

# Introduction to Bayesian Analysis Using Stata

Chuck Huber, PhD  
StataCorp  
[chuber@stata.com](mailto:chuber@stata.com)

# Download Website

- You can download all of the slides, datasets and do-files here:

**<https://tinyurl.com/IntroToBayes>**

# Outline

- Introduction to Bayesian Analysis
- Coin Toss Example
- Priors, Likelihoods, and Posteriors
- Markov Chain Monte Carlo (MCMC)
- Bayesian Linear Regression
- Advantages and Disadvantages of Bayes

# STATA BAYESIAN ANALYSIS REFERENCE MANUAL RELEASE 14

## Title

**bayesmh** — Bayesian regression using Metropolis–Hastings algorithm

## Description

`bayesmh` fits a variety of Bayesian models using an adaptive Metropolis–Hastings (MH) algorithm. It provides various likelihood models and prior distributions for you to choose from. Likelihood models include univariate normal linear and nonlinear regressions, multivariate normal linear and nonlinear regressions, generalized linear models such as logit and Poisson regressions, and multiple-equations linear models. Prior distributions include continuous distributions such as uniform, Jeffreys, normal, gamma, multivariate normal, and Wishart and discrete distributions such as Bernoulli and Poisson. You can also program your own Bayesian models; see [[BAYES](#)] **bayesmh evaluators**.

bayesmh - Bayesian regression using Metropolis-Hastings algorithm

Model Model 2 if/in Weights Simulation Adaptation Reporting Advanced

Syntax:

Univariate distributions

- Univariate linear models
- Multivariate normal linear regression with common regressors
- Multivariate normal linear regression with outcome-specific regressors
- Multiple-equation linear models
- Univariate nonlinear regression
- Multivariate normal nonlinear regression
- Univariate distributions
- Multiple-equation distribution specifications

Distribution

--> Exponential distribution  
--> Bernoulli distribution  
--> Binomial distribution  
--> Poisson distribution

Success probability:  
 Create...

Priors of model parameters

Press "Create" to define a prior distribution

Show model summary without estimation

bayesmh - Bayesian regression using Metropolis-Hastings algorithm

Model Model 2 if/in Weights Simulation Adaptation Reporting Advanced

Syntax:

Univariate distributions

Model

Dependent variable:  
heads

Distribution

--> Exponential distribution  
--> Bernoulli distribution  
--> Binomial distribution  
--> Poisson distribution

Success probability:  
{theta} Create...

Priors of model parameters

Prior 1

Create... Edit Disable Enable

prior({theta}, beta(1,1))

Show model summary without estimation

?

OK Cancel Submit

Prior 1

Parameters specification:  
{theta}

Choose a prior distribution:

Univariate continuous

- > Normal distribution
- > Lognormal distribution
- > Uniform distribution
- > Gamma distribution
- > Inverse gamma distribution
- > Exponential distribution
- > Beta distribution
- > Chi-squared distribution
- > Jeffreys prior for variance of normal distribution

Multivariate continuous

- > Multivariate normal distribution
- > Multivariate normal distribution with zero mean
- > Zellner's g-prior
- > Zellner's g-prior with zero mean
- > Wishart distribution
- > Inverse Wishart distribution
- > Jeffreys prior for covariance of multivariate normal

Discrete

- > Bernoulli distribution
- > Discrete index distribution
- > Poisson distribution

Generic

- > Flat prior (with a density of 1)
- > Generic density
- > Generic log density

Shape a:  
1 Create...

Shape b:  
1 Create...

?

OK Cancel

# The **bayesmh** Command

```
bayesmh sbp age sex bmi,                                ///
likelihood(normal({sigma2}))                         ///
prior({sbp:_cons}, normal(0,100))                   ///
prior({sbp:age}, normal(0,100))                      ///
prior({sbp:sex}, normal(0,100))                      ///
prior({sbp:bmi}, normal(0,100))                      ///
prior({sigma2}, igamma(1,1))
```

# STATA BAYESIAN ANALYSIS REFERENCE MANUAL RELEASE 15

## Title

**bayes** — Bayesian regression models using the bayes prefix

## Description

The `bayes` prefix fits Bayesian regression models. It provides Bayesian support for many likelihood-based estimation commands. The `bayes` prefix uses default or user-supplied priors for model parameters and estimates parameters using MCMC by drawing simulation samples from the corresponding posterior model. Also see [BAYES] `bayesmh` and [BAYES] `bayesmh evaluators` for fitting more general Bayesian models.

# The **bayes** Prefix

**regress sbp age sex**

**bayes: regress sbp age sex**

**logistic highbp age sex**

**bayes: logistic highbp age sex**

# Two Paradigms

## Frequentist Statistics

Model parameters are considered to be unknown but fixed constants and the observed data are viewed as a repeatable random sample.

## Bayesian Statistics

Model parameters are random quantities which have a posterior distribution formed by combining prior knowledge about parameters with the evidence from the observed data sample.

# Reverend Thomas Bayes



- 1701 – born in London
- Presbyterian Minister
- Amateur Mathematician
- Published one paper on theology and one on mathematics
- 1761 – died in Kent
- 1763 - “Bayes Theorem” paper published by friend Richard Price

# Coin Toss Example



What is the probability of heads ( $\theta$ )?

# Prior Distribution

Prior distributions are probability distributions of model parameters based on some a priori knowledge about the parameters.

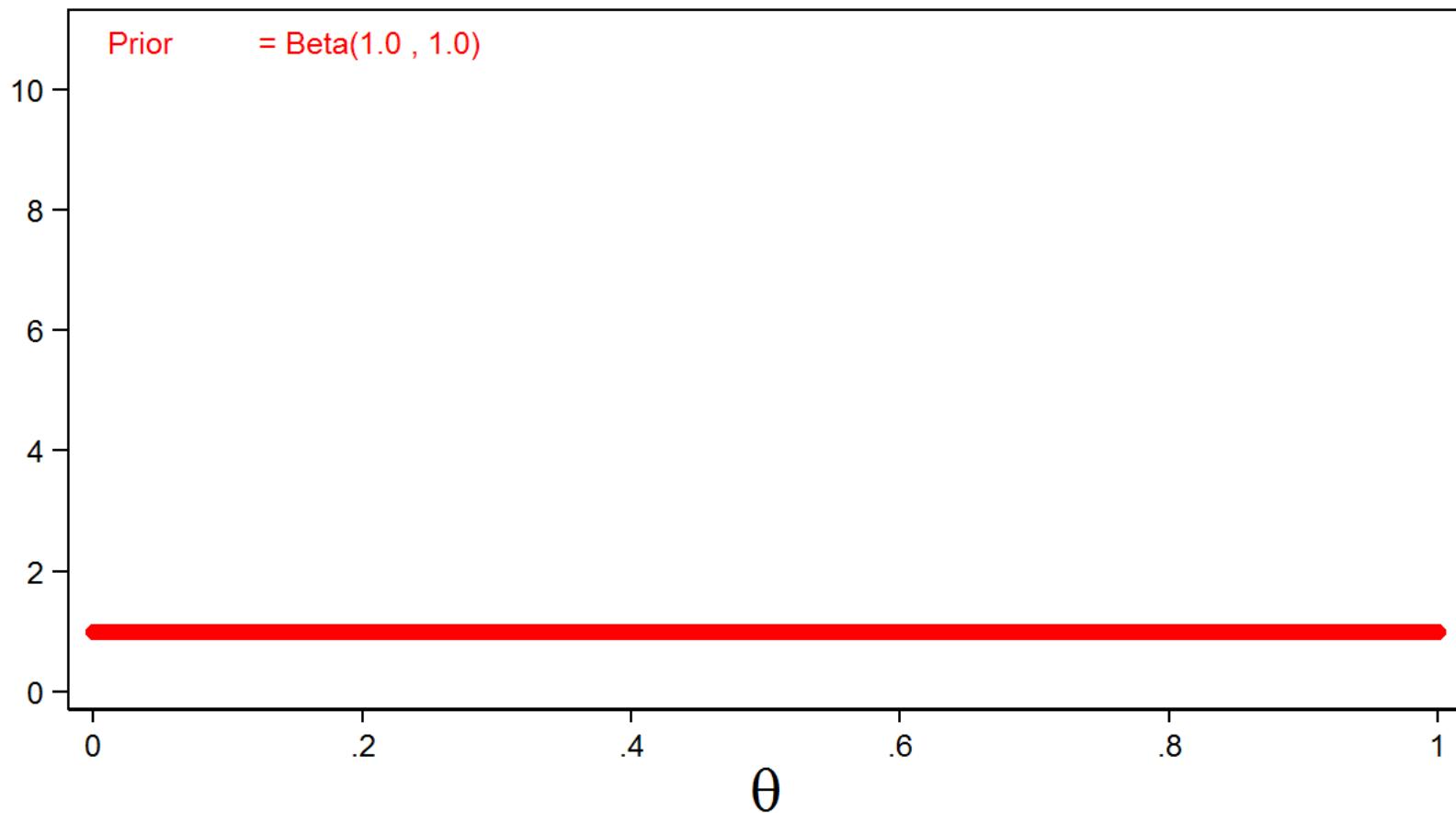
Prior distributions are independent of the observed data.

# Beta Prior for $\theta$

$$P(\theta) = \text{Beta}(\alpha, \beta)$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{(\alpha-1)}(1 - \theta)^{(\beta-1)}$$

