Bayesian analysis using Stata

Yulia Marchenko

Director of Biostatistics
StataCorp LP

2015 Nordic and Baltic Stata Users Group meeting
Bayesian analysis using Stata

Outline

1 Introduction
   • What is Bayesian analysis?
   • Why Bayesian analysis?
   • Components of Bayesian analysis
   • Motivating example: Beta-binomial model

2 Stata’s Bayesian suite
   • Commands
   • Graphical user interface (GUI)

3 Examples
   • Beta-binomial model (revisited)
   • Power priors
   • Model comparison
   • User-defined models: Hurdle model

4 Summary

5 Forthcoming

6 References
More examples (extra)

- Normal linear regression
- Random-intercept model
- Random-coefficient model
- Meta analysis
- Nonlinear Poisson model: Change-point analysis
- Bioequivalence in a crossover trial
Bayesian analysis is a statistical paradigm that answers research questions about unknown parameters using probability statements.
What is the probability that a person accused of a crime is guilty?

What is the probability that treatment A is more cost effective than treatment B for a specific health care provider?

What is the probability that the odds ratio is between 0.3 and 0.5?

What is the probability that three out of five quiz questions will be answered correctly by students?
You may be interested in Bayesian analysis if

- you have some prior information available from previous studies that you would like to incorporate in your analysis. For example, in a study of preterm birthweights, it would be sensible to incorporate the prior information that the probability of a mean birthweight above 15 pounds is negligible. Or,

- your research problem may require you to answer a question: What is the probability that my parameter of interest belongs to a specific range? For example, what is the probability that an odds ratio is between 0.2 and 0.5? Or,

- you want to assign a probability to your research hypothesis. For example, what is the probability that a person accused of a crime is guilty?

- And more.
Observed data sample $D$ is fixed and model parameters $\theta$ are random.

$D$ is viewed as a result of a one-time experiment.

A parameter is summarized by an entire distribution of values instead of one fixed value as in classical frequentist analysis.
There is some prior (before seeing the data!) knowledge about $\theta$ formulated as a **prior distribution** $p(\theta)$.

After data $D$ are observed, the information about $\theta$ is updated based on the **likelihood** $f(D|\theta)$.

Information is updated by using the Bayes rule to form a **posterior distribution** $p(\theta|D)$:

$$p(\theta|D) = \frac{f(D|\theta)p(\theta)}{p(D)}$$

where $p(D)$ is the **marginal distribution** of the data $D$. 
Estimating a posterior distribution \( p(\theta|D) \) is at the heart of Bayesian analysis.

Various summaries of this distribution are used for inference.

- Point estimates: posterior means, modes, medians, percentiles.
- Interval estimates: **credible intervals** (Crl)—(fixed) ranges to which a parameter is known to belong with a pre-specified probability.
- Monte-Carlo standard error (MCSE)—represents precision about posterior mean estimates.
- Hypothesis testing—assign probability to any hypothesis of interest
- Model comparison: model posterior probabilities, Bayes factors
Potential subjectivity in specifying prior information—noninformative priors or sensitivity analysis to various choices of informative priors.

Computationally demanding—involves intractable integrals that can only be computed using intensive numerical methods such as Markov chain Monte Carlo (MCMC).
Research problem

- Prevalence of a rare infectious disease in a small city (Hoff 2009)
- A sample of 20 subjects is checked for infection
- Parameter $\theta$ is the proportion of infected individuals in the city
- Outcome $y$ is the # of infected individuals in the sample
Model

- Likelihood, $f(y|\theta)$: Binomial
- Prior, $p(\theta)$: Infection rate ranged between 0.05 and 0.20, with an average prevalence of 0.10, in other similar cities
- Bayesian model:

\[
y|\theta \sim \text{Binomial}(20, \theta)
\]
\[
\theta \sim \text{Beta}(2, 20)
\]

- Posterior: $\theta|y \sim \text{Beta}(2 + y, 20 + 20 - y)$
Observed data

- We sample individuals and observe none who have an infection, $y = 0$
- Posterior: $\theta | y \sim \text{Beta}(2, 40)$
- Prior mean: $E(\theta) = 2/(2+20) = 0.09$
- Posterior mean: $E(\theta | y) = 2/(2+40) = 0.048$
- Posterior probability: $P(\theta < 0.10) = 0.93$
Introduction

Motivating example: Beta-binomial model

Prior and posterior distributions of $\theta$

Proportion infected in the population, $\theta$

$p(\theta)$ and $p(\theta|y)$
• Fit beta-binomial model using **bayesmh** (variable y has one observation equal to 0)

• **MCMC method: adaptive Metropolis-Hastings (MH)**

```plaintext
. set seed 14
. bayesmh y, likelihood(binlogit(20), noglmtransform) ///
  >     prior({y:_cons}, beta(2,20))

Model summary

Likelihood:
  y ~ binomial({y:_cons}, 20)

Prior:
  {y:_cons} ~ beta(2, 20)

Bayesian binomial regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 1
Acceptance rate = .4205
Efficiency = .1401

Log marginal likelihood = -1.1714402

Equal-tailed

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>95% Cred. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>.0466517</td>
<td>.0316076</td>
<td>.000844</td>
<td>.0391639</td>
<td>.0058112 .1260038</td>
</tr>
</tbody>
</table>
Compute posterior probability

. bayestest interval {y:_cons}, upper(0.1)
Interval tests MCMC sample size = 10,000
prob1 : {y:_cons} < 0.1

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.9299</td>
<td>0.25533</td>
<td>.006074</td>
</tr>
<tr>
<td>Command</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesmh</td>
<td>Bayesian regression using MH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesmh <em>evaluators</em></td>
<td>User-written Bayesian models using MH</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Postestimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesgraph</td>
<td>Graphical convergence diagnostics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesstats ess</td>
<td>Effective sample sizes and more</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesstats summary</td>
<td>Summary statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayesstats ic</td>
<td>Information criteria and Bayes factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayestest model</td>
<td>Model posterior probabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bayestest interval</td>
<td>Interval hypothesis testing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Models

- 10 built-in likelihoods: normal, logit, ologit, Poisson, ...
- 18 built-in priors: normal, gamma, Wishart, Zellner’s g, ...
- Continuous, binary, ordinal, and count outcomes
- Univariate, multivariate, and multiple-equation models
- Linear, nonlinear, and canonical generalized nonlinear models
- Continuous univariate, multivariate, and discrete priors
- User-defined models

MCMC methods

- Adaptive MH
- Adaptive MH with Gibbs updates—hybrid
- Full Gibbs sampling for some models
Built-in models

bayesmh ..., likelihood() prior() ...

User-defined models

bayesmh ..., {evaluator() | llevaluator() prior()} ...

Postestimation features are the same whether you use a built-in model or program your own!
Perform Bayesian analysis by using the command line
Or, use a powerful point-and-click interface
You can access the GUI by typing

. db bayesmh

or from the Statistics menu
Recall the beta-binomial model from the motivating example.
Let’s store the estimation results for future comparison.
`estimates store` requires first saving `bayesmh`’s MCMC data.
Use option `saving()` during estimation or on replay:

