</tbody>
</table>

Note: Default priors are used for model parameters.
Example 2: Histograms for random effects

- `bayesgraph histogram`

`bayesgraph histogram {U0[1/6]}, name(g1 g2 g3 g4 g5 g6, replace)`

`. graph combine g1 g2 g3 g4 g5 g6, ///
> title("Histograms for first 6 random effects")`
Example 2: Histograms for random effects

- `bayesgraph histogram`

  ```
  . bayesgraph histogram {U0[1/6]},name(g1 g2 g3 g4 g5 g6,replace)
  . graph combine g1 g2 g3 g4 g5 g6, ///
  >  title("Histograms for first 6 random effects")
  ```

![Histograms for first 6 random effects](image-url)
Example 3: Change-point model
Example 3: Change-point model

- Let’s work now with an example where we write our model using a substitutable expression.
- We have average oil prices for January 1986 to December 2015:

The series has a significant increase around 2005.
- We may consider fitting a change-point model.
Example 3: Gibbs sampling

Change-point model specification with blocking

\[
\text{bayesmh oilprice} = (\{\mu_1\} \times \text{sign} (\text{year} < \{\text{cp}\})) + (\{\mu_2\} \times \text{sign} (\text{year} \geq \{\text{cp}\})),
\]

\[
\text{likelihood} (\text{normal} (\{\text{var}\}))
\]

\[
\text{prior} (\{\mu_1\}, \text{normal} (0, 50))
\]

\[
\text{prior} (\{\mu_2\}, \text{normal} (50, 150))
\]

\[
\text{prior} (\{\text{cp}\}, \text{uniform} (\text{tm} (1986m1), 2015m12))
\]

\[
\text{prior} (\{\text{var}\}, \text{igamma} (0.01, 0.01))
\]

\[
\text{initial} (\{\mu_1\} = 15, \{\mu_2\} = 100, \{\text{cp}\} = \text{tm} (1986m1))
\]

\[
\text{block} (\{\text{var}\}, \text{gibbs}) \quad \text{block} (\{\text{cp}\}) \quad \text{blocksummary}
\]

\[
\text{rseed} (123) \quad \text{mcmcsizes} (40000)
\]

\[
\text{dots} (500, \text{every} (5000))
\]

quietly {
    matrix mean = e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] ///
    _n _col(17) "Cut point (Month): " %tm mean[1,1]
}

Example 3: Gibbs sampling

Change-point model specification with blocking

```
bayesmh oilprice = \{\mu_1\} \times \text{sign}(\text{year}<\{\text{cp}\})
+ \{\mu_2\} \times \text{sign}(\text{year}>={\text{cp}})),

\text{likelihood(normal(\{var\}))}
\text{prior(\{\mu_1\}, normal(0,50))}
\text{prior(\{\mu_2\}, normal(50,150))}
\text{prior(\{\text{cp}\}, uniform(tm(1986m1),2015m12))}
\text{prior(\{var\}, igamma(.01,.01))}

\text{initial(\{\mu_1\} =15 \{\mu_2\} =100 \{\text{cp}\} =tm(1986m1))}
\text{block(\{var\}, gibbs) block(\{\text{cp}\}) blocksummary}
\text{rseed(123) mcmcsizer(40000)}
\text{dots(500,every(5000))}
```

quietly {
    \begin{verbatim}
    matrix mean=e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] \\
    _n _col(17) "Cut point (Month): " \%tm mean[1,1]
    \end{verbatim}
}
Example 3: Gibbs sampling

Change-point model specification with blocking

```stata
bayesmh oilprice = (\{mu1\} * sign(year<{cp}))
+ (\{mu2\} * sign(year>={cp})),
likelihood(normal(\{var\}))
prior(\{mu1\}, normal(0,50))
prior(\{mu2\}, normal(50,150))
prior(\{cp\}, uniform(tm(1986m1),2015m12))
prior(\{var\}, igamma(.01,.01))
initial(\{mu1\} =15 \{mu2\} =100 \{cp\} =tm(1986m1))
block(\{var\}, gibbs) block(\{cp\}) blocksummary
rseed(123) mcmcsize(40000)
dots(500,every(5000))
quietly {
    matrix mean=e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] _n _col(17) "Cut point (Month): " %tm mean[1,1]
}
```