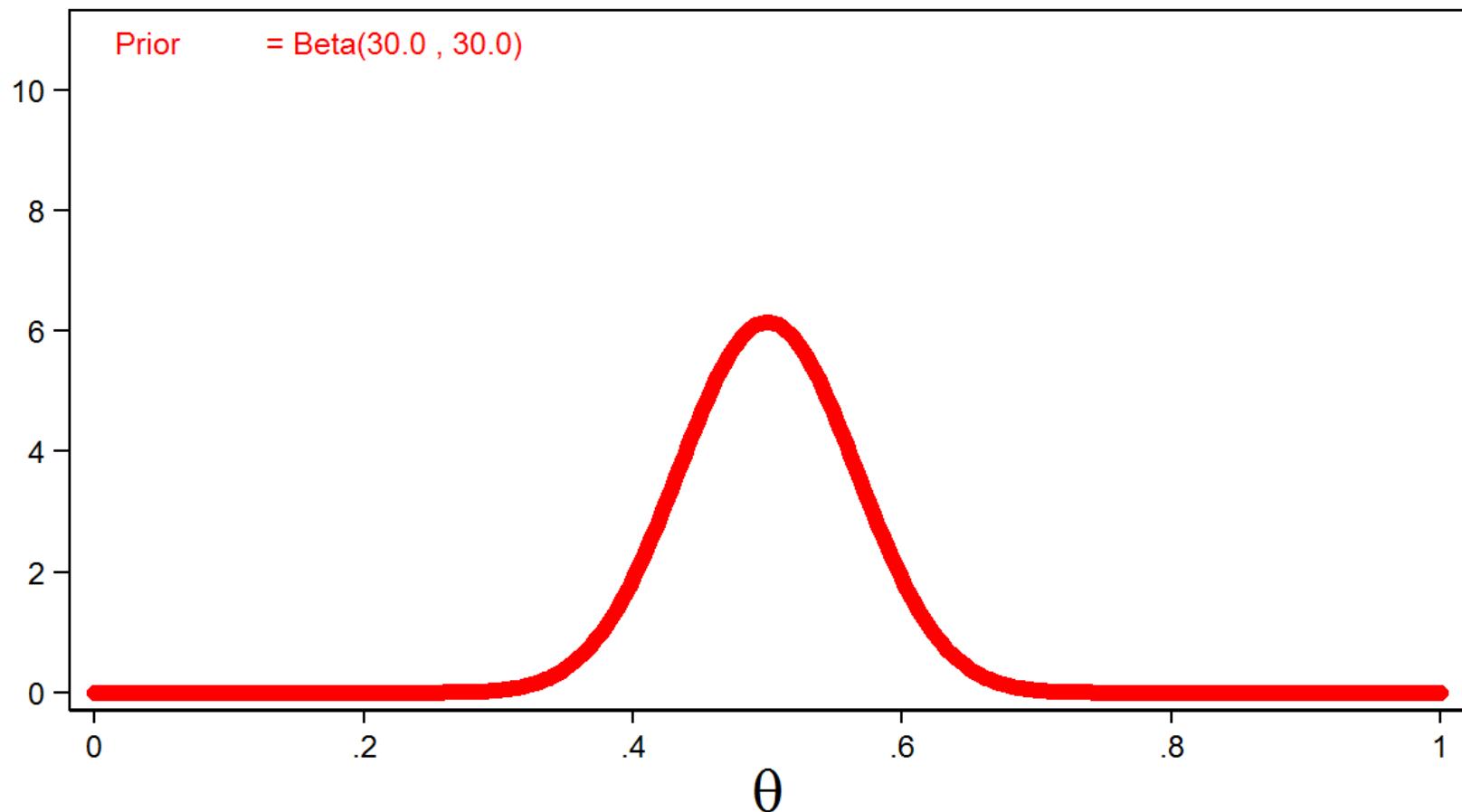
# Uninformative Prior



# Different Priors



# Informative Prior



# Coin Toss Experiment

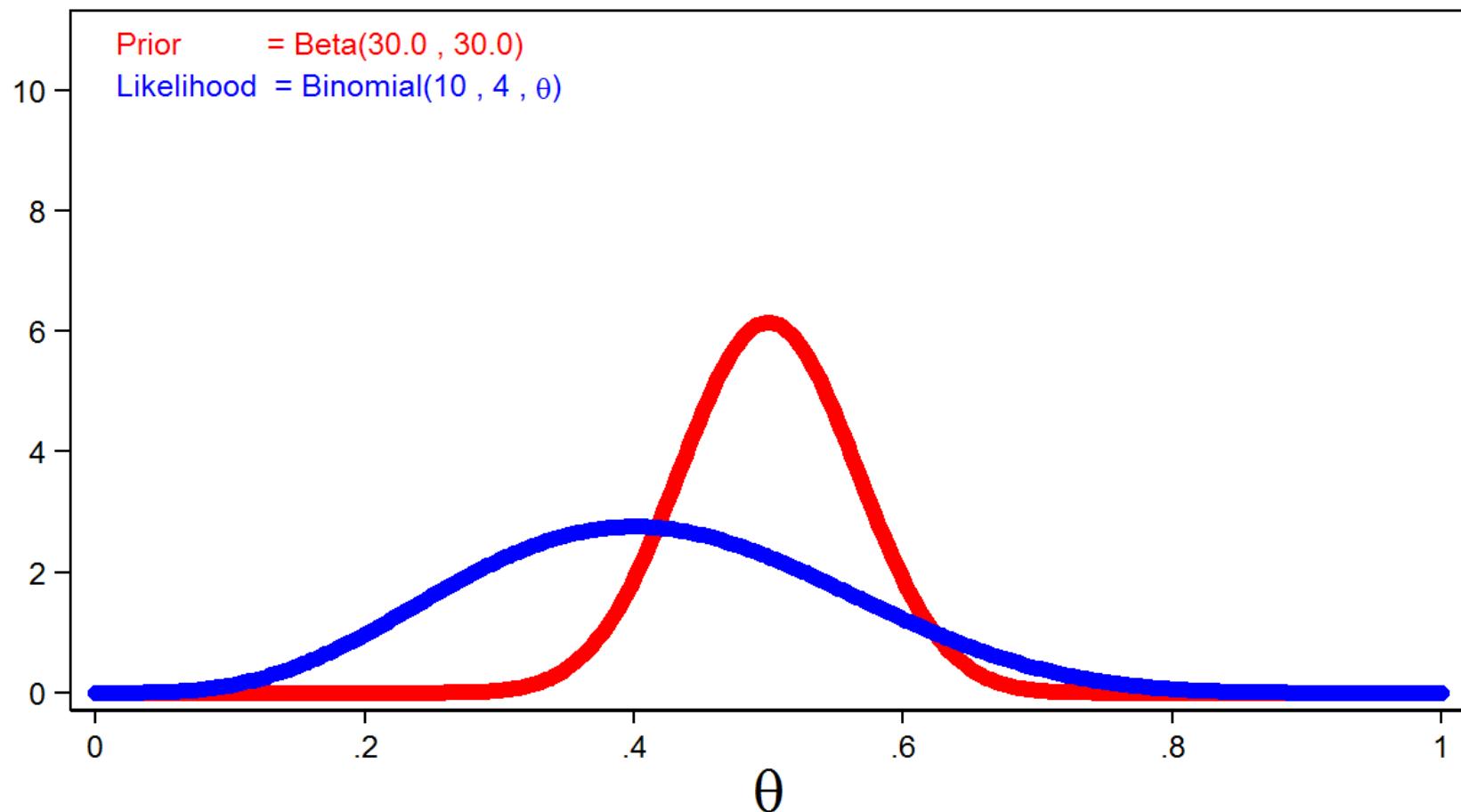


# Likelihood Function for the Data

$$P(y|\theta) = \text{Binomial}(n, \theta)$$

$$= \binom{n}{y} \theta^y (1 - \theta)^{(n-y)}$$

# Prior and Likelihood



# Posterior Distribution

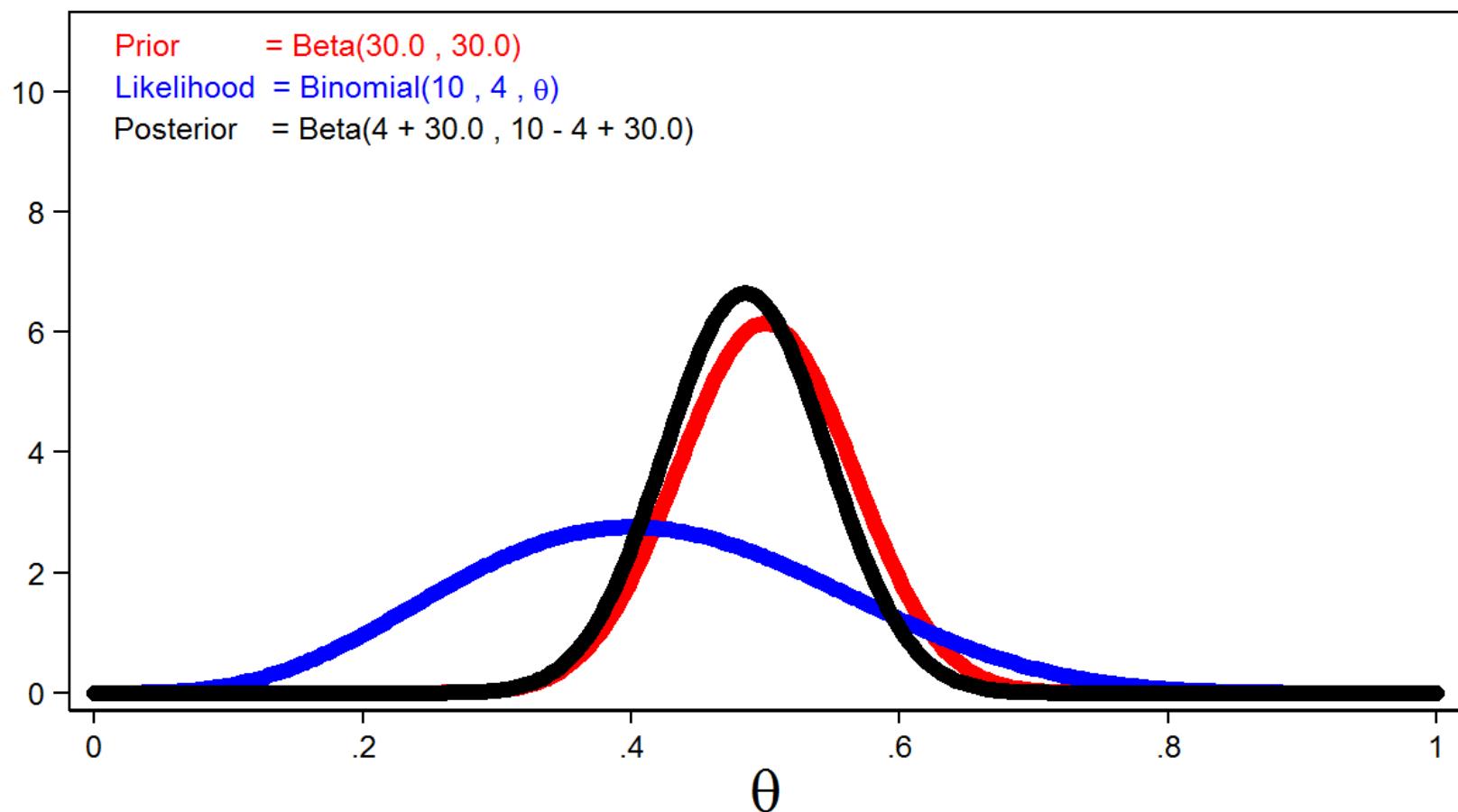
*Posterior = Prior × Likelihood*

$$P(\theta|y) = P(\theta)P(y|\theta)$$

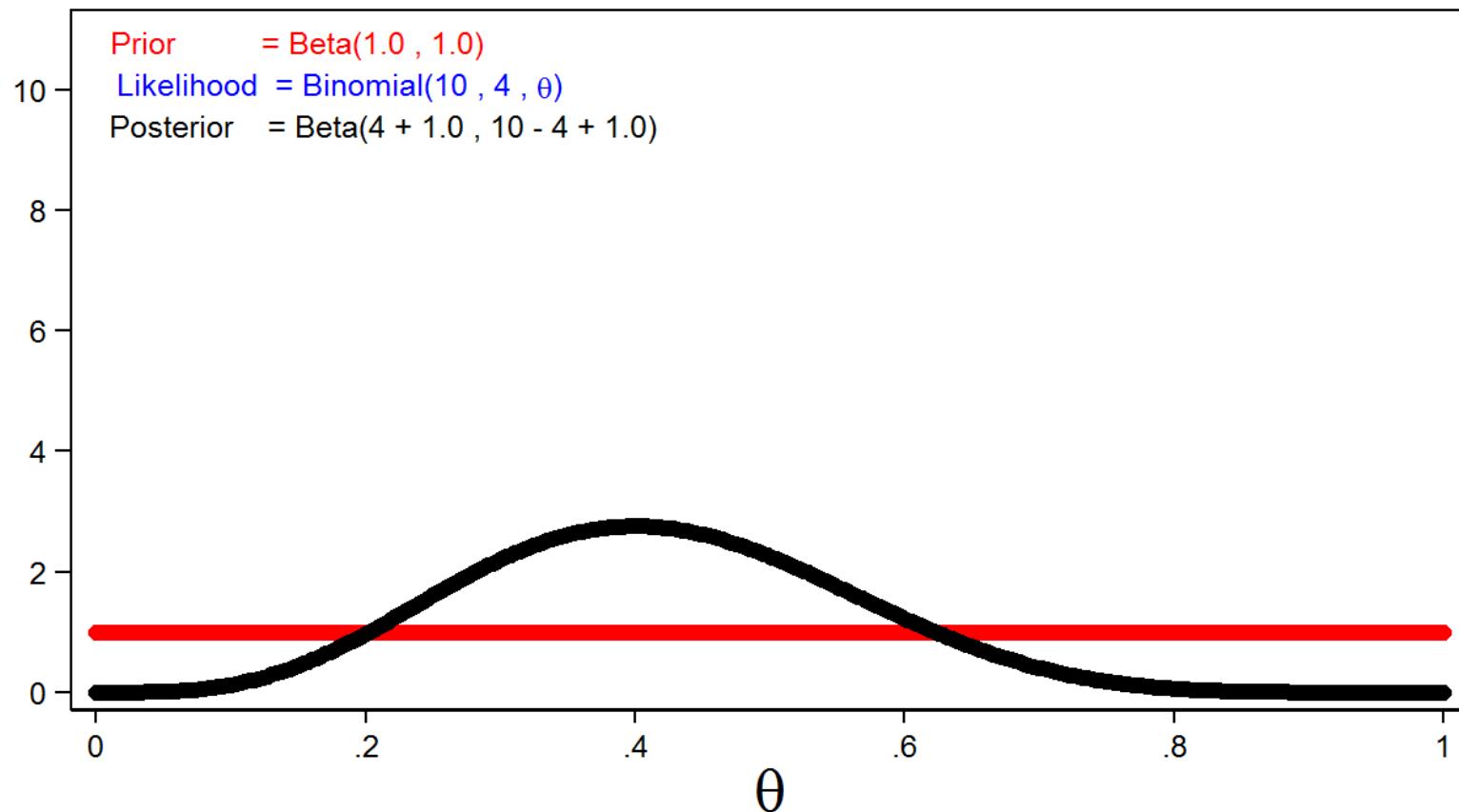
$$P(\theta|y) = \text{Beta}(\alpha, \beta) \times \text{Binomial}(n, \theta)$$

$$= \text{Beta}(y + \alpha, n - y + \beta)$$

# Posterior Distribution



# Effect of Uninformative Prior



# Effect of Informative Prior



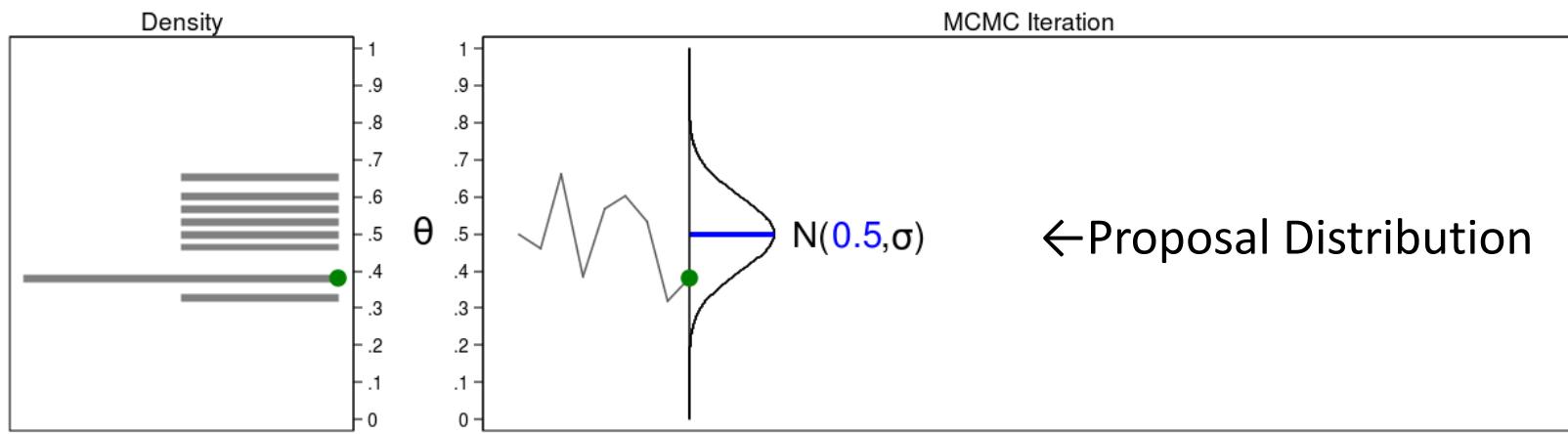
# Markov Chain Monte Carlo

Often the posterior distribution does not have a simple form. We can use Markov Chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm to generate a sample from the posterior distribution.