```
.bayesmh, saving(betabin)
note: file betabin.dta saved
.estimates store betabin
```
Check MCMC convergence

```
.bayesgraph diagnostics {y:_cons}
```
Check MCMC sampling efficiency

`. bayesstats ess {y:_cons}`

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>_cons</td>
<td>1400.87</td>
<td>7.14</td>
<td>0.1401</td>
</tr>
</tbody>
</table>
Test an interval hypothesis

. bayestest interval {y:_cons}, upper(0.1)

<table>
<thead>
<tr>
<th>prob1</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.9299</td>
<td>0.25533</td>
<td>.006074</td>
</tr>
</tbody>
</table>
Motivating example used a beta prior for $\theta$

Sensitivity analysis to the choice of the priors is very important in Bayesian analysis

Consider an alternative prior—a power prior
Based on similar historical data $y_0$

Idea: raise the likelihood function of the historical data to the power $\alpha_0$, where $0 \leq \alpha_0 \leq 1$.

$\alpha_0$ quantifies the uncertainty in $y_0$ by controlling the heaviness of the tails of the prior distribution.

$\alpha_0 = 0$ means no information from the historical data and $\alpha_0 = 1$ means that the historical data have as much weight as the observed data.
Suppose that in another similar city, a random sample of 15 subjects was selected and 1 subject had a disease.

Let’s consider a competing power prior:

\[ p(\theta) \sim \{\text{BinomialPMF}(15, 1, \theta)\}^{\alpha_0} \]

Let \( \alpha_0 = 0.5 \).
bayesmh does not have built-in power priors but we can use prior()’s suboption density() to specify our own prior.

```
. set seed 14
. bayesmh y, likelihood(binlogit(20), noglmtransform)       ///
  >     prior({y:_cons}, density(sqrt(binomialp(15,1,{y:_cons})))) ///
  >     saving(powerbin)
```

Model summary

Likelihood:
  y ~ binomial({y:_cons},20)
Prior:
  {y:_cons} ~ density(sqrt(binomialp(15,1,{y:_cons})))
Bayesian binomial regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 1
Acceptance rate = .4294
Efficiency = .1228

Log marginal likelihood = -3.4630512

<table>
<thead>
<tr>
<th>y</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>_cons</td>
<td>.0501507</td>
<td>.0392846</td>
<td>.001121</td>
<td>.0401686</td>
<td>.0040134 .1521774</td>
</tr>
</tbody>
</table>

file powerbin.dta not found; file saved
.estimates store powerbin
Compute model posterior probabilities

```
. bayestest model powerbin betabin
Bayesian model tests

|       | log(ML) | P(M)   | P(M|y)  |
|-------|---------|--------|---------|
| powerbin | -3.4631 | 0.5000 | 0.0918  |
| betabin | -1.1714 | 0.5000 | 0.9082  |
```

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
Compute the Bayes factor—the ratio of the marginal likelihoods of the two models calculated using the same data.