Example 3: Gibbs sampling

Change-point model specification with blocking

```stata
bayesmh oilprice = (\{mu1\} * sign(year<{cp}))
+ (\{mu2\} * sign(year>={cp})),
likelihood(normal(\{var\}))
prior(\{mu1\}, normal(0,50))
prior(\{mu2\}, normal(50,150))
prior(\{cp\}, uniform(tm(1986m1),2015m12))
prior(\{var\}, igamma(.01,.01))
initial(\{mu1\} =15 \{mu2\} =100 \{cp\} =tm(1986m1))
block(\{var\}, gibbs) block(\{cp\}) blocksummary
rseed(123) mcmcsize(40000)
dots(500,every(5000))
quietly {
    matrix mean=e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] _n _col(17) "Cut point (Month): " %tm mean[1,1]
}
```
Example 3: Gibbs sampling

Change-point model specification with blocking

```
bayesmh oilprice = ({mu1}*sign(year<{cp}) + {mu2}*sign(year>={cp})),
   likelihood(normal({var}))
   prior({mu1}, normal(0,50))
   prior({mu2}, normal(50,150))
   prior({cp}, uniform(tm(1986m1),2015m12))
   prior({var}, igamma(.01,.01))
   initial({mu1} =15 {mu2} =100 {cp} =tm(1986m1))
   block({var}, gibbs) block({cp}) blocksummary
   rseed(123) mcmcsize(40000)
   dots(500,every(5000))
```

quently {
  matrix mean=e(mean)
  noisily display _n _col(10) "Date: " mean[1,1]  
    _n _col(17) "Cut point (Month): " %tm mean[1,1]
}
```
Example 3: Gibbs sampling

Change-point model specification with blocking

bayesmh oilprice = ({mu1}*sign(year<{cp})
+ {mu2}*sign(year>={cp})),
likelihood(normal({var}))
prior({mu1}, normal(0,50))
prior({mu2}, normal(50,150))
prior({cp}, uniform(tm(1986m1),2015m12))
prior({var}, igamma(.01,.01))
initial({mu1} =15 {mu2} =100 {cp} =tm(1986m1))
block({var}, gibbs) blocksummary
rseed(123) mcmcsize(40000)
dots(500,every(5000))

quietly {
    matrix mean=e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] _n _col(17) "Cut point (Month): " %tm mean[1,1]
}

Example 3: Gibbs sampling

Change-point model specification with blocking

bayesmh oilprice = ({mu1}*sign(year<{cp})
+ {mu2}*sign(year>={cp})),
likelihood(normal({var}))
prior({mu1}, normal(0,50))
prior({mu2}, normal(50,150))
prior({cp}, uniform(tm(1986m1),2015m12))
prior({var}, igamma(.01,.01))
initial({mu1} =15 {mu2} =100 {cp} =tm(1986m1))
block({var}, gibbs) blocksummary
rseed(123) mcmcsize(40000)
dots(500,every(5000))

quietly {
    matrix mean=e(mean)
    noisily display _n _col(10) "Date: " mean[1,1] _n _col(17) "Cut point (Month): " %tm mean[1,1]
}
Example 3: Gibbs sampling

Change-point model specification with blocking