# MCMC and Metropolis-Hastings

1. Monte Carlo
2. Markov Chains
3. Metropolis-Hastings

# Monte Carlo

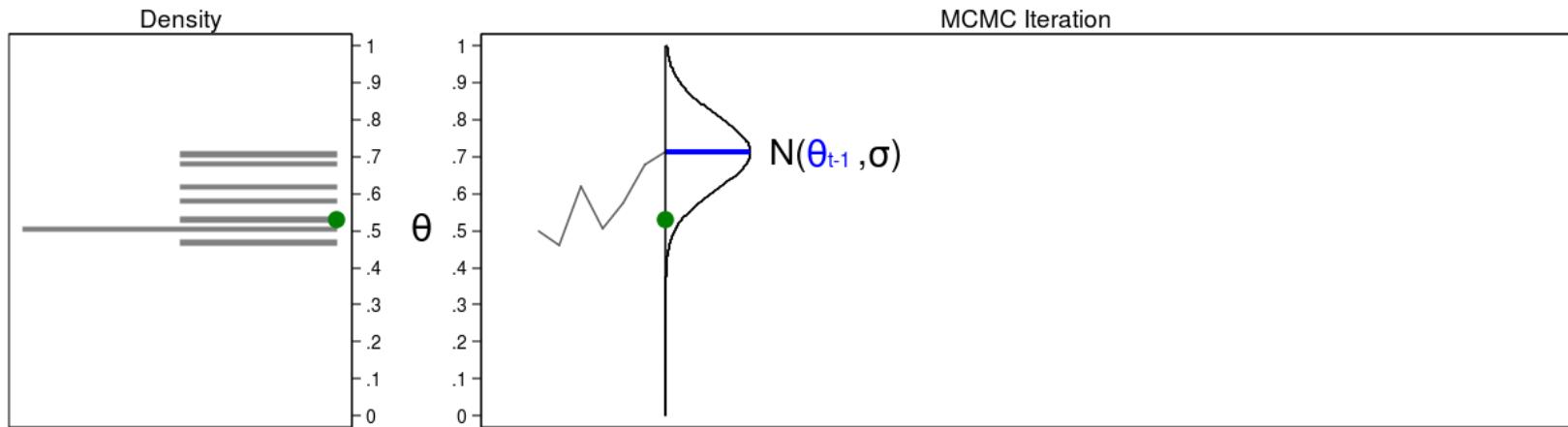


Draw  $\theta_t \sim \text{Normal}(0.5, \sigma) = 0.381$

# Monte Carlo



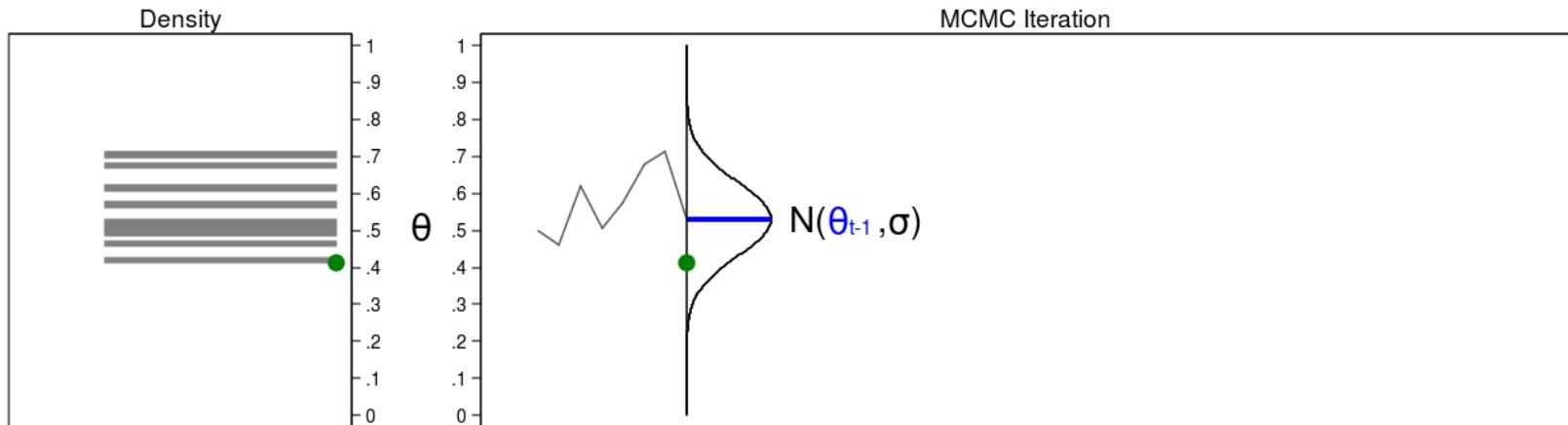
# Markov Chain Monte Carlo



Draw  $\theta_t \sim \text{Normal}(\theta_{t-1}, \sigma)$

$\text{Normal}(0.712, \sigma) = 0.530$

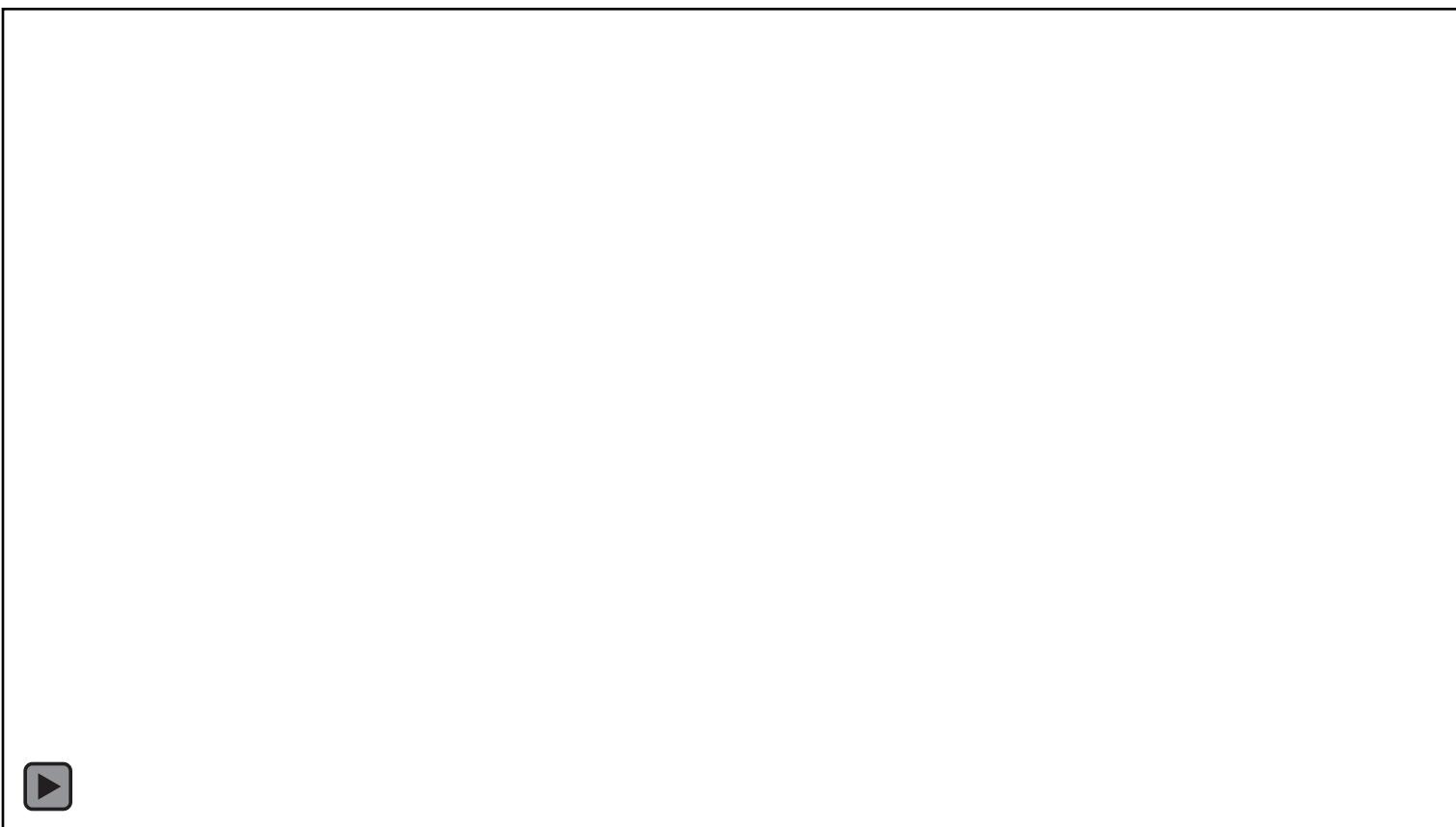
# Markov Chain Monte Carlo



Draw  $\theta_t \sim \text{Normal}(\theta_{t-1}, \sigma)$

$\text{Normal}(0.530, \sigma) = 0.411$

# Markov Chain Monte Carlo



# MCMC with Metropolis-Hastings

$$\begin{aligned} r(\theta_{new}, \theta_{t-1}) &= \frac{\text{Posterior probability of } \theta_{new}}{\text{Posterior probability of } \theta_{t-1}} \\ &= \frac{\text{Beta}(1,1, \theta_{new}) \times \text{Binomial}(10,4, \theta_{new})}{\text{Beta}(1,1, \theta_{t-1}) \times \text{Binomial}(10,4, \theta_{t-1})} \end{aligned}$$

# MCMC with Metropolis-Hastings

*acceptance probability* =  $\alpha(\theta_{new}, \theta_{t-1})$

=  $\min[ r(\theta_{new}, \theta_{t-1}), 1 ]$

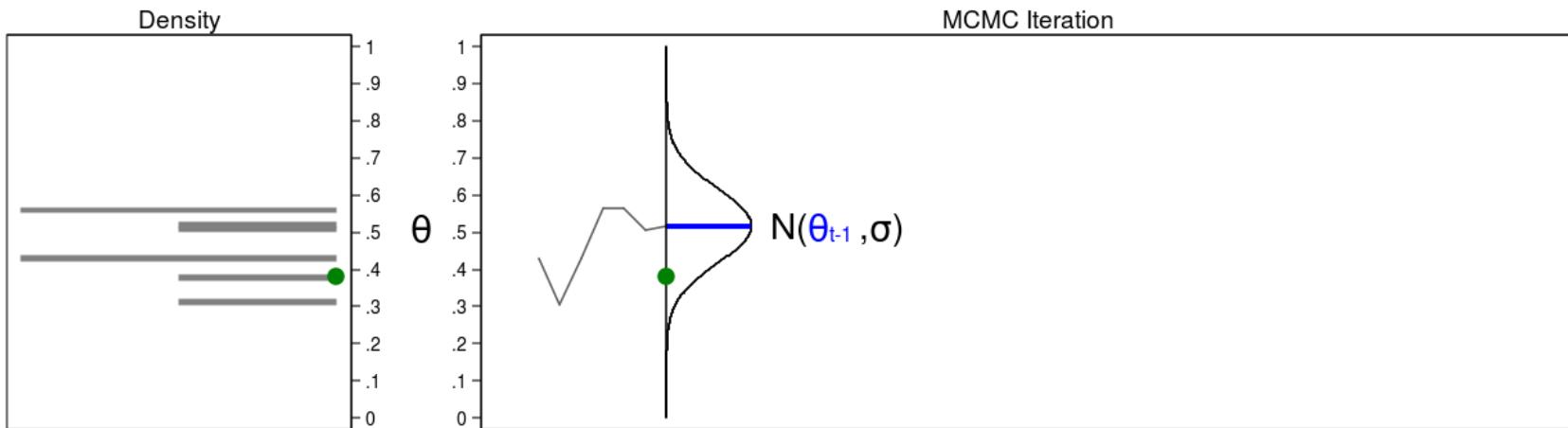
# MCMC with Metropolis-Hastings

*Draw  $u \sim Uniform(0,1)$*

*If     $u < \alpha(\theta_{new}, \theta_{t-1})$     Then  $\theta_t = \theta_{new}$*

*Otherwise  $\theta_t = \theta_{t-1}$*

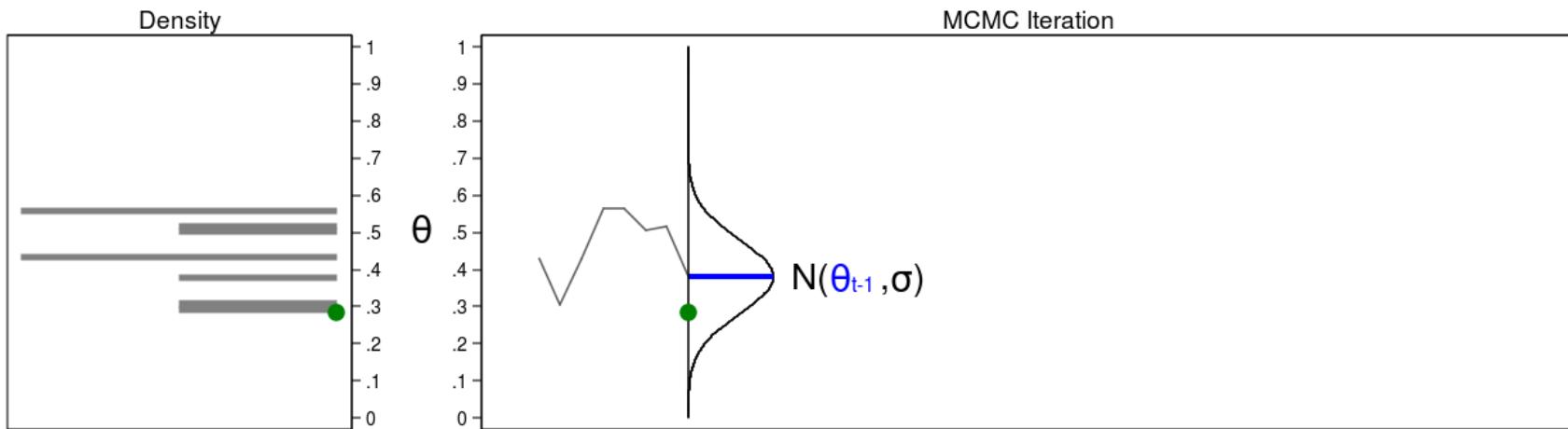
# MCMC with Metropolis-Hastings



$$\text{Step 1: } r(\theta_{\text{new}}, \theta_{t-1}) = \frac{\text{Posterior}(\theta_{\text{new}})}{\text{Posterior}(\theta_{t-1})} = \frac{\text{Beta}(1,1, 0.380) \times \text{Binomial}(10,4, 0.380)}{\text{Beta}(1,1, 0.517) \times \text{Binomial}(10,4, 0.517)} = 1.307$$

$$\text{Step 2: Acceptance probability } \alpha(\theta_{\text{new}}, \theta_{t-1}) = \min\{r(\theta_{\text{new}}, \theta_{t-1}), 1\} = \min\{1.000, 1\} = 1.000$$