```
.bayesstats ic powerbin betabin
Bayesian information criteria

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>log(ML)</th>
<th>log(BF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>powerbin</td>
<td>2.129576</td>
<td>-3.463051</td>
<td>.</td>
</tr>
<tr>
<td>betabin</td>
<td>1.956201</td>
<td>-1.17144</td>
<td>2.291611</td>
</tr>
</tbody>
</table>
```

Note: Marginal likelihood (ML) is computed using Laplace-Metropolis approximation.
In addition to the many built-in models, you can also program your own models.

Program log likelihood and use one of the built-in priors:
. bayesmh ..., llevaluator(llprogname) prior() ...

Or, program the log posterior:
. bayesmh ..., evaluator(lpprogramname) ...
One of the questions we received shortly after releasing `bayesmh` is “How do I fit Bayesian hurdle models?”

A hurdle model (Cragg model) is used to model a bounded dependent variable. It combines a selection model that determines the boundary points of the dependent variable with an outcome model that determines its nonbounded values.

Hurdle models are not currently among the built-in `bayesmh` models.

But, we can program them using `bayesmh`’s used-defined evaluators.

Below I provide two types of log-likelihood evaluators: one using Stata’s command `churdle` (new in Stata 14) to compute the log likelihood and the other programming the log likelihood from scratch.
We consider a subset of the fitness data from [R] churdle.

We consider a simple linear hurdle model.

We model the decision to exercise or not as a function of an individual’s average commute to work.

Once a decision to exercise is made, we model the number of hours an individual exercises per day as a function of age.

. webuse fitness
. set seed 17653
. sample 10
(17,848 observations deleted)
We use \texttt{churdle} to compute the log-likelihood values at each MCMC iteration.

\begin{verbatim}
. program mychurdle1
  1.     version 14.0
  2.     args llf
  3.     tempname b
  4.     mat `b´ = ($MH\textunderscore b, $MH\textunderscore p)
  5.     cap churdle linear $MH\textunderscore y1 $MH\textunderscore y1x1 if $MH\textunderscore touse, ///
            select($MH\textunderscore y2x1) ll(0) from(`b´) iterate(0)

    >         scalar `llf´ = .
  6.     if _rc {
  7.         if (_rc==1) { // handle break key
  8.             exit _rc
  9.         }
 10.         scalar `llf´ = .
 11.     }
 12.     else {
 13.         scalar `llf´ = e(ll)
 14.     }
 15.     end
\end{verbatim}
. set seed 14
. gen byte hours0 = (hours==0)
. bayesmh (hours age) (hours0 commute), ///
  llevaluator(mychurdle1, parameters({lnsig})) ///
  prior({hours:} {hours0:} {lnsig}, flat) dots

Burn-in 2500 aaaaaaaaa1000aaaaaaaaaa2000aaaaa. done
> 000..........6000..........7000..........8000..........9000..........10000 done

Model summary

Likelihood:
  hours hours0 ~ mychurdle1(xb_hours,xb_hours0,{lnsig})

Priors:
  {hours:age _cons} ~ 1 (flat) (1)
  {hours0:commute _cons} ~ 1 (flat) (2)
  {lnsig} ~ 1 (flat)

(1) Parameters are elements of the linear form xb_hours.
(2) Parameters are elements of the linear form xb_hours0.
Bayesian regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 1,983
Acceptance rate = .2752
Efficiency: min = .04197
avg = .06659
max = .08861

Log marginal likelihood = -2772.4136

<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>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0051872</td>
<td>.0027702</td>
<td>.000093</td>
<td>.0052248</td>
<td>-.0002073 - .0104675</td>
</tr>
<tr>
<td>_cons</td>
<td>1.163384</td>
<td>.1219417</td>
<td>.005135</td>
<td>1.16685</td>
<td>.9203519 - 1.388663</td>
</tr>
<tr>
<td>hours0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>commute</td>
<td>-.0716184</td>
<td>.1496757</td>
<td>.005623</td>
<td>-.0758964</td>
<td>-.3733355 - .2181717</td>
</tr>
<tr>
<td>_cons</td>
<td>.1454332</td>
<td>.084041</td>
<td>.003066</td>
<td>.1451574</td>
<td>-.0222543 - .3128047</td>
</tr>
<tr>
<td>lnsig</td>
<td>.1341657</td>
<td>.034162</td>
<td>.001668</td>
<td>.1336526</td>
<td>.0634106 - .2021694</td>
</tr>
</tbody>
</table>

This model took 25 minutes
The corresponding log likelihood programmed from scratch

. program mychurdle2
1.     version 14.0
2.     args lnf xb xg lnsig
3.     tempname sig
4.     scalar `sig´ = exp(`lnsig´)
5.     tempvar lnfj
6.     qui gen double `lnfj´ = normal(`xg´) if $MH_touse
7.     qui replace `lnfj´ = log(1 - `lnfj´) if $MH_y1 <= 0 & $MH_touse
8.     qui replace `lnfj´ = log(`lnfj´) - log(normal(`xb´/`sig´)) ///
   >     + log(normalden($MH_y1,`xb´,`sig´)) ///
   >     if $MH_y1 > 0 & $MH_touse
9.     summarize `lnfj´ if $MH_touse, meanonly
10.    if r(N) < $MH_n {
11.        scalar `lnf´ = .
12.        exit
13.    }
14.    scalar `lnf´ = r(sum)
15.    end
. set seed 14
. bayesmh (hours age) (hours0 commute), ///
> llevaluator(mychurdle2, parameters({lnsig}) ) ///
> prior({hours:} {hours0:} {lnsig}, flat) dots

Burn-in 2500 aaaaaaaaa1000aaaaaaaaaa2000aaaa. done
Simulation 10000 ........1000........2000........3000........4000........5
> 000........6000........7000........8000........9000........10000 done

Model summary

Likelihood:
  hours hours0 ~ mychurdle2(xb_hours,xb_hours0,{lnsig})

Priors:
  {hours:age _cons} ~ 1 (flat) (1)
  {hours0:commute _cons} ~ 1 (flat) (2)
  {lnsig} ~ 1 (flat)

(1) Parameters are elements of the linear form xb_hours.
(2) Parameters are elements of the linear form xb_hours0.
Bayesian analysis using Stata

Examples

User-defined models: Hurdle model programmed from scratch

Bayesian regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 1,983
Acceptance rate = .2752
Efficiency: min = .04197
avg = .06659
max = .08861

Log marginal likelihood = -2772.4136

<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>hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>.0051872</td>
<td>.0027702</td>
<td>.000093</td>
<td>.0052248</td>
<td>-.0002073 to .0104675</td>
</tr>
<tr>
<td>_cons</td>
<td>1.163384</td>
<td>.1219417</td>
<td>.005135</td>
<td>1.16685</td>
<td>.9203519 to 1.388663</td>
</tr>
<tr>
<td>hours0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>commute</td>
<td>-.0716184</td>
<td>.1496757</td>
<td>.005623</td>
<td>-.0758964</td>
<td>-.3733355 to .2181717</td>
</tr>
<tr>
<td>_cons</td>
<td>.1454332</td>
<td>.084041</td>
<td>.003066</td>
<td>.1451574</td>
<td>-.0222543 to .3128047</td>
</tr>
<tr>
<td>ln sigma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This model took only 20 seconds!
Bayesian analysis is a powerful tool that allows you to incorporate prior information about model parameters into your analysis.

It provides intuitive and direct interpretations of results by using probability statements about parameters.

It provides a way to assign an actual probability to any hypothesis of interest.
Use `bayesmh` for estimation: choose one of the built-in models or program your own.

Use postestimation features for checking MCMC convergence, estimating functions of model parameters, and performing hypothesis testing and model comparison.

Work interactively using the command line or use the point-and-click interface.

Check out the “More examples” section and the `[BAYES] Bayesian analysis` manual for more examples and details about Bayesian analysis.
More computationally efficient handling of multilevel ("random-effects") models—option `reffects()` for two-level models and option `block(, reffects)` for models with more than two levels.

For example, Bayesian IRT 1PL models with more than 32,000 subjects are now feasible:

```
.bayesmh y i.item, noconstant reffects(id) likelihood(logit) ///
   >   prior({y:i.id}, normal(0, {var})) ///
   >   prior({y:i.item}, normal(0, 10)) ///
   >   prior({var}, igamma(0.01,0.01)) ///
   >   block({y:i.item}, reffects) ///
   >   block({var})
```
Straightforward specification of unstructured covariances between random-effects parameters—prior distribution `mvnormal()` is now row-column conformable.

For example,