```
.bayesmh oilprice=({mu1}*sign(month<{cp})+{mu2}*sign(month>=>{cp})), ///
>   likelihood(normal({var})) ///
>   prior({mu1}, normal(0,50)) ///
>   prior({mu2}, normal(50,150)) ///
>   prior({cp}, uniform(tm(1986m1),tm(2015m12))) ///
>   prior({var}, igamma(.01,.01)) ///
>   initial({mu1} =15 {mu2} =100 {cp} =tm(1986m1) rseed(123) ///
>   block({var}, gibbs) block({cp}) blocksummary ///
>   mcmcsize(20000) dots(500, every(5000))
```

Burn-in 2500 aaaaa done
Simulation 20000 ........5000............10000............15000............20000 done

Model summary

**Likelihood:**
```
oilprice ~ normal({mu1}*sign(month<{cp})+{mu2}*sign(month>=>{cp})),{var})
```

**Priors:**
```
{var} ~ igamma(.01,.01)
{mu1} ~ normal(0,50)
{mu2} ~ normal(50,150)
{cp} ~ uniform(tm(1986m1),tm(2015m12))
```

**Block summary**
```
1: {var}  
2: {cp}   
3: {mu1} {mu2}  (Gibbs)
```
Example 3: Gibbs sampling

Change-point model specification with blocking

```
.bayesmh oilprice=({mu1}*sign(month<{cp})+{mu2}*sign(month>={cp})), likelihood(normal({var}))
  > prior({mu1}, normal(0,50))
  > prior({mu2}, normal(50,150))
  > prior({cp}, uniform(tm(1986m1),tm(2015m12))
  > prior({var}, igamma(.01,.01))
  > initial({mu1} =15 {mu2} =100 {cp} =tm(1986m1)) rseed(123)
  > block({var}, gibbs) block({cp}) blocks
mcmcsize(20000) dots(500, every(5000))
```

Bayesian normal regression

Metropolis-Hastings and Gibbs sampling

MCMC iterations = 22,500
Burn-in = 2,500
MCMC sample size = 20,000
Number of obs = 360
Acceptance rate = .5632
Efficiency: min = .09094
          avg = .3304
          max = 1

Log marginal likelihood = -1481.9487

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>Equal-tailed [95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>cp</td>
<td>541.5063</td>
<td>1.806737</td>
<td>.037169</td>
<td>541.4515</td>
<td>536.7238 to 544.9228</td>
</tr>
<tr>
<td>mu1</td>
<td>22.07432</td>
<td>.936419</td>
<td>.01974</td>
<td>22.09333</td>
<td>20.23623 to 23.85525</td>
</tr>
<tr>
<td>mu2</td>
<td>78.69139</td>
<td>1.259118</td>
<td>.029524</td>
<td>78.67589</td>
<td>76.2043 to 81.19035</td>
</tr>
<tr>
<td>var</td>
<td>197.286</td>
<td>14.80914</td>
<td>.104716</td>
<td>196.6902</td>
<td>169.991 to 228.0003</td>
</tr>
</tbody>
</table>

. quietly {

    elapsed date: 541.50629
    Cut point (Month): 2005m2

```
Example 3: `bayesgraph trace`

- Use `bayesgraph trace` to look at the trace for all the parameters.

```
.bayesgraph trace _all,combine
```

The plots indicate that convergence seems to be achieved.
Example 3: `bayesgraph ac`

- Use `bayesgraph ac` to look at the autocorrelation for all the parameters.
  
  ```stata
  . bayesgraph ac _all,combine
  ```

- Autocorrelation quickly becomes negligible for all the parameters.
Example 3: \textit{bayesgraph matrix}

- Use \texttt{bayesgraph matrix} to look at pairwise correlation for the parameters.

\begin{verbatim}
.bayesgraph matrix _all
\end{verbatim}

- The plots seem to indicate that there are no significant pairwise correlations among the parameters.
Bayesian analysis

Summing up

- Bayesian analysis: A statistical approach that can be used to answer questions about unknown parameters in terms of probability statements.

- It can be used when we have prior information on the distribution of the parameters involved in the model.

- Alternative approach or complementary approach to classic/frequentist approach?
Reference


Links

https://www.stata.com/meeting/uk17/slides/uk17_Marchenko.pdf


https://www.stata.com/meeting/spain18/slides/spain18_Sanchez.pdf