# MCMC with Metropolis-Hastings



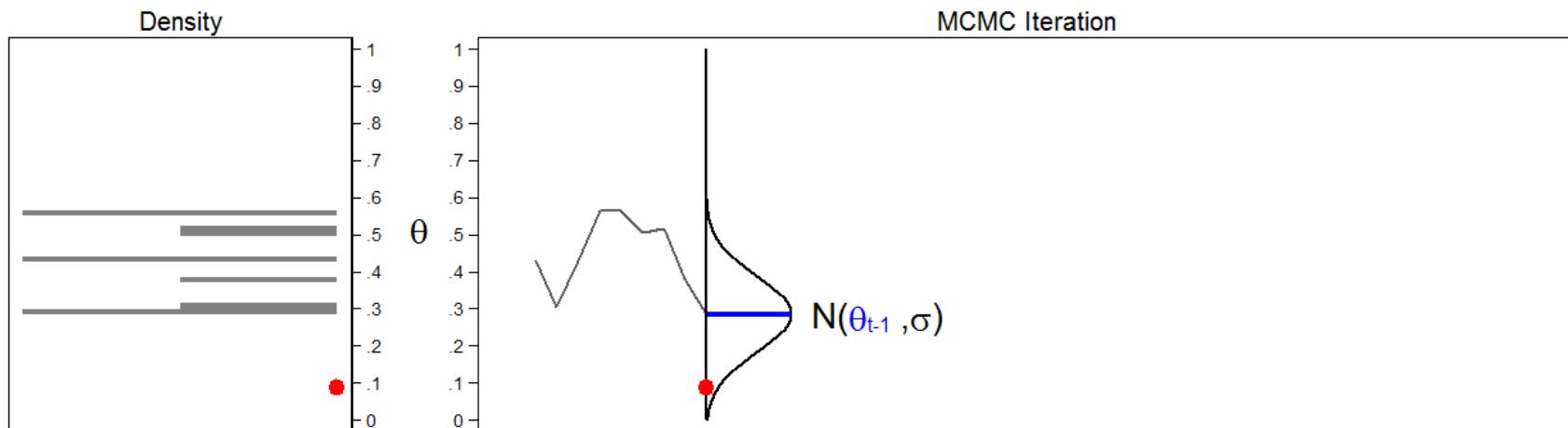
Step 1:  $r(\theta_{\text{new}}, \theta_{t-1}) = \frac{\text{Posterior}(\theta_{\text{new}})}{\text{Posterior}(\theta_{t-1})} = \frac{\text{Beta}(1,1, 0.286) \times \text{Binomial}(10,4, 0.286)}{\text{Beta}(1,1, 0.380) \times \text{Binomial}(10,4, 0.380)} = 0.747$

Step 2: Acceptance probability  $\alpha(\theta_{\text{new}}, \theta_{t-1}) = \min\{r(\theta_{\text{new}}, \theta_{t-1}), 1\} = \min\{0.747, 1\} = 0.747$

Step 3: Draw  $u \sim \text{Uniform}(0,1) = 0.094$

Step 4: If  $u < \alpha(\theta_{\text{new}}, \theta_{t-1}) \rightarrow$  If  $0.094 < 0.747$  Then  $\theta_t = \theta_{\text{new}} = 0.286$   
Otherwise  $\theta_t = \theta_{t-1} = 0.380$

# MCMC with Metropolis-Hastings



Step 1:  $r(\theta_{\text{new}}, \theta_{t-1}) = \frac{\text{Posterior}(\theta_{\text{new}})}{\text{Posterior}(\theta_{t-1})} = \frac{\text{Beta}(1,1, 0.088) \times \text{Binomial}(10,4, 0.088)}{\text{Beta}(1,1, 0.286) \times \text{Binomial}(10,4, 0.286)} = 0.039$

Step 2: Acceptance probability  $\alpha(\theta_{\text{new}}, \theta_{t-1}) = \min\{r(\theta_{\text{new}}, \theta_{t-1}), 1\} = \min\{0.039, 1\} = 0.039$

Step 3: Draw  $u \sim \text{Uniform}(0,1) = 0.247$

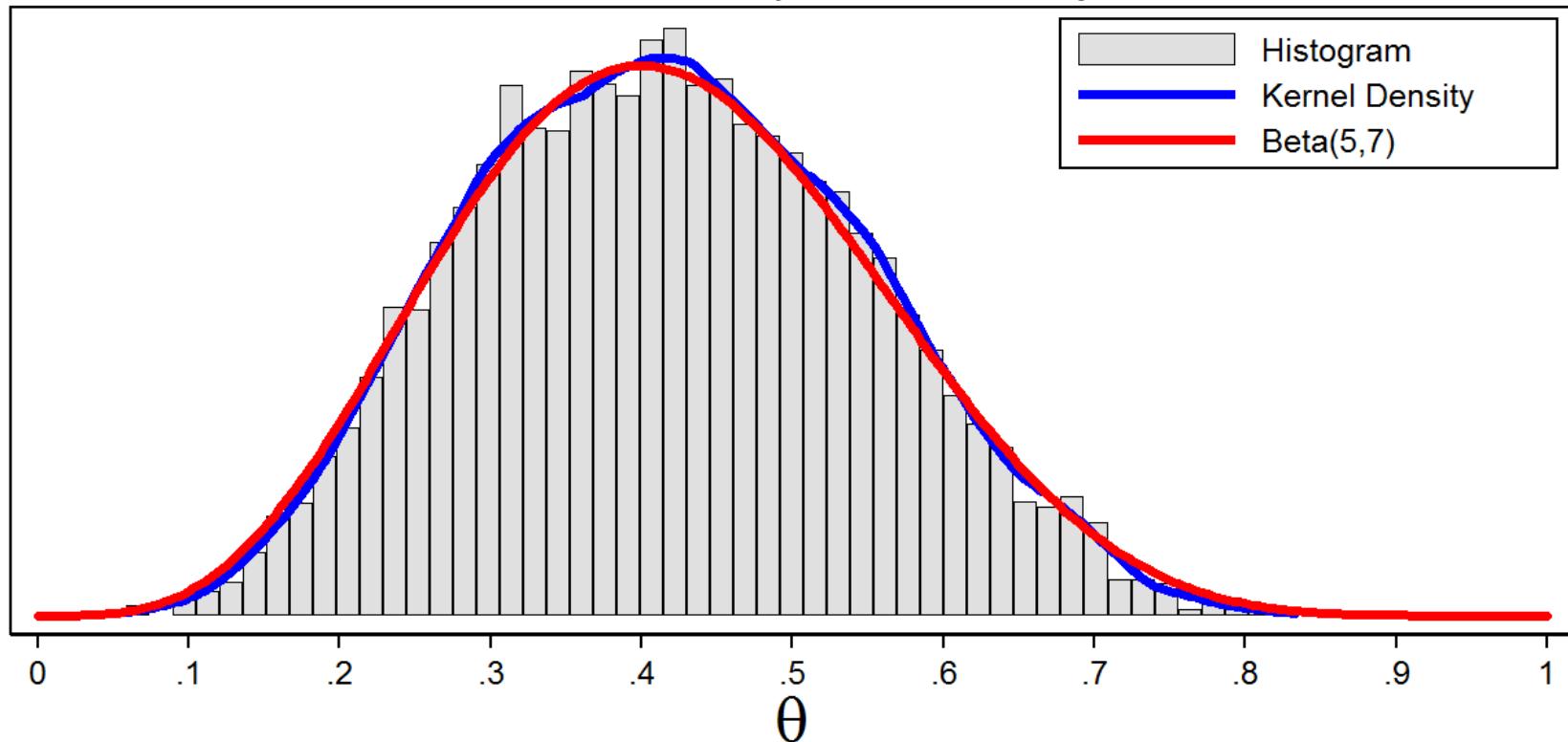
Step 4: If  $u < \alpha(\theta_{\text{new}}, \theta_{t-1}) \rightarrow$  If  $0.247 < 0.039$  Then  $\theta_t = \theta_{\text{new}} = 0.088$   
Otherwise  $\theta_t = \theta_{t-1} = 0.286$

# MCMC with Metropolis-Hastings



# MCMC with Metropolis-Hastings

Comparison of the MCMC sample and  
the theoretical posterior density



# MCMC with Gibbs Sampling



# The **bayesmh** command

```
bayesmh heads,                                ///
    likelihood(dbernoulli({theta}))           ///
    prior({theta}, beta(1,1))
```

```
. bayesmh heads, likelihood(dbernoulli({theta})) prior({theta}, beta(1,1))
```

Burn-in ...

Simulation ...

Model summary

---

Likelihood:

```
heads ~ bernoulli({theta})
```

Prior:

```
{theta} ~ beta(1,1)
```

---

Bayesian Bernoulli model

Random-walk Metropolis-Hastings sampling

MCMC iterations = 12,500

Burn-in = 2,500

MCMC sample size = 10,000

Number of obs = 10

Acceptance rate = .4823

Efficiency = .2291

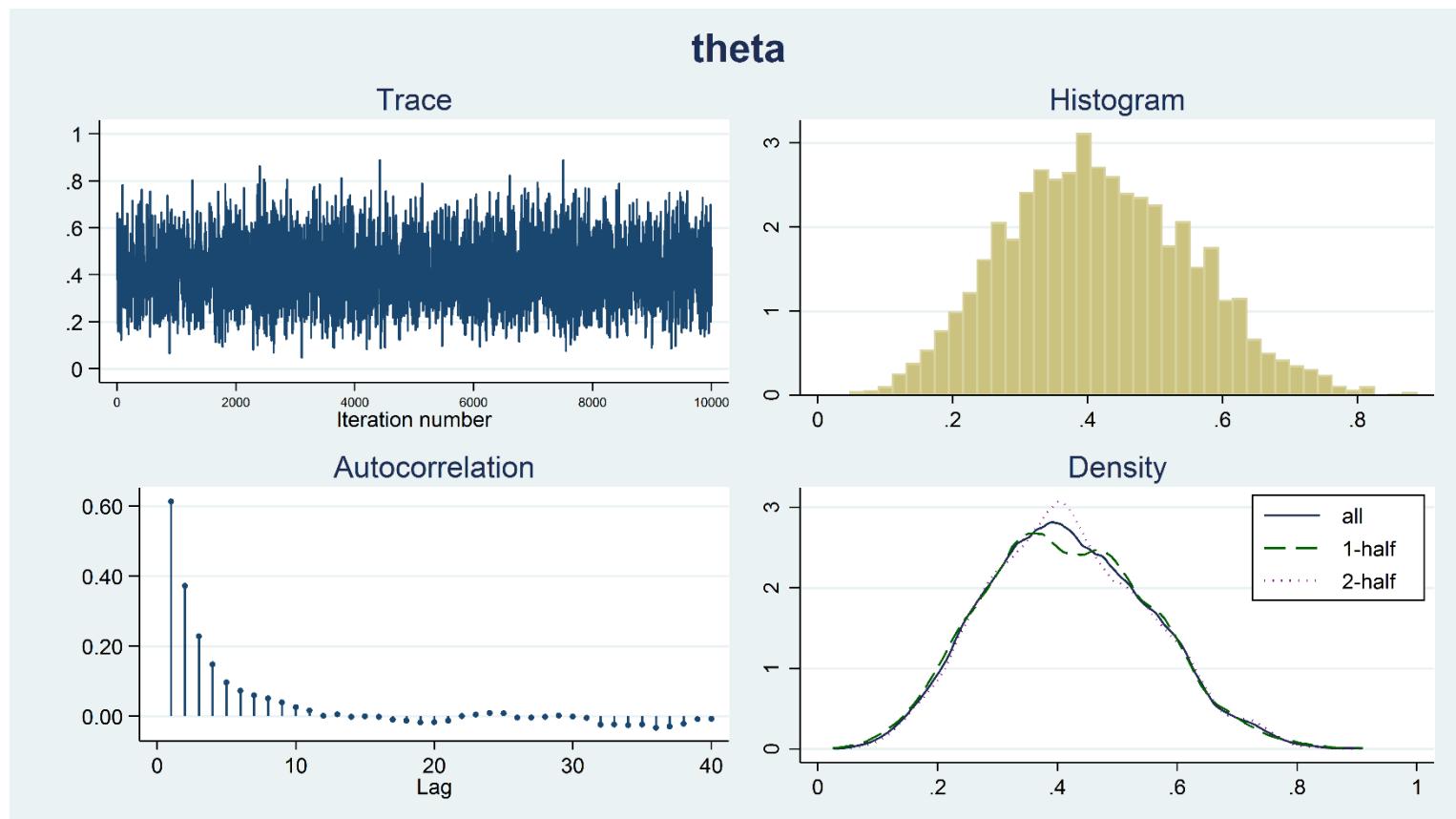
Log marginal likelihood = -7.8194591

---

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
theta	.4187117	.1342192	.002804	.4152274	.1746616	.6876875

---

# Diagnostic Plots



```
bayesgraph diagnostics {theta}
```

# Outline

- Introduction to Bayesian Analysis
- Coin Toss Example
- Priors, Likelihoods, and Posteriors
- Markov Chain Monte Carlo (MCMC)
- **Bayesian Linear Regression**
- Advantages and Disadvantages of Bayes

# Bayesian Linear Regression

```
. describe
```

Contains data from test\_1level.dta

obs: 1,000

vars: 5

15 Jun 2019 11:10

variable	name	storage	display	value	variable	label
		type	format	label		
student		float	%9.0g		Student Identification Number	
sex		float	%9.0g	sex	Sex	
age		float	%9.0g		Age (years)	
cage		float	%9.0g		age - 16	
score		float	%9.0g		Math Test Score	

Sorted by: student

$$score_i = \beta_0 + \beta_1 age_i + \beta_2 sex_i + e_i$$

. regress score age sex

Source	SS	df	MS	Number of obs	=	1,000
Model	1694.46256	2	847.231282	F(2, 997)	=	32.96
Residual	25626.6374	997	25.7037487	Prob > F	=	0.0000
Total	27321.1	999	27.3484484	R-squared	=	0.0620

score	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.4631148	.1112826	4.16	0.000	.2447399 .6814897
sex	2.209304	.3207331	6.89	0.000	1.579915 2.838694
_cons	69.99239	1.786002	39.19	0.000	66.48763 73.49714

```
. bayes: regress score age sex
```

Burn-in ...