```
.bayesmh ..., ... prior({y:i.id i.id#c.x}, mvnormal(2,{b0},{b1},{Sigma,matrix}))
```
models the unstructured covariance between random intercepts and random coefficients for `x` associated with the levels of `id`. 


Data: weight measurements of 48 pigs on 9 successive weeks (e.g., Diggle et al. (2002)).

Ignore the grouping structure of the data for now

Likelihood model:

\[
\text{weight}_{ij} = \beta_0 + \beta_1 \text{week}_{ij} + \epsilon_{ij}
\]

\[
\epsilon_{ij} \sim \text{Normal}(0, \sigma^2)
\]

where \( i = 1, \ldots, 9 \) and \( j = 1, \ldots, 48 \).

Noninformative prior distributions:

\[
\beta_i \sim \text{Normal}(0, 100), \quad i = 0, 1
\]

\[
\sigma^2 \sim \text{InvGamma}(0.001, 0.001)
\]
. webuse pig  
(Longitudinal analysis of pig weights)
. set seed 14
. bayesmh weight week, likelihood(normal({var}))      ///
    >         prior({weight:}, normal(0,100))     ///
    >         prior({var},                 igamma(0.001,0.001))

Burn-in ...  
Simulation ...  
Model summary

Likelihood:  
    weight ~ normal(xb_weight,{var})

Priors:  
    {weight:week _cons} ~ normal(0,100)               (1)
          {var} ~ igamma(0.001,0.001)

(1) Parameters are elements of the linear form xb_weight.
Bayesian normal regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 432
Acceptance rate = .2291
Efficiency: min = .0692
avg = .08122
max = .09538

Log marginal likelihood = -1270.6848

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.214291</td>
<td>.0787262</td>
<td>.002549</td>
<td>6.214297</td>
<td>6.055505 6.364085</td>
</tr>
<tr>
<td>_cons</td>
<td>19.32917</td>
<td>.4468276</td>
<td>.015889</td>
<td>19.31526</td>
<td>18.47262 20.22465</td>
</tr>
<tr>
<td>var</td>
<td>19.50327</td>
<td>1.33882</td>
<td>.050894</td>
<td>19.44994</td>
<td>17.09487 22.30596</td>
</tr>
</tbody>
</table>
Bayesian analysis using Stata

More examples (extra)

Normal linear regression: Gibbs sampling

. set seed 14
. bayesmh weight week, likelihood(normal({var})) ///
>    prior({weight:}, normal(0,100)) ///
>    prior({var},     igamma(0.001,0.001)) ///
>    block({weight:}, gibbs)  ///
>    block({var},     gibbs) nomodelsummary

Burn-in ...
Simulation ...
Bayesian normal regression
Gibbs sampling

<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.216249</td>
<td>.0816994</td>
<td>.000817</td>
<td>6.216445</td>
<td>6.053813-6.377687</td>
</tr>
<tr>
<td>_cons</td>
<td>19.31436</td>
<td>.4619975</td>
<td>.004539</td>
<td>19.31138</td>
<td>18.41486-20.22794</td>
</tr>
</tbody>
</table>

Yulia Marchenko (StataCorp)  September 4, 2015
Measurements within a pig are correlated—introduce a random intercept

Likelihood model:

\[
\text{weight}_{ij} = \beta_0 + u_{0j} + \beta_1 \text{week}_{ij} + \epsilon_{ij}
\]

\[
\epsilon_{ij} \sim \text{Normal}(0,\sigma^2)
\]

\[
u_{0j} \sim \text{Normal}(0,\sigma_0^2)
\]

where \(i = 1, \ldots, 9\) and \(j = 1, \ldots, 48\).

Prior distributions:

\[
\beta_i \sim \text{Normal}(0,100), \ i = 0, 1
\]

\[
\sigma^2 \sim \text{InvGamma}(0.001, 0.001)
\]

\[
\sigma_0^2 \sim \text{InvGamma}(0.001, 0.001)
\]
Alternative model formulation

- Let $\tau_{0j} = \beta_0 + u_{0j}$
- Alternative likelihood model formulation:

$$
\text{weight}_{ij} = \tau_{0j} + \beta_1 \text{week}_{ij} + \epsilon_{ij}
$$

$$
\epsilon_{ij} \sim \text{Normal}(0,\sigma^2)
$$

$$
\tau_{0j} \sim \text{Normal}(\beta_0,\sigma_0^2)
$$
Default MH sampling

```
. webuse pig  
(Longitudinal analysis of pig weights)
. fvset base none id
. set seed 14
. bayesmh weight week i.id, likelihood(normal({var})) noconstant ///
>    prior({weight:i.id}, normal({weight:cons},{var_0})) ///
>    prior({weight:week}, normal(0,100)) ///
>    prior({weight:cons}, normal(0,100)) ///
>    prior({var},     igamma(0.001,0.001)) ///
>    prior({var_0},   igamma(0.001,0.001)) ///
>    noshow({weight:i.id})
```
Model summary

- Burn-in ...
- Simulation ...
- Model summary

Likelihood:
  weight ~ normal(xb_weight,{var})

Priors:
  {weight:week} ~ normal(0,100) \hspace{1em} (1)
  {weight:i.id} ~ normal({weight:cons},{var_0}) \hspace{1em} (1)
  \{var\} ~ igamma(0.001,0.001)

Hyperpriors:
  \{weight:cons\} ~ normal(0,100)
  \{var_0\} ~ igamma(0.001,0.001)

(1) Parameters are elements of the linear form xb_weight.
Bayesian analysis using Stata
More examples (extra)
Random-intercept model

Bayesian normal regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 432
Acceptance rate = .2341
Efficiency: min = .001963
avg = .005539
max = .01159

Log marginal likelihood = -1338.2346

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.257469</td>
<td>.0273198</td>
<td>.002538</td>
<td>6.256309</td>
<td>6.205179 - 6.309333</td>
</tr>
<tr>
<td></td>
<td>var</td>
<td>8.895206</td>
<td>.6146577</td>
<td>8.844657</td>
<td>7.799991 - 10.25156</td>
</tr>
<tr>
<td></td>
<td>var_0</td>
<td>12.36591</td>
<td>.35361</td>
<td>12.36093</td>
<td>11.66033 - 13.05275</td>
</tr>
</tbody>
</table>

Note: There is a high autocorrelation after 500 lags.
Default MH sampling is very inefficient in this example

Improve MCMC efficiency by blocking of parameters

Use `block()`’s suboption `split` to block random-effects parameters—very important with many random effects

```stata
.set seed 14
.bayesmh weight week i.id, likelihood(normal({var})) noconstant ///
>    prior({weight:i.id}, normal({weight:cons},{var_0})) ///
>    prior({weight:week}, normal(0,100)) ///
>    prior({weight:cons}, normal(0,100)) ///
>    prior({var}, igamma(0.001,0.001)) ///
>    prior({var_0}, igamma(0.001,0.001)) ///
>    block({var}) block({var_0}) ///
>    block({weight:week}) block({weight:cons}) ///
>    block({weight:i.id}, split) ///
>    nomodelsummary notable
```
Blocking improved MCMC efficiency

Burn-in ...
Simulation ...
Bayesian normal regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 432
Acceptance rate = .4447
Efficiency: min = .02386
avg = .1491
max = .1953

Log marginal likelihood = -1052.2375
Estimates are more similar to the frequentist results (see \texttt{[ME] mixed})

```
.bayesstats summary \{weight:week cons\} \{var_0\} \{var\}
```

<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>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.203559</td>
<td>.0382251</td>
<td>.002475</td>
<td>6.20247</td>
<td>6.132607 - 6.279994</td>
</tr>
<tr>
<td>cons</td>
<td>19.353</td>
<td>.6176088</td>
<td>.019352</td>
<td>19.3461</td>
<td>18.15131 - 20.57819</td>
</tr>
<tr>
<td>var</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var_0</td>
<td>15.88671</td>
<td>3.595539</td>
<td>.094179</td>
<td>15.32318</td>
<td>10.62316 - 24.33477</td>
</tr>
<tr>
<td>var</td>
<td>4.427113</td>
<td>.3264523</td>
<td>.007969</td>
<td>4.404244</td>
<td>3.835123 - 5.102618</td>
</tr>
</tbody>
</table>

Posterior summary statistics

MCMC sample size = 10,000
- Including random effects as a factor variable is not feasible with tens of thousands of random effects.
- Option `split` is very time consuming.
- Forthcoming option `reffects()` is an alternative.
- Replace `i.id` in the model formulation with option `reffects(id)` and remove block(`weight:i.id, split`)

```stata
. set seed 14
. bayesmh weight week, likelihood(normal({var})) noconstant reffects(id) ///
    > prior({weight:i.id}, normal({weight:cons},{var_0})) ///
    > prior({weight:week}, normal(0,100)) ///
    > prior({weight:cons}, normal(0,100)) ///
    > prior({var}, igamma(0.001,0.001)) ///
    > prior({var_0}, igamma(0.001,0.001)) ///
    > block({var}) block({var_0}) ///
    > block({weight:week}) block({weight:cons}) ///
    > nomodelsummary notable
```
MCMC sampling efficiencies are slightly smaller

Bayesian normal regression
Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 432
Acceptance rate = .3788
Efficiency: min = .01923
avg = .0944
max = .1566

Log marginal likelihood = -1077.2283
Estimates are similar to previous estimates

\[
. \text{bayesstats summary \{weight:week cons\} \{var_0\} \{var\}}
\]
Posterior summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.215106</td>
<td>.0378704</td>
<td>.002731</td>
<td>6.214882</td>
<td>6.139693</td>
</tr>
<tr>
<td>cons</td>
<td>19.25063</td>
<td>.6306763</td>
<td>.02307</td>
<td>19.24458</td>
<td>18.00894</td>
</tr>
<tr>
<td>var_0</td>
<td>16.00539</td>
<td>3.739944</td>
<td>.104932</td>
<td>15.44782</td>
<td>10.336</td>
</tr>
<tr>
<td>var</td>
<td>4.432357</td>
<td>.3225202</td>
<td>.00815</td>
<td>4.416106</td>
<td>3.836758</td>
</tr>
</tbody>
</table>
We can use Gibbs sampling for some of the parameters to further improve MCMC sampling.

Average MCMC sampling efficiency increased from 9% to 30%.

```stata
set seed 14
.bayesmh weight week, likelihood(normal({var})) noconstant reffects(id) ///
   > prior({weight:i.id}, normal({weight:cons},{var_0})) ///
   > prior({weight:week}, normal(0,100)) ///
   > prior({weight:cons}, normal(0,100)) ///
   > prior({var}, igamma(0.001,0.001)) ///
   > prior({var_0}, igamma(0.001,0.001)) ///
   > block({var}, gibbs) block({var_0}, gibbs) ///
   > block({weight:week}, gibbs) block({weight:cons}, gibbs) ///
   > nomodelsummary notable

Burn-in ...
Simulation ...
Bayesian normal regression
Metropolis-Hastings and Gibbs sampling
MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of obs = 432
Acceptance rate = .8235
Efficiency: min = .02439
case avg = .2851
max = .6009