Simulation ...

Model summary

---

Likelihood:

```
score ~ regress(xb_score,{sigma2})
```

Priors:

{score:age sex _cons} ~ normal(0,10000)	(1)
{sigma2} ~ igamma(.01,.01)	

---

(1) Parameters are elements of the linear form xb\_score.

Bayesian linear regression	MCMC iterations = 12,500
Random-walk Metropolis-Hastings sampling	Burn-in = 2,500
	MCMC sample size = 10,000
	Number of obs = 1,000
	Acceptance rate = .3355
	Efficiency: min = .07576
	avg = .1215
Log marginal-likelihood = -3066.8217	max = .2365

---

	Mean	Std. Dev.	MCSE	Median	Equal-tailed	
					[95% Cred. Interval]	
score						
age	.4616019	.1120901	.003935	.4674614	.247683	.6673255
sex	2.212278	.3130193	.010299	2.207893	1.613654	2.828866
_cons	70.02365	1.800004	.065397	69.93342	66.66458	73.56361
sigma2	25.77379	1.187874	.024424	25.71326	23.60044	28.2026

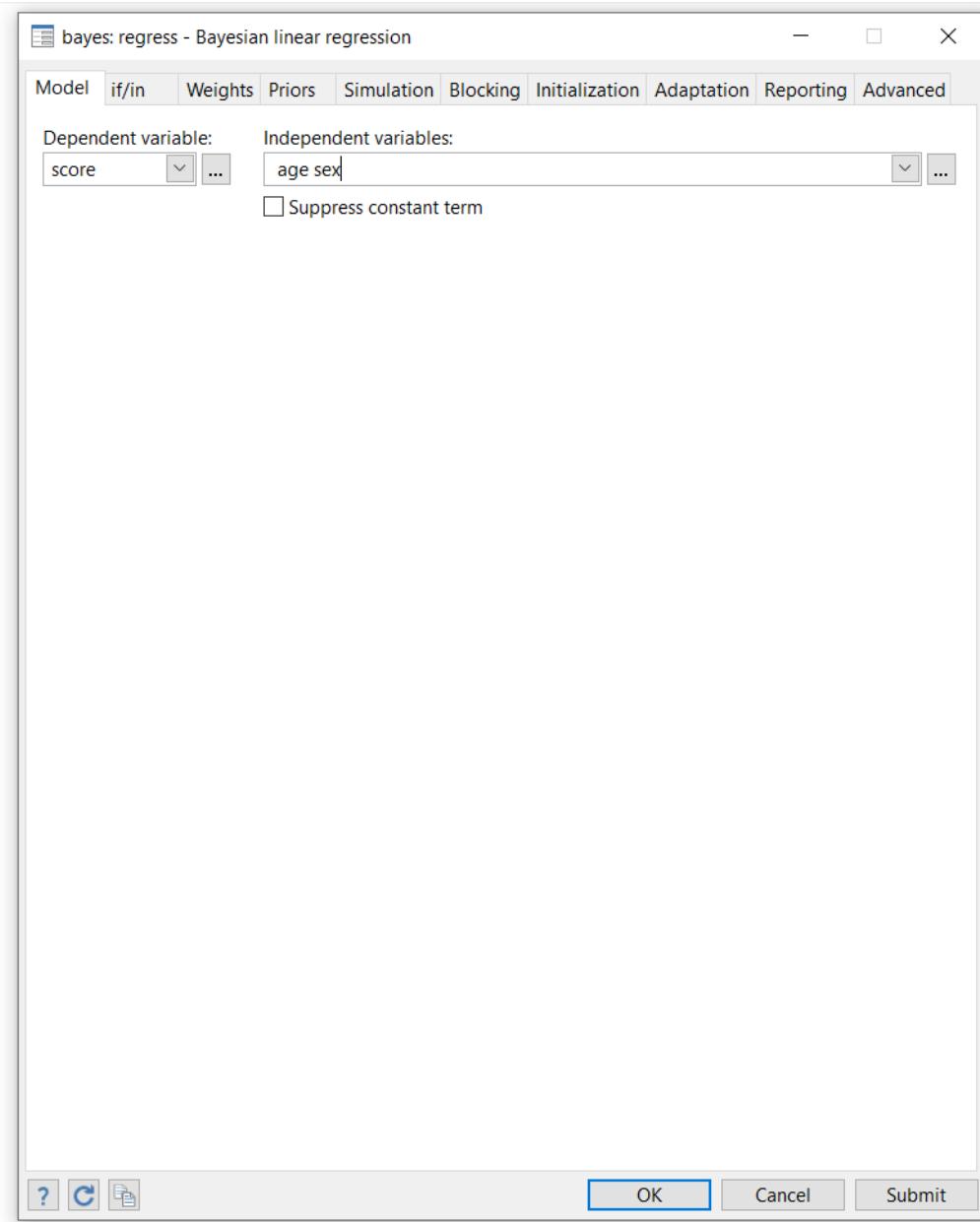
---

Note: Default priors are used for model parameters.

```
. bayesstats summary {score:_cons age sex} {sigma2}
```

Posterior summary statistics                                    MCMC sample size = 10,000

	Mean	Std. Dev.	MCSE	Median	Equal-tailed	
					[95% Cred. Interval]	
score						
_cons	70.00025	1.745976	.06218	69.99261	66.46578	73.53115
age	.4623579	.1084987	.00368	.4605215	.2424589	.6794042
sex	2.223806	.3206964	.012034	2.220988	1.595755	2.848643
sigma2	25.78444	1.185872	.025164	25.75575	23.62021	28.25372



bayes: regress - Bayesian linear regression

Model if/in Weights Priors Simulation Blocking Initialization Adaptation Reporting Advanced

Use Gibbs sampling to update model parameters

Default priors

Normal prior for coefficients  
100 Standard deviation

Inverse-gamma prior for variance components  
0.01 Shape  
0.01 Scale

Custom priors for model parameters

Press "Create" to define a prior distribution

Show model summary without estimation

?  

OK Cancel Submit

bayes: regress - Bayesian linear regression

Model if/in Weights Priors Simulation Blocking Initialization Adaptation Reporting Advanced

Use Gibbs sampling to update model parameters

Default priors

Normal prior for coefficients  
100 Standard deviation

Inverse-gamma prior for variance components  
0.01 Shape  
0.01 Scale

Custom priors for model parameters

Press "Create" to define a prior distribution

Show model summary without estimation

Prior 1

Parameters specification:  
(score:\_cons)

Choose a prior distribution:

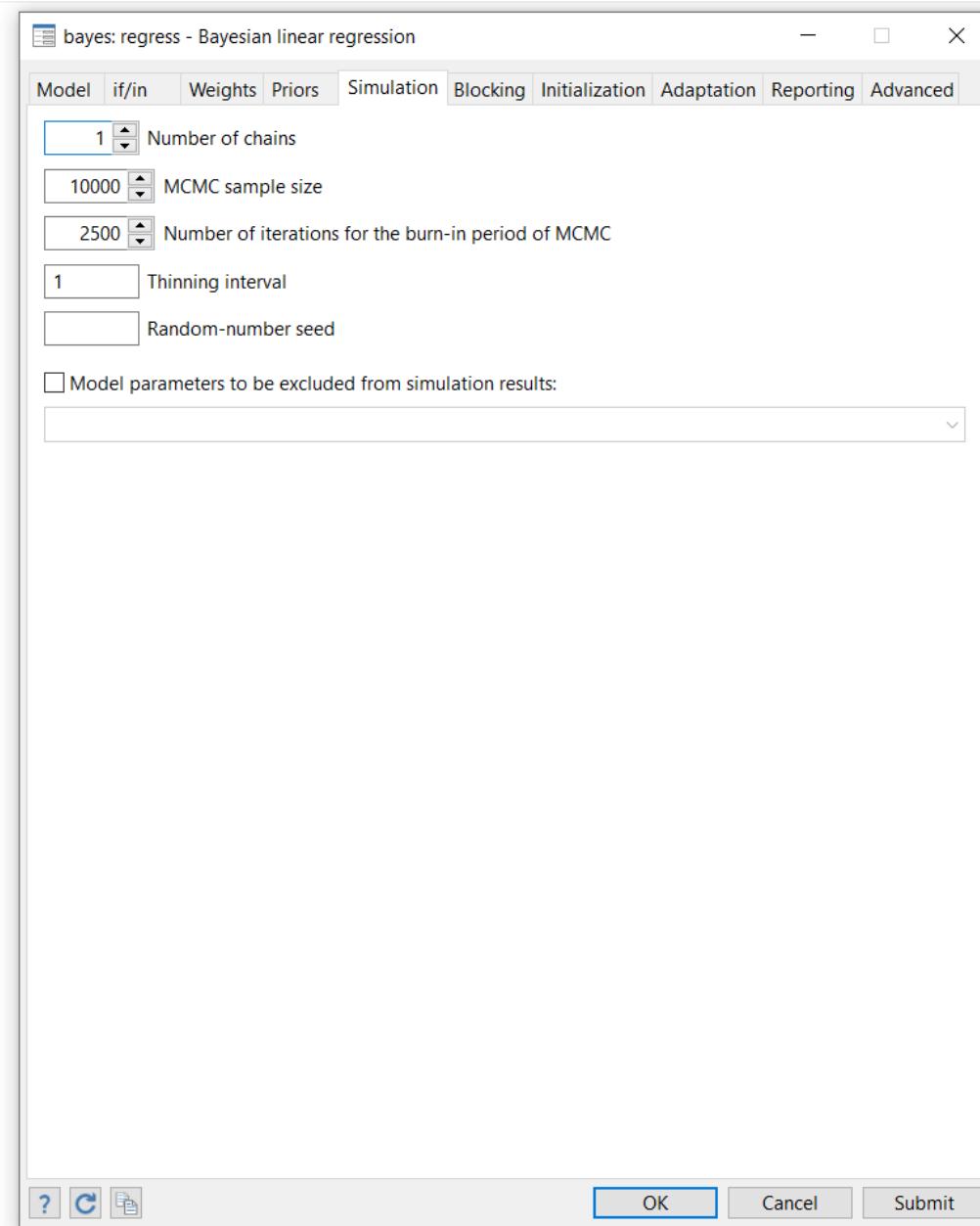
Univariate continuous  
--> Normal distribution  
--> Generalized t distribution  
--> Lognormal distribution  
--> Uniform distribution  
--> Gamma distribution  
--> Inverse gamma distribution  
--> Exponential distribution  
--> Laplace distribution  
--> Cauchy distribution  
--> Beta distribution  
--> Chi-squared distribution  
--> Pareto distribution  
--> Jeffreys prior for variance of normal distribution

Multivariate continuous  
--> Multivariate normal distribution  
--> Multivariate normal distribution with zero mean  
--> Zellner's g-prior  
--> Zellner's g-prior with zero mean  
--> Dirichlet distribution  
--> Wishart distribution  
--> Inverse Wishart distribution  
--> Jeffreys prior for covariance of multivariate normal

Lower endpoint:  
-1000

Upper endpoint:  
1000

?



bayes: regress - Bayesian linear regression

Model if/in Weights Priors Simulation Blocking Initialization Adaptation Reporting Advanced

Adaptive MCMC procedure

100 Adaptation interval  
25 Maximum number of adaptive iterations  
5 Minimum number of adaptive iterations  
0.75 Parameter controlling acceptance rate, alpha()  
0.8 Parameter controlling proposal covariance, beta()  
0 Parameter controlling adaptation rate, gamma()  
Target acceptance rate for all blocks of model parameters  
0.01 Tolerance for acceptance rate

2.38 Initial multiplier for the scale factor for all blocks  
Input matrix... Scale matrix for initial proposal covariance

?

OK Cancel Submit

# bayes options

. bayes: regress score age sex, noheader

Burn-in ...

Simulation ...

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4678708	.1105634	.003533	.4727448	.2565501	.6867787
sex	2.204918	.3185867	.010859	2.215316	1.586366	2.825122
_cons	69.91938	1.765032	.054985	69.88605	66.44711	73.31774
sigma2	25.7486	1.156099	.022763	25.66118	23.55735	28.13625

# bayes options

. bayes, rseed(15): regress score age sex, noheader

Burn-in ...

Simulation ...

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4616019	.1120901	.003935	.4674614	.247683	.6673255
sex	2.212278	.3130193	.010299	2.207893	1.613654	2.828866
_cons	70.02365	1.800004	.065397	69.93342	66.66458	73.56361
sigma2	25.77379	1.187874	.024424	25.71326	23.60044	28.2026

# bayes options

```
. bayes, mcmcsize(20000) rseed(15): regress score age sex, noheader
```

Burn-in ...

Simulation ...

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4625807	.1121395	.002667	.4621873	.2485448	.6801277
sex	2.20712	.3192993	.007611	2.206704	1.585289	2.830943
_cons	70.00382	1.795897	.043426	70.0035	66.53197	73.47022
sigma2	25.75867	1.180937	.016997	25.70115	23.59171	28.19241

# bayes options

. bayes, prior({score:\_cons}, uniform(0,100000)): regress score age sex, noheader

Burn-in ...

Simulation ...

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4062103	.0826128	.022212	.4147312	.2580627	.5411578
sex	2.162167	.3420588	.049342	2.159435	1.453007	2.785148
_cons	70.92467	1.338897	.361782	70.86285	68.74958	73.29666
sigma2	25.81098	1.202767	.025871	25.78483	23.54569	28.2141

# bayes options

```
. bayes, prior({score:}, flat) rseed(15): regress score age sex, noheader
```

Burn-in ...

Simulation ...

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4623579	.1084987	.00368	.4605215	.2424589	.6794042
sex	2.223806	.3206964	.012034	2.220988	1.595755	2.848643
_cons	70.00025	1.745976	.06218	69.99261	66.46578	73.53115
sigma2	25.78444	1.185872	.025164	25.75575	23.62021	28.25372

# bayes options

```
. bayes, nchains(5) rseed(15) : regress score age sex
```

Chain 1

  Burn-in ...

  Simulation ...

Chain 2

  Burn-in ...

  Simulation ...

Chain 3

  Burn-in ...

  Simulation ...

Chain 4

  Burn-in ...

  Simulation ...

Chain 5

  Burn-in ...

  Simulation ...

# bayes options

Bayesian linear regression

Random-walk Metropolis-Hastings sampling

Number of chains = 5

Per MCMC chain:

Iterations = 12,500

Burn-in = 2,500

Sample size = 10,000

Number of obs = 1,000

Avg acceptance rate = .3393

Avg efficiency: min = .08578

avg = .1255

max = .228

Avg log marginal-likelihood = -3066.8405

Max Gelman-Rubin Rc = 1.001

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
score						
age	.4656843	.1128116	.001633	.4653546	.2438611	.6862797
sex	2.209279	.3171939	.004843	2.207893	1.59085	2.842695
_cons	69.95217	1.814761	.026652	69.94621	66.38668	73.56682
sigma2	25.76854	1.16008	.010866	25.73012	23.60044	28.13613

Note: Default priors are used for model parameters.

Note: Default initial values are used for multiple chains.

# Checking “Convergence” of the Chain

- Gelman-Rubin Convergence Diagnostics
- Effective Sample Size
- Trace Plots
- Histograms
- Correlograms
- Scatterplot Matrices

# Checking “Convergence” of the Chain

```
. bayesstats grubin _all
```

Gelman-Rubin convergence diagnostic

Number of chains = 5  
MCMC size, per chain = 10,000  
Max Gelman-Rubin Rc = 1.000553

	Rc
score	
age	1.000363
sex	1.000553
_cons	1.000417
sigma2	1.000228

Convergence rule: Rc < 1.1

# Checking “Convergence” of the Chain

```
. bayesstats ess
```

```
Efficiency summaries      MCMC sample size =    10,000
```

	ESS	Corr. time	Efficiency
score			
age	869.23	11.50	0.0869
sex	710.17	14.08	0.0710
_cons	788.46	12.68	0.0788
sigma2	2220.80	4.50	0.2221

(Single chain)

# Checking “Convergence” of the Chain

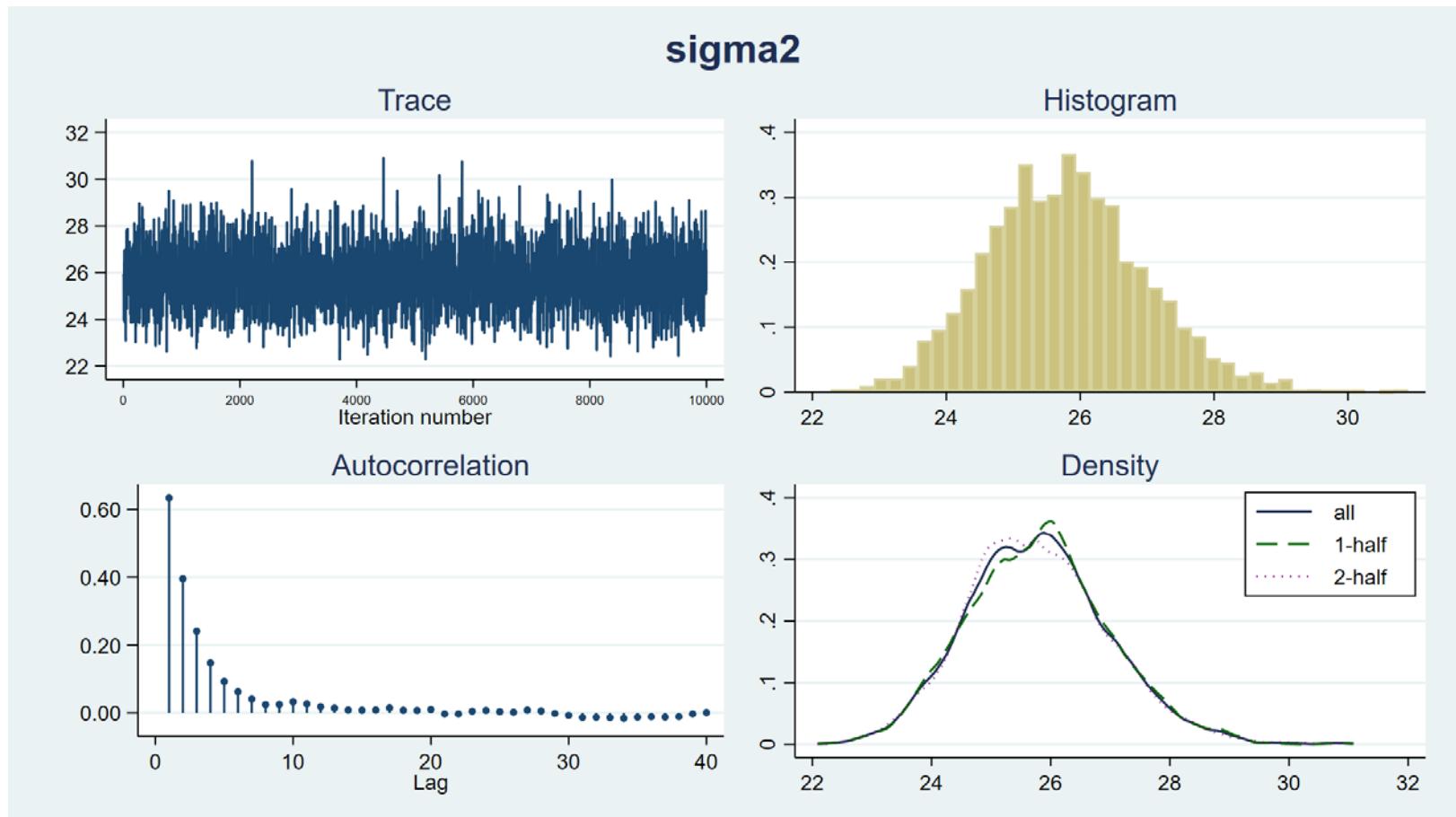
```
. bayesstats ess
```

```
Efficiency summaries      Number of chains =      5
                           MCMC sample size = 50,000
                           Efficiency: min = .08578
                                         avg = .1255
                                         max = .228
```

	ESS	Corr. time	Efficiency
score			
age	4774.05	10.47	0.0955
sex	4288.94	11.66	0.0858
_cons	4636.49	10.78	0.0927
sigma2	11399.18	4.39	0.2280

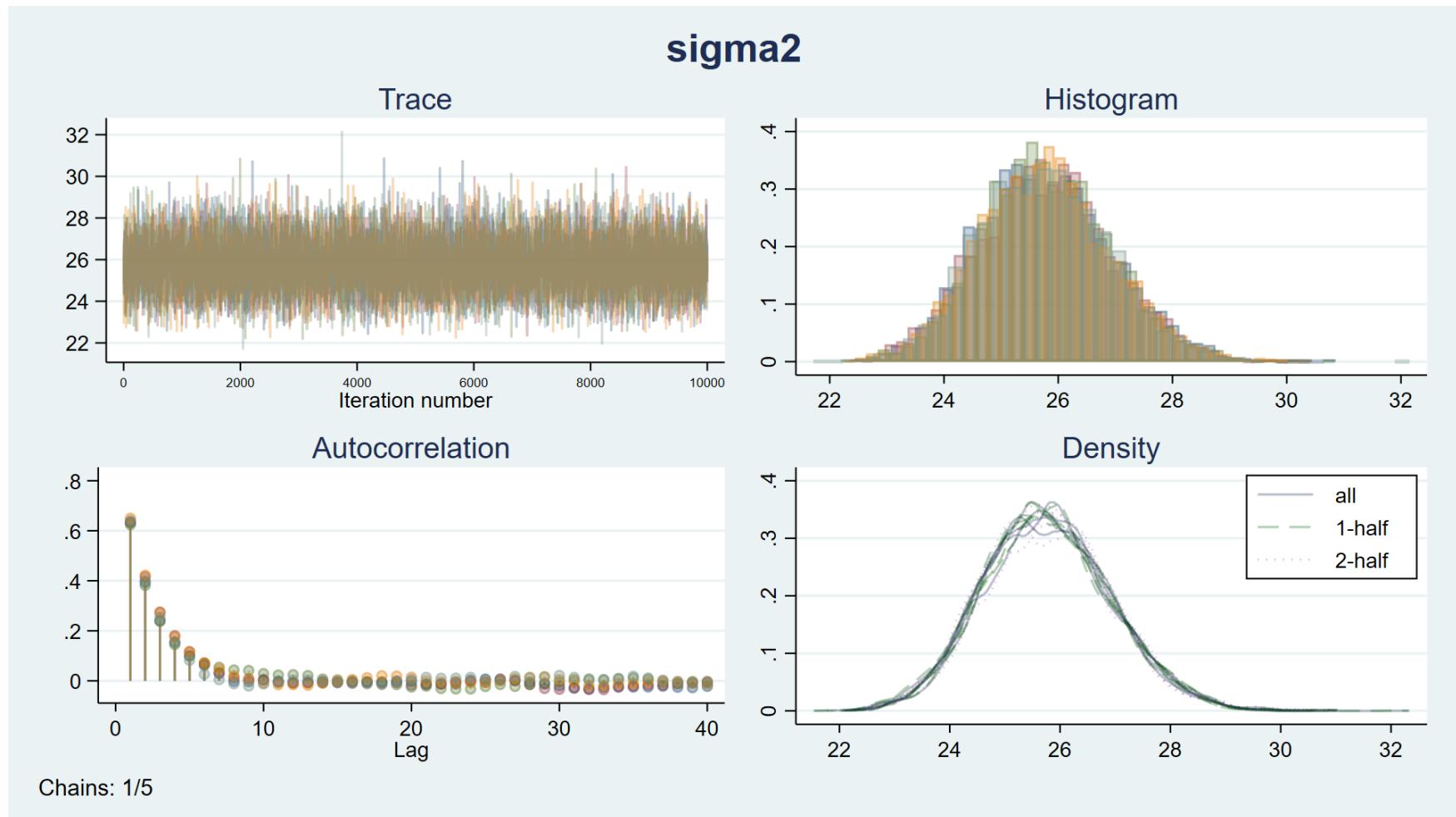
(Multiple chains)