Log marginal likelihood = -1077.0036
```
. bayesstats summary {weight:week cons} {var_0} {var}

Posterior summary statistics

MCMC sample size = 10,000

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Equal-tailed</td>
</tr>
<tr>
<td>week</td>
<td>6.216461</td>
<td>0.0383844</td>
<td>0.002458</td>
<td>6.217039</td>
<td>6.139121 6.291271</td>
</tr>
<tr>
<td>cons</td>
<td>19.24988</td>
<td>0.6046734</td>
<td>0.015102</td>
<td>19.24786</td>
<td>18.06586 20.46588</td>
</tr>
<tr>
<td>var_0</td>
<td>15.78329</td>
<td>3.541348</td>
<td>0.045683</td>
<td>15.32768</td>
<td>10.28163 24.15133</td>
</tr>
<tr>
<td>var</td>
<td>4.423026</td>
<td>0.3241646</td>
<td>0.005444</td>
<td>4.409645</td>
<td>3.824604 5.100363</td>
</tr>
</tbody>
</table>
Pig-specific slopes—random coefficient on week

Likelihood model:

\[ \text{weight}_{ij} = \beta_0 + u_{0j} + (\beta_1 + u_{1j})\text{week}_{ij} + \epsilon_{ij} \]

\[ \epsilon_{ij} \sim \text{Normal}(0, \sigma^2) \]

\[ u_{0j} \sim \text{Normal}(0, \sigma_0^2) \]

\[ u_{1j} \sim \text{Normal}(0, \sigma_1^2) \]

where \( i = 1, \ldots, 9 \) and \( j = 1, \ldots, 48 \).

Prior distributions:

\[ \beta_i \sim \text{Normal}(0, 100), \ i = 0, 1 \]

\[ \sigma^2 \sim \text{InvGamma}(0.001, 0.001) \]

\[ \sigma_0^2 \sim \text{InvGamma}(0.001, 0.001) \]

\[ \sigma_1^2 \sim \text{InvGamma}(0.001, 0.001) \]
Alternative model formulation

- Let $\tau_{0j} = \beta_0 + u_{0j}$ and $\tau_{1j} = \beta_1 + u_{1j}$
- Alternative likelihood model formulation:
  \[
  \text{weight}_{ij} = \tau_{0j} + \tau_{ij\text{ week}_{ij}} + \epsilon_{ij}
  \]
  \[
  \epsilon_{ij} \sim \text{Normal}(0, \sigma^2)
  \]
  \[
  \tau_{0j} \sim \text{Normal}(\beta_0, \sigma_{0}^2)
  \]
  \[
  \tau_{1j} \sim \text{Normal}(\beta_1, \sigma_{1}^2)
  \]
Option `reffects()` supports only (two-level) random-intercept models.

Must use the factor-variable specification.

But can replace time-consuming splitting with the forthcoming suboption `reffects in a block()`.
. webuse pig
(Longitudinal analysis of pig weights)
. fvset base none id
. set seed 14
. bayesmh weight i.id i.id#c.week, likelihood(normal({var})) noconstant ///
>    prior({weight:i.id}, normal({weight:cons},{var_0})) ///
>    prior({weight:i.id#c.week}, normal({weight:week},{var_1})) ///
>    prior({weight:week}, normal(0,100)) ///
>    prior({weight:cons}, normal(0,100)) ///
>    prior({var}, igamma(0.001,0.001)) ///
>    prior({var_0}, igamma(0.001,0.001)) ///
>    prior({var_1}, igamma(0.001,0.001)) ///
>    block({weight:i.id}, reffects) ///
>    block({weight:i.id#c.week}, reffects) ///
>    block({var}, gibbs) block({var_0}, gibbs) block({var_1}, gibbs) ///
>    block({weight:week}, gibbs) block({weight:cons}, gibbs) ///
>    burnin(10000) noshow({weight:i.id i.id#c.week}) dots
Model summary

Burn-in 10000 aaaaaaaaaa1000aa........2000...........3000...........4000...........5000 > ........6000...........7000...........8000...........9000...........10000 done
Simulation 10000 ...........1000...........2000...........3000...........4000...........5 > 000...........6000...........7000...........8000...........9000...........10000 done

Model summary

Likelihood:
  weight ~ normal(xb_weight,{var})

Priors:
  \{weight:i.id\} ~ normal(\{weight:cons\},{var_0})         \(1\)
  \{weight:i.id#c.week\} ~ normal(\{weight:week\},{var_1})\(1\)
  \{var\} ~ igamma(0.001,0.001)

Hyperpriors:
  \{weight:week cons\} ~ normal(0,100)
  \{var_0 var_1\} ~ igamma(0.001,0.001)

(1) Parameters are elements of the linear form xb_weight.
. bayesstats summary {weight:week cons} {var_0} {var_1} {var}
Posterior summary statistics

<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>weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.206141</td>
<td>.0934325</td>
<td>.002277</td>
<td>6.206412</td>
<td>6.02124</td>
</tr>
<tr>
<td>cons</td>
<td>19.33658</td>
<td>.4127152</td>
<td>.013154</td>
<td>19.33267</td>
<td>18.52088</td>
</tr>
<tr>
<td>var_0</td>
<td>7.192013</td>
<td>1.73689</td>
<td>.080111</td>
<td>6.972026</td>
<td>4.541918</td>
</tr>
<tr>
<td>var_1</td>
<td>.391377</td>
<td>.0897799</td>
<td>.00281</td>
<td>.3801791</td>
<td>.2507229</td>
</tr>
<tr>
<td>var</td>
<td>1.616059</td>
<td>.1252948</td>
<td>.004119</td>
<td>1.608114</td>
<td>1.389298</td>
</tr>
</tbody>
</table>
Relax the assumption of independence between random intercepts and random coefficients

Likelihood model:

\[
\text{weight}_{ij} = \tau_{0j} + \tau_{ij}\text{week}_{ij} + \epsilon_{ij}
\]

\[
\epsilon_{ij} \sim \text{Normal}(0, \sigma^2)
\]

\[
\begin{pmatrix}
\tau_{0j} \\
\tau_{1j}
\end{pmatrix} \sim \text{MVN}\left\{\begin{pmatrix}
\beta_0 \\
\beta_1
\end{pmatrix}, \Sigma = \begin{pmatrix}
\sigma_0^2 & \sigma_{01} \\
\sigma_{01} & \sigma_1^2
\end{pmatrix}\right\}
\]

where \( i = 1, \ldots, 9 \) and \( j = 1, \ldots, 48 \).

Prior distributions:

\[
\beta_i \sim \text{Normal}(0, 100), \ i = 0, 1
\]

\[
\sigma^2 \sim \text{InvGamma}(0.001, 0.001)
\]

\[
\Sigma \sim \text{InvWishart}(3, I(2))
\]
Forthcoming specification of the `mvnormal()` prior for specifying an unstructured covariance for multiple sets of random effects

```stata
.set seed 14
.bayesmh weight i.id i.id#c.week, likelihood(normal({var})) noconstant ///
  >  prior({weight:i.id i.id#c.week},
    >  mvnnormal(2,{weight:cons},{weight:week},{Sigma, matrix}))
  >  prior({weight:week cons}, normal(0,100))
  >  prior({var}, igamma(0.001,0.001))
  >  prior({Sigma,m}, iwishart(2,3,I(2)))
  >  block({weight:i.id}, reffects)
  >  block({weight:i.id#c.week}, reffects)
  >  block({var}, gibbs)
  >  block({Sigma,m}, gibbs)
  >  burnin(10000) nomodelsummary notable dots
```

**Burn-in ...**

**Simulation ...**

Bayesian normal regression

Metropolis-Hastings and Gibbs sampling