# Checking “Convergence” of the Chain



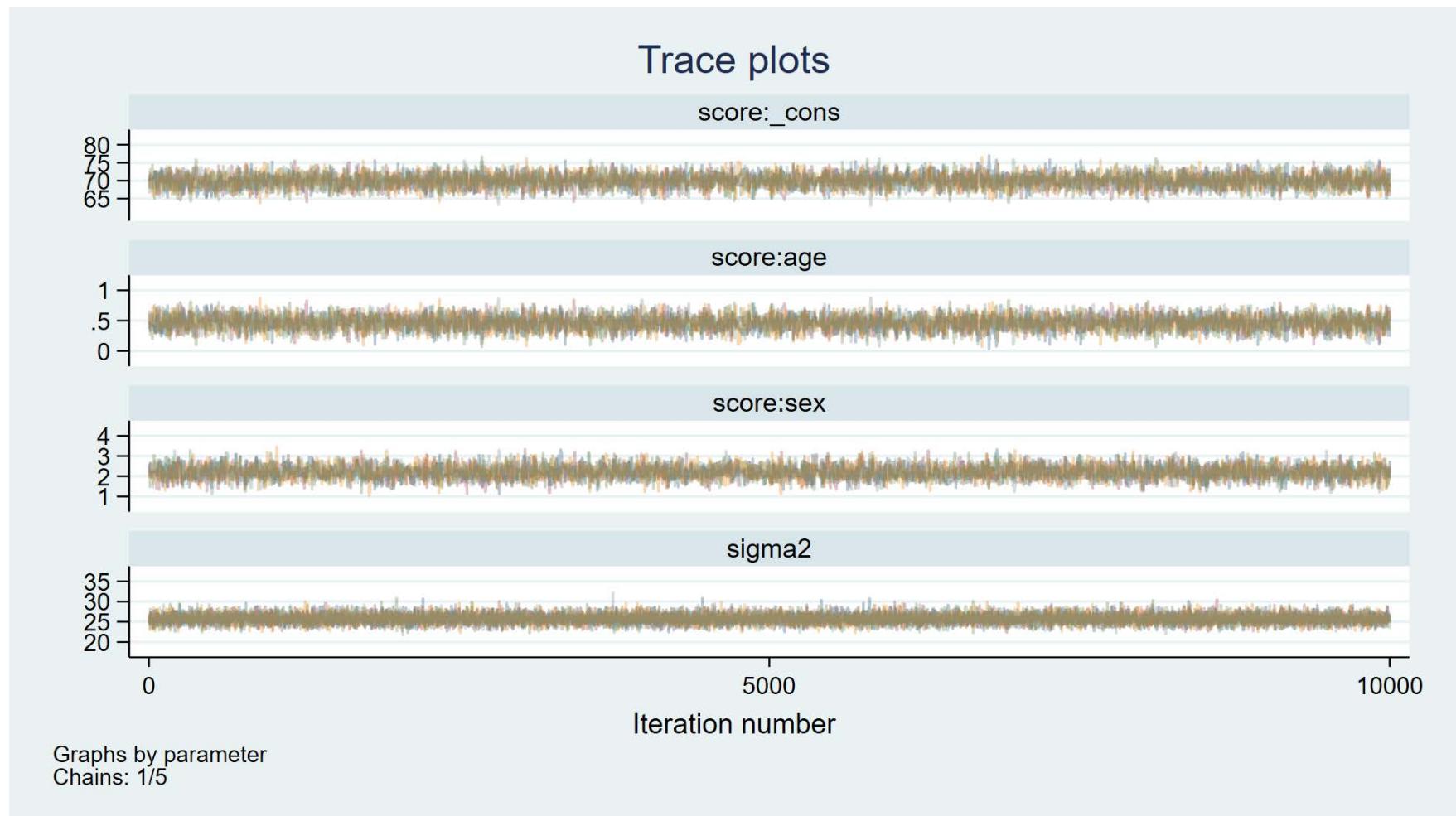
`bayesgraph diagnostics {sigma2}`

# Checking “Convergence” of the Chain



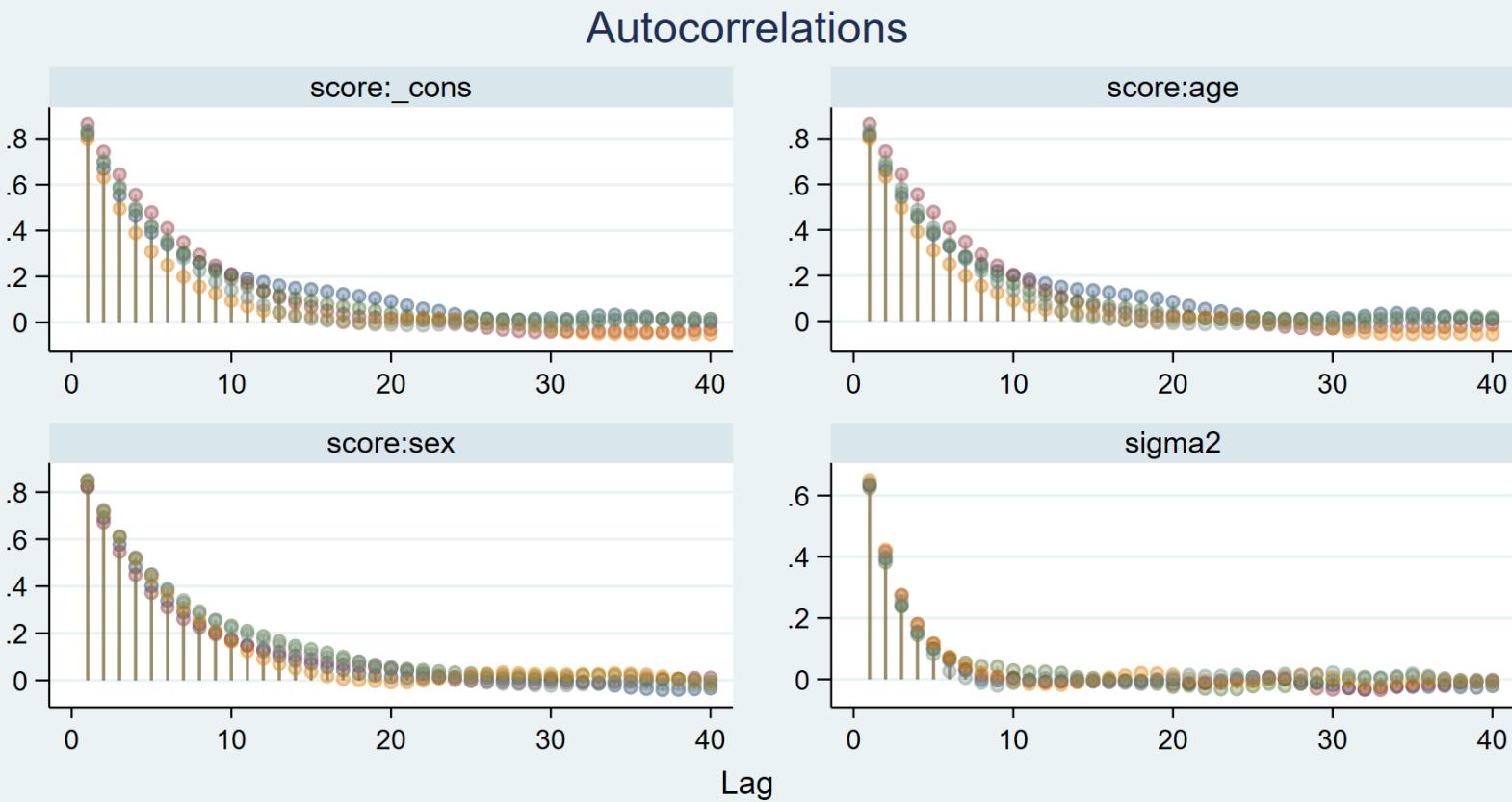
**bayesgraph diagnostics {sigma2}**

# Checking “Convergence” of the Chain



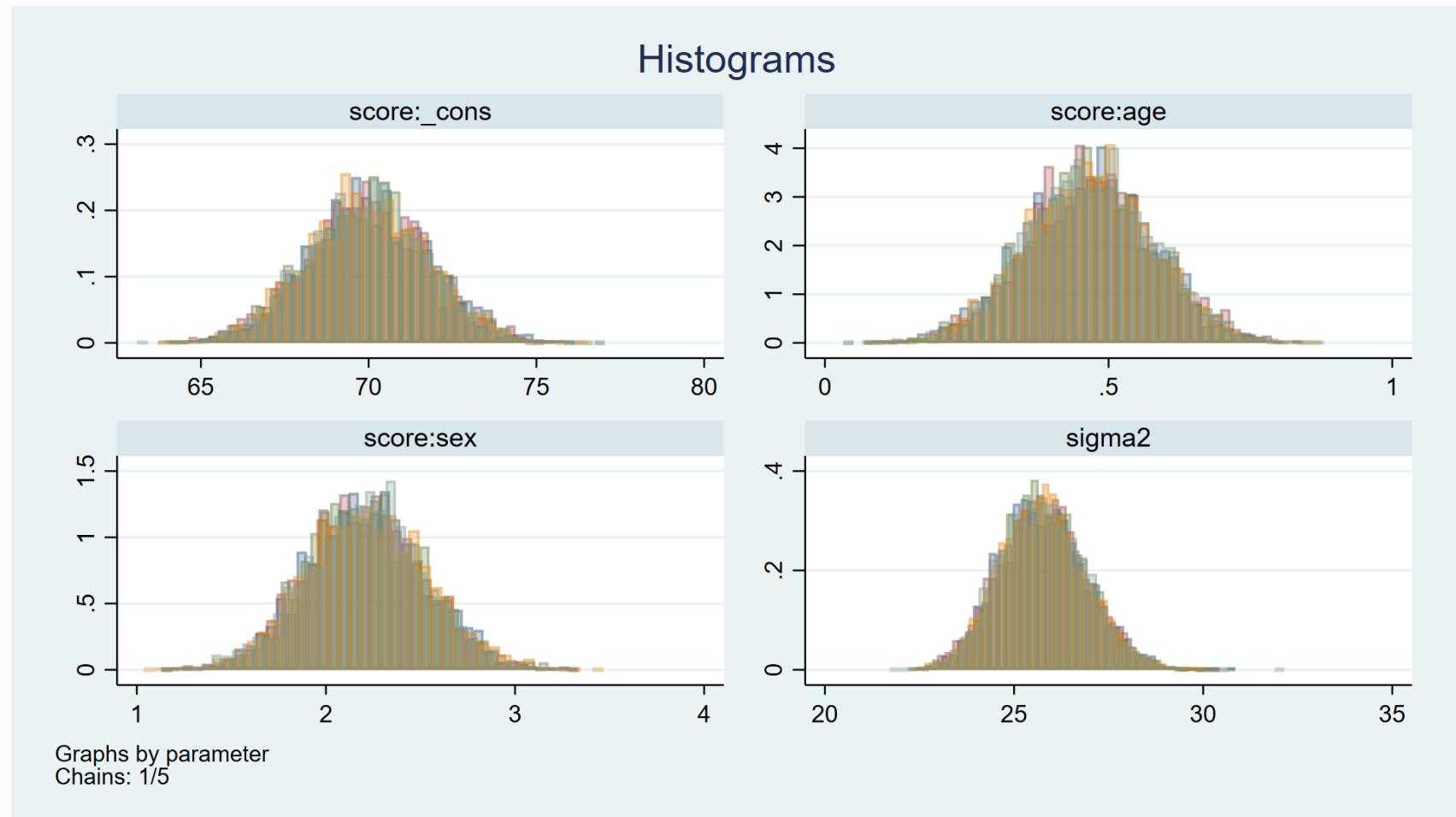
```
bayesgraph trace {score:_cons age sex} {sigma2}, byparm
```

# Checking “Convergence” of the Chain



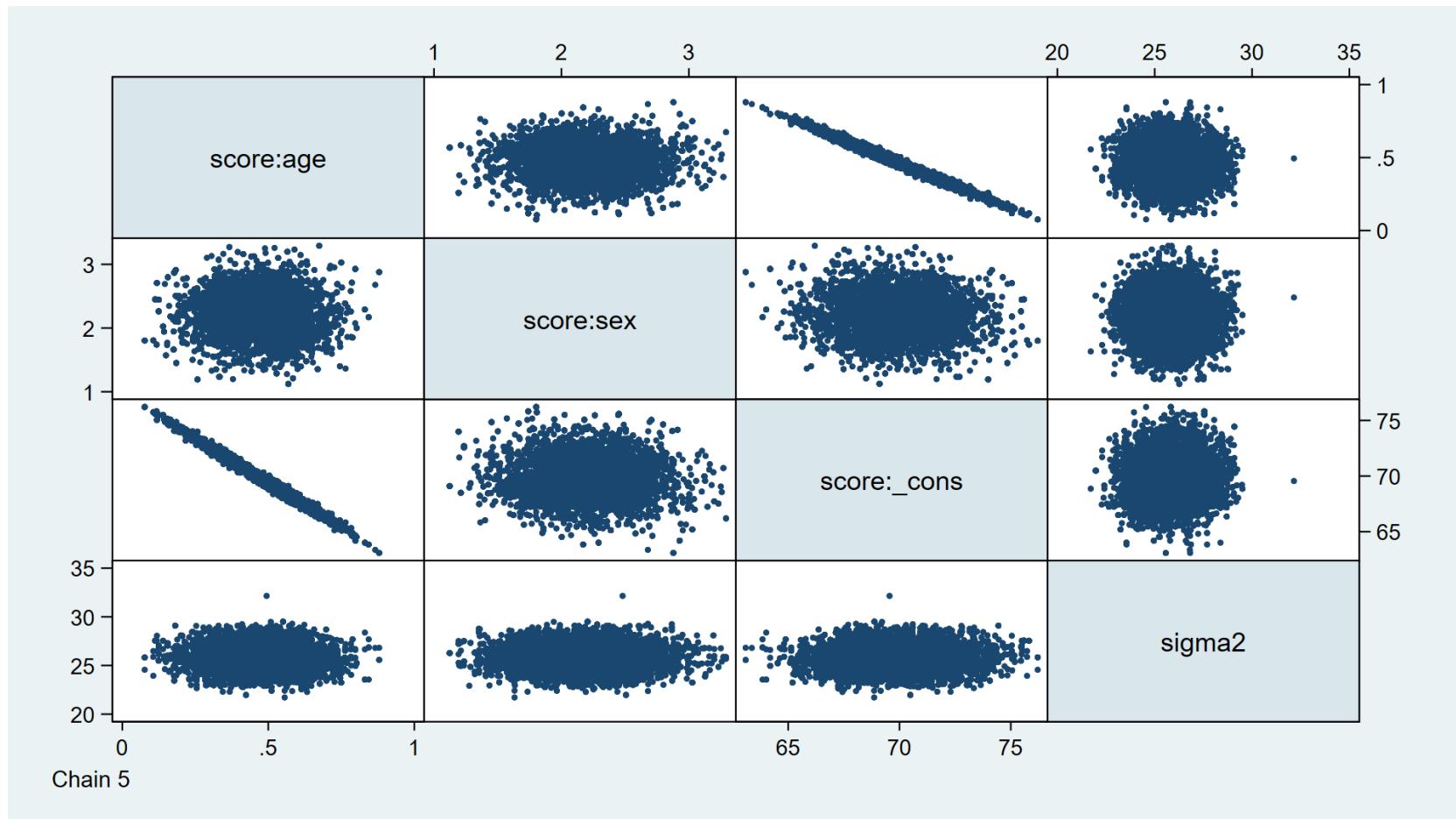
```
bayesgraph ac {score:_cons age sex} {sigma2}, byparm
```

# Checking “Convergence” of the Chain



```
bayesgraph histogram {score:_cons age sex} {sigma2}, byparm
```

# Checking “Convergence” of the Chain



```
bayesgraph matrix _all
```

# Bayesian Model Selection

```
quietly {  
    bayes, rseed(15) saving(age, replace): regress score age  
    estimates store age  
  
    bayes, rseed(15) saving(sex, replace): regress score sex  
    estimates store sex  
  
    bayes, rseed(15) saving(full, replace): regress score age sex  
    estimates store full  
}
```

# Bayesian Model Selection

```
. bayesstats ic age sex full, diconly
```

Deviance information criterion

	DIC
age	6134.169
sex	6104.963
full	6089.611

# Bayesian Model Selection

```
. bayestest model age sex full
```

Bayesian model tests

	log(ML)	P(M)	P(M y)
age	-3.08e+03	0.3333	0.0000
sex	-3.07e+03	0.3333	0.1407
full	-3.07e+03	0.3333	0.8593

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

# Bayesian Model Selection

```
. bayesstats ic age sex full, basemodel(full) bayesfactor
```

Bayesian information criteria

	DIC	log(ML)	BF
age	6134.169	-3084.262	2.67e-08
sex	6104.963	-3068.631	.1637183
full	6089.611	-3066.822	.