```
MCMC iterations   =   20,000
Burn-in           =    10,000
MCMC sample size  =    10,000
Number of obs     =     432
Acceptance rate   =   .5005
Efficiency: min   =   .005916
                  avg =   .01594
Log marginal likelihood =  -924.64857
```
. bayesstats summary {weight:week cons} {Sigma} {var}

Posterior summary statistics

<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><strong>weight</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>6.212649</td>
<td>.0965009</td>
<td>.003403</td>
<td>6.214282</td>
<td>6.016377 6.390494</td>
</tr>
<tr>
<td><strong>Sigma</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigma_1_1</td>
<td>6.938195</td>
<td>1.637985</td>
<td>.076558</td>
<td>6.735893</td>
<td>4.42667 10.71507</td>
</tr>
<tr>
<td>Sigma_2_1</td>
<td>-.0926991</td>
<td>.2678663</td>
<td>.009932</td>
<td>-.0843172</td>
<td>-.656516 .4238284</td>
</tr>
<tr>
<td>Sigma_2_2</td>
<td>.3997822</td>
<td>.0893766</td>
<td>.002398</td>
<td>.3879609</td>
<td>.2610762 .6069753</td>
</tr>
<tr>
<td><strong>var</strong></td>
<td>1.612773</td>
<td>.1277831</td>
<td>.004689</td>
<td>1.607633</td>
<td>1.385116 1.881754</td>
</tr>
</tbody>
</table>
Meta analysis is a statistical analysis that involves summarizing results from similar but independent studies.

Consider data from nine clinical trials that examined the effect of taking diuretics during pregnancy on the risk of preeclampsia (Tanner et al. 2000).

Data contain estimates of treatment effects expressed as log odds-ratios ($\ln \text{OR}$) and their respective variances ($\text{var}$).

Negative $\ln \text{OR}$ values indicate that taking diuretics lowers the risk of preeclampsia.
• Likelihood model:

\[ y_i \sim \text{Normal}(\mu_i, \text{var}_i) \]
\[ \mu_i \sim \text{Normal}(\theta, \tau^2) \]

where \( i = 1, \ldots, 9 \).

• Prior distributions:

\[ \theta \sim \text{Normal}(0, 10000) \]
\[ \tau^2 \sim \text{InvGamma}(0.0001, 0.0001) \]
. use meta
(Meta analysis of clinical trials studying diuretics and pre-eclampsia)
. set seed 14
. fvset base none trial
. bayesmh lnOR i.trial, noconstant likelihood(normal(var)) ///
   > prior({lnOR:i.trial}, normal({theta},{tau2})) ///
   > prior({theta}, normal(0,10000)) ///
   > prior({tau2}, igamma(0.0001,0.0001)) ///
   > block({lnOR:i.trial}, split) ///
   > block({theta}, gibbs) ///
   > block({tau2}, gibbs)

Burn-in ...  
Simulation ...  
Model summary

<table>
<thead>
<tr>
<th>Likelihood:</th>
<th>lnOR ~ normal(xb_lnOR,var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior:</td>
<td>{lnOR:i.trial} ~ normal({theta},{tau2}) (1)</td>
</tr>
<tr>
<td>Hyperpriors:</td>
<td>{theta} ~ normal(0,10000)</td>
</tr>
<tr>
<td></td>
<td>{tau2} ~ igamma(0.0001,0.0001)</td>
</tr>
</tbody>
</table>

(1) Parameters are elements of the linear form xb_lnOR.
Bayesian normal regression
Metropolis-Hastings and Gibbs sampling

- MCMC iterations = 12,500
- Burn-in = 2,500
- MCMC sample size = 10,000
- Number of obs = 9
- Acceptance rate = .6353
- Efficiency: min = .01537
  avg = .0647
  max = .1798

Log marginal likelihood = 8.2435069

<table>
<thead>
<tr>
<th>trial</th>
<th>lnOR</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.2074594</td>
<td>0.3233577</td>
<td>0.014264</td>
<td>-0.2390982</td>
<td>-0.7840912</td>
<td>.4732284</td>
</tr>
<tr>
<td>2</td>
<td>-0.7422326</td>
<td>0.3059792</td>
<td>0.014353</td>
<td>-0.7277104</td>
<td>-1.352696</td>
<td>-0.2290158</td>
</tr>
<tr>
<td>3</td>
<td>-0.8101728</td>
<td>0.3579343</td>
<td>0.019156</td>
<td>-0.7938089</td>
<td>-1.557279</td>
<td>-0.2024199</td>
</tr>
<tr>
<td>4</td>
<td>-0.8860118</td>
<td>0.4367827</td>
<td>0.027156</td>
<td>-0.8529495</td>
<td>-1.824588</td>
<td>-0.1811792</td>
</tr>
<tr>
<td>5</td>
<td>-1.032694</td>
<td>0.3685822</td>
<td>0.029732</td>
<td>-1.046375</td>
<td>-1.738105</td>
<td>-0.3787439</td>
</tr>
<tr>
<td>6</td>
<td>-0.3225829</td>
<td>0.0969534</td>
<td>0.003571</td>
<td>-0.3241207</td>
<td>-0.5102041</td>
<td>-0.1320317</td>
</tr>
<tr>
<td>7</td>
<td>-0.3476522</td>
<td>0.2873013</td>
<td>0.008138</td>
<td>-0.3712284</td>
<td>-0.8994376</td>
<td>0.2624625</td>
</tr>
<tr>
<td>8</td>
<td>-0.0831874</td>
<td>0.5189861</td>
<td>0.019312</td>
<td>-0.1686125</td>
<td>-0.9203838</td>
<td>1.128532</td>
</tr>
<tr>
<td>9</td>
<td>-0.0531772</td>
<td>0.268729</td>
<td>0.016447</td>
<td>-0.0631959</td>
<td>-0.5078684</td>
<td>0.5056795</td>
</tr>
</tbody>
</table>

| theta | -0.499449 | 0.2307223 | 0.005441 | -0.4849543 | -0.9790357 | -0.0413009 |
| tau2  | 0.3385446 | 0.4122769 | 0.016601 | 0.2325792 | 0.0003896 | 1.332994 |

Note: Adaptation tolerance is not met in at least one of the blocks.
Test whether taking diuretics reduces the risk of preeclampsia

```stata
. bayestest interval {theta}, upper(0)
Interval tests      MCMC sample size = 10,000
prob1 : {theta} < 0

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob1</td>
<td>.9825</td>
<td>0.13113</td>
<td>.0017971</td>
</tr>
</tbody>
</table>
```
 Plot posterior distributions of trial-specific effects

. bayesgraph histogram {lnOR:i.trial}, ///
  byparm(legend(off) noxrescale noyrescale) ///
  title(Posterior distributions of trial effects) ///
  normal addplot(pci 0 -0.51 4 -0.51, lcolor(red))
### Posterior distributions of trial effects

<table>
<thead>
<tr>
<th>InOR:1bn.trial</th>
<th>InOR:2.trial</th>
<th>InOR:3.trial</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
<tr>
<td>InOR:4.trial</td>
<td>InOR:5.trial</td>
<td>InOR:6.trial</td>
</tr>
<tr>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>InOR:7.trial</td>
<td>InOR:8.trial</td>
<td>InOR:9.trial</td>
</tr>
<tr>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
<td><img src="image9.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

Graphs by parameter
British coal mining disaster dataset from 1851 to 1962 (Carlin, Gelfand, and Smith 1992)

- Outcome count: number of disasters involving 10 or more deaths
- There was a fairly abrupt decrease in the rate of disasters around 1887–1895.
- Estimate the date, change point $cp$, when the rate of disasters changed.
Bayesian analysis using Stata

More examples (extra)

Nonlinear Poisson model: Change-point analysis

- Likelihood model:

\[
\begin{align*}
\text{counts}_i &\sim \text{Poisson}(\mu_1), \text{ if } \text{year}_i < cp \\
\text{counts}_i &\sim \text{Poisson}(\mu_2), \text{ if } \text{year}_i \geq cp
\end{align*}
\]

where \( i = 1, \ldots, 112. \)

- Prior distributions:

\[
\begin{align*}
\mu_1 &\sim 1 \\
\mu_2 &\sim 1 \\
cp &\sim \text{Uniform}(1851, 1962)
\end{align*}
\]
. webuse coal  
(British coal-mining disaster data, 1851-1962)  
. set seed 14  
. bayesmh count = ({mu1}*sign(year<{cp})+{mu2}*sign(year>={cp})), ///  
>  likelihood(poisson, noglmtransform) ///  
>  prior({mu1} {mu2}, flat) ///  
>  prior({cp}, uniform(1851,1962)) ///  
>  initial({mu1} 1 {mu2} 1 {cp} 1906) ///  
>  title(Change-point analysis)

Burn-in ...
Simulation ...
Model summary

Likelihood:
  count ~ poisson({mu1}*sign(year<{cp})+{mu2}*sign(year>={cp}))

Priors:
  {mu1 mu2} ~ 1 (flat)
  {cp} ~ uniform(1851,1962)
Estimate the ratio between the two means

After 1890, the mean number of disasters decreased by a factor of about 3.4 with a 95% credible range of [2.47, 4.55].

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>95% Cred. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu1</td>
<td>3.118753</td>
<td>.3001234</td>
<td>.015504</td>
<td>3.110907</td>
<td>2.545246</td>
</tr>
<tr>
<td>cp</td>
<td>1890.362</td>
<td>2.4808</td>
<td>.071835</td>
<td>1890.553</td>
<td>1886.065</td>
</tr>
<tr>
<td>mu2</td>
<td>.9550596</td>
<td>.1209208</td>
<td>.005628</td>
<td>.9560248</td>
<td>.7311639</td>
</tr>
</tbody>
</table>
. bayesstats summary (ratio:{mu1}/{mu2})

Posterior summary statistics

<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>ratio</td>
<td>3.316399</td>
<td>.5179103</td>
<td>.027848</td>
<td>3.270496</td>
<td>2.404047  -  4.414975</td>
</tr>
</tbody>
</table>

MCMC sample size = 10,000
Crossover design is a repeated-measures design in which patients crossover from one treatment to another during the course of the study.

Crossover designs are widely used for testing the efficacy of new drugs.

Consider a two-treatment, two-period crossover trial comparing two Carbamazepine tablets: A—new and B—standard (Gelfand et al. 1990).

10 subjects were randomized to two treatment sequences: AB and BA.

Outcome: logarithms of maxima of concentration-time curves.
Bayesian analysis using Stata

More examples (extra)

Bioequivalence in a crossover trial

- Likelihood model:

\[
Y_{i(jk)} = \mu + (-1)^{j-1} \frac{\phi}{2} + (-1)^{k-1} \frac{\pi}{2} + d_i + \epsilon_{i(jk)} = \mu_{i(jk)} + \epsilon_{i(jk)}
\]

\[
\epsilon_{i(jk)} \sim \text{Normal}(0, \sigma^2)
\]

\[
d_i \sim \text{Normal}(0, \tau^2)
\]

where \( i = 1, \ldots, 10, \, j = 1, 2 \) is the treatment group (sequence), and \( k = 1, 2 \) is the period.

- Prior distributions:

\[
\mu, \, \phi, \, \pi \sim \text{Normal}(0, 10^6)
\]

\[
\sigma^2 \sim \text{InvGamma}(0.001, 0.001)
\]

\[
\tau^2 \sim \text{InvGamma}(0.001, 0.001)
\]
. webuse bioequiv
   (Bioequivalent study of Carbamazepine tablets)
. set seed 14
. fvset base none id
. bayesmh y = ({mu}+(-1)^(treat-1)*{phi}/2+(-1)^(period-1)*{pi}/2+{y:i.id}), ///
   likelihood(normal({var})) ///
   prior({y:i.id}, normal(0,{tau2})) ///
   prior({tau2}, igamma(0.001,0.001)) ///
   prior({var}, igamma(0.001,0.001)) ///
   prior({mu} {phi} {pi}, normal(0,1e6)) ///
   block({y:i.id}, reffects) ///
   block({tau2}, gibbs) ///
   block({var}, gibbs) ///
   adaptation(every(200) maxiter(50)) burnin(10000) ///
   noshow({y:i.id})
Model summary

Likelihood:
   \( y \sim \text{normal}(\text{expr1}, \text{var}) \)

Priors:
   \text{var} \sim \text{igamma}(0.001, 0.001)
   \{y:i.id\} \sim \text{normal}(0, \text{tau2})
   \{\mu, \phi, \pi\} \sim \text{normal}(0, 1e6)

Hyperprior:
   \text{tau2} \sim \text{igamma}(0.001, 0.001)

Expression:
   \text{expr1} : \{\mu\}+(-1)^{(\text{treat}-1)\times\phi}/2+(-1)^{(\text{period}-1)\times\pi}/2+(\{y:1bn.id\}+1bn.id+\{y:2.id\}+\{y:3.id\}+\{y:4.id\}+\{y:5.id\}+\{y:6.id\}+\{y:7.id\}+\{y:8.id\}+\{y:9.id\}+\{y:10.id\})
Bayesian normal regression
Metropolis-Hastings and Gibbs sampling

MCMC iterations = 20,000
Burn-in = 10,000
MCMC sample size = 10,000
Number of obs = 20
Acceptance rate = .5959
Efficiency: min = .01359
        avg = .03528
        max = .0511

Log marginal likelihood = -8.6538165

<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>mu</td>
<td>1.444575</td>
<td>.0492361</td>
<td>.004224</td>
<td>1.444955</td>
<td>1.350906 - 1.54269</td>
</tr>
<tr>
<td>phi</td>
<td>-.0092691</td>
<td>.0537334</td>
<td>.00255</td>
<td>-.0087842</td>
<td>-.1126505 - 0.0939082</td>
</tr>
<tr>
<td>pi</td>
<td>-.1768478</td>
<td>.0517259</td>
<td>.002288</td>
<td>-.1785769</td>
<td>-.2839622 - -.0668874</td>
</tr>
<tr>
<td>var</td>
<td>.0136361</td>
<td>.0090926</td>
<td>.000637</td>
<td>.0109485</td>
<td>.004295 - .0377165</td>
</tr>
<tr>
<td>tau2</td>
<td>.02173</td>
<td>.0175663</td>
<td>.000811</td>
<td>.017856</td>
<td>.0023005 - .0647257</td>
</tr>
</tbody>
</table>
\[ \theta = \exp(\phi) \] is commonly used as a measure of bioequivalence.

Bioequivalence is declared whenever \( \theta \) lies in the interval \((0.8, 1.2)\) with a high posterior probability.

```
. bayesstats summary (equiv:exp({phi})>0.8 & exp({phi})<1.2)
Posterior summary statistics               MCMC sample size =  10,000
equiv : exp({phi})>0.8 & exp({phi})<1.2

<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>equiv</td>
<td>.9937</td>
<td>.0791261</td>
<td>.003951</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
```