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

# Tests

```
. bayestest interval {score:sex}, lower(1.5) upper(2.5)
```

Interval tests      MCMC sample size =      10,000

prob1 : 1.5 < {score:sex} < 2.5

	Mean	Std. Dev.	MCSE
prob1	.8004	0.39972	.0115299

# Predictions

## Frequentist

```
. lincom _b[_cons] + _b[age]*16 + _b[sex]*1
( 1) 16*age + sex + _cons = 0
```

score	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	79.61153	.2281184	348.99	0.000	79.16388 80.05917

## Bayesian

```
. bayesstats summary (predicted: {score:_cons} + {score:age}*16 + {score:sex}*1)
```

Posterior summary statistics    MCMC sample size = 10,000

predicted : {score:\_cons} + {score:age}\*16 + {score:sex}\*1

	Equal-tailed				
	Mean	Std. Dev.	MCSE	Median	[95% Cred. Interval]
predicted	79.62178	.2310121	.007836	79.62428	79.17674 80.06427

# Predictions

```
. bayesstats summary (pred14: {score:_cons} + {score:age}*14 + {score:sex}*1) ///
>                      (pred15: {score:_cons} + {score:age}*15 + {score:sex}*1) ///
>                      (pred16: {score:_cons} + {score:age}*16 + {score:sex}*1) ///
>                      (pred17: {score:_cons} + {score:age}*17 + {score:sex}*1) ///
>                      (pred18: {score:_cons} + {score:age}*18 + {score:sex}*1)
```

Posterior summary statistics

MCMC sample size = 10,000

```
pred14 : {score:_cons} + {score:age}*14 + {score:sex}*1
pred15 : {score:_cons} + {score:age}*15 + {score:sex}*1
pred16 : {score:_cons} + {score:age}*16 + {score:sex}*1
pred17 : {score:_cons} + {score:age}*17 + {score:sex}*1
pred18 : {score:_cons} + {score:age}*18 + {score:sex}*1
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred. Interval]	
pred14	78.69706	.3194793	.010116	78.70842	78.06708	79.31235
pred15	79.15942	.2567973	.00833	79.16805	78.64546	79.65786
pred16	79.62178	.2310121	.007836	79.62428	79.17674	80.06427
pred17	80.08414	.2536382	.008898	80.08001	79.5942	80.57721
pred18	80.5465	.3143914	.01207	80.54443	79.9303	81.16742

```
. bayesstats summary (pred14_m: {score:_cons} + {score:age}*14 + {score:sex}*1) ///
>                                (pred15_m: {score:_cons} + {score:age}*15 + {score:sex}*1) ///
>                                (pred16_m: {score:_cons} + {score:age}*16 + {score:sex}*1) ///
>                                (pred17_m: {score:_cons} + {score:age}*17 + {score:sex}*1) ///
>                                (pred18_m: {score:_cons} + {score:age}*18 + {score:sex}*1) ///
>                                (pred14_f: {score:_cons} + {score:age}*14 + {score:sex}*0) ///
>                                (pred15_f: {score:_cons} + {score:age}*15 + {score:sex}*0) ///
>                                (pred16_f: {score:_cons} + {score:age}*16 + {score:sex}*0) ///
>                                (pred17_f: {score:_cons} + {score:age}*17 + {score:sex}*0) ///
>                                (pred18_f: {score:_cons} + {score:age}*18 + {score:sex}*0)
```

Posterior summary statistics

MCMC sample size = 10,000

pred14\_m : {score:\_cons} + {score:age}\*14 + {score:sex}\*1  
 pred15\_m : {score:\_cons} + {score:age}\*15 + {score:sex}\*1  
 pred16\_m : {score:\_cons} + {score:age}\*16 + {score:sex}\*1  
 pred17\_m : {score:\_cons} + {score:age}\*17 + {score:sex}\*1  
 pred18\_m : {score:\_cons} + {score:age}\*18 + {score:sex}\*1  
 pred14\_f : {score:\_cons} + {score:age}\*14 + {score:sex}\*0  
 pred15\_f : {score:\_cons} + {score:age}\*15 + {score:sex}\*0  
 pred16\_f : {score:\_cons} + {score:age}\*16 + {score:sex}\*0  
 pred17\_f : {score:\_cons} + {score:age}\*17 + {score:sex}\*0  
 pred18\_f : {score:\_cons} + {score:age}\*18 + {score:sex}\*0

	Equal-tailed					
	Mean	Std. Dev.	MCSE	Median	[95% Cred. Interval]	
pred14_m	78.69706	.3194793	.010116	78.70842	78.06708	79.31235
pred15_m	79.15942	.2567973	.00833	79.16805	78.64546	79.65786
pred16_m	79.62178	.2310121	.007836	79.62428	79.17674	80.06427
pred17_m	80.08414	.2536382	.008898	80.08001	79.5942	80.57721
pred18_m	80.5465	.3143914	.01207	80.54443	79.9303	81.16742
pred14_f	76.47326	.3095034	.012827	76.48103	75.86406	77.05269
pred15_f	76.93562	.2479293	.010275	76.93282	76.44855	77.40938
pred16_f	77.39797	.2251431	.008618	77.39667	76.95665	77.85437
pred17_f	77.86033	.2519004	.008338	77.86333	77.37567	78.3648
pred18_f	78.32269	.3158515	.009502	78.32296	77.70053	78.96945

# Predictions

```
. return list
```

scalars:

```
r(nchains) = 1  
r(mcmc size) = 10000  
r(clevel) = 95  
    r(hpd) = 0  
r(batch) = 0  
r(skip) = 0  
r(corrlag) = 500  
r(corrtol) = .01
```

macros:

```
r(names) : ""pred14_m" "pred15_m" "pred16_m" "pred17_m" "pred18_m" "pred14_f" "pred15_f" "pred16_f" "p.."  
r(exprnames) : ""pred14_m" "pred15_m" "pred16_m" "pred17_m" "pred18_m" "pred14_f" "pred15_f" "pred16_f" "p.."  
r(expr_10) : "{score:_cons} + {score:age}*18 + {score:sex}*0"  
r(expr_9) : "{score:_cons} + {score:age}*17 + {score:sex}*0"  
r(expr_8) : "{score:_cons} + {score:age}*16 + {score:sex}*0"  
r(expr_7) : "{score:_cons} + {score:age}*15 + {score:sex}*0"  
r(expr_6) : "{score:_cons} + {score:age}*14 + {score:sex}*0"  
r(expr_5) : "{score:_cons} + {score:age}*18 + {score:sex}*1"  
r(expr_4) : "{score:_cons} + {score:age}*17 + {score:sex}*1"  
r(expr_3) : "{score:_cons} + {score:age}*16 + {score:sex}*1"  
r(expr_2) : "{score:_cons} + {score:age}*15 + {score:sex}*1"  
r(expr_1) : "{score:_cons} + {score:age}*14 + {score:sex}*1"
```

matrices:

```
r(summary) : 10 x 6
```

# Predictions

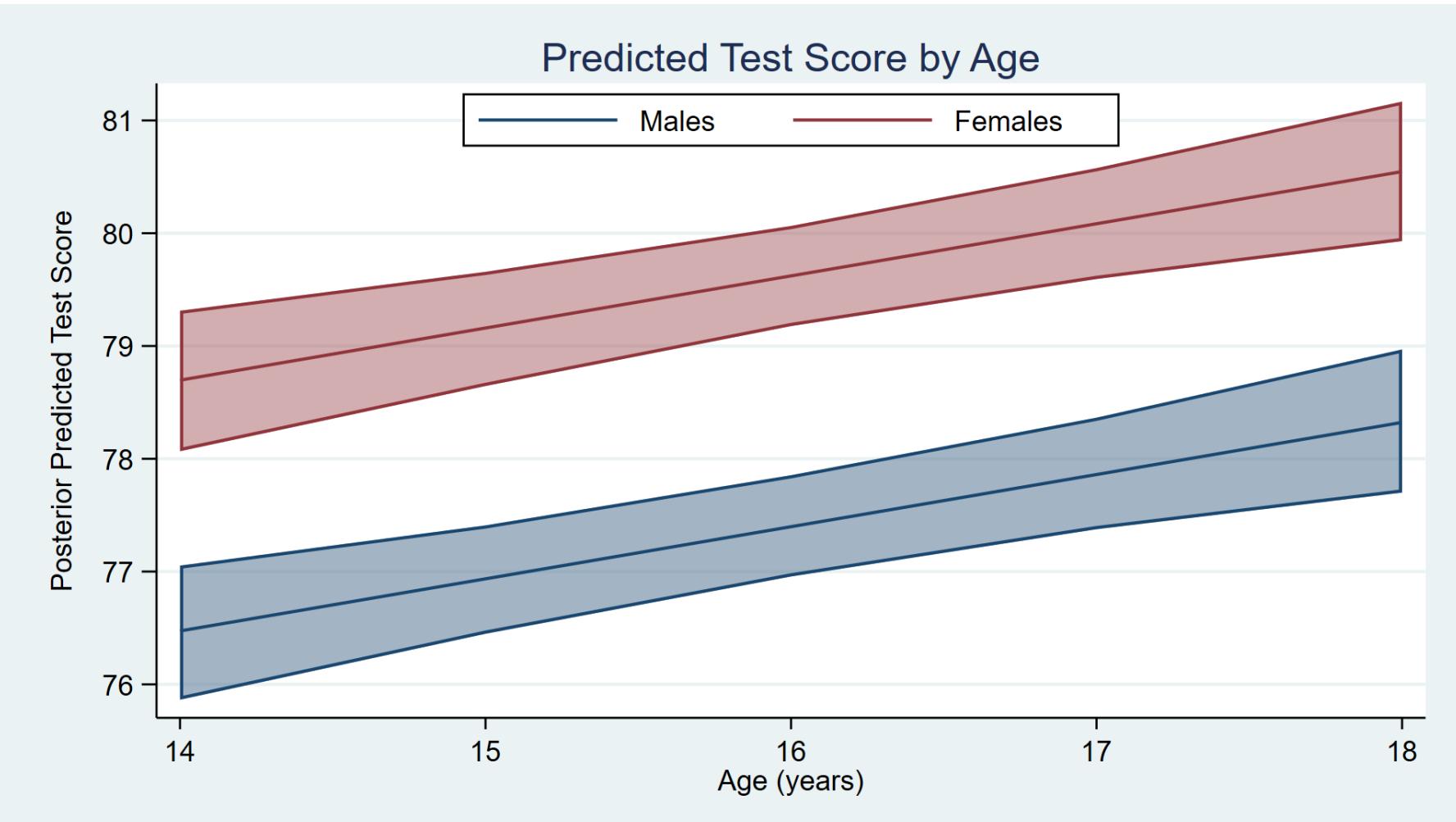
```
. matlist r(summary)
```

	Mean	Std Dev	MCSE	Median	CrI lower	CrI upper
pred14_m	78.69706	.3194793	.0101164	78.70842	78.06708	79.31235
pred15_m	79.15942	.2567973	.0083296	79.16805	78.64546	79.65786
pred16_m	79.62178	.2310121	.0078362	79.62428	79.17674	80.06427
pred17_m	80.08414	.2536382	.008898	80.08001	79.5942	80.57721
pred18_m	80.5465	.3143914	.0120702	80.54443	79.9303	81.16742
pred14_f	76.47326	.3095034	.0128274	76.48103	75.86406	77.05269
pred15_f	76.93562	.2479293	.0102751	76.93282	76.44855	77.40938
pred16_f	77.39797	.2251431	.0086175	77.39667	76.95665	77.85437
pred17_f	77.86033	.2519004	.0083378	77.86333	77.37567	78.3648
pred18_f	78.32269	.3158515	.0095017	78.32296	77.70053	78.96945

# Convert the Matrix to a Dataset

```
matrix pred = r(summary)
clear
svmat pred          /* Convert matrix to a dataset */
rename pred1 mean
rename pred2 stddev
rename pred3 mcse
rename pred4 median
rename pred5 lower
rename pred6 upper
gen sex = _n>5
label define sex 0 "Female" 1 "Male"
label values sex sex
gen age = _n + 13 if _n<6
replace age = _n + 8 if _n>5
```

# Predictions



# The **bayes** Prefix

Linear regression models

<b>regress</b>	<b>[BAYES] bayes:</b> <b>regress</b>	Linear regression
<b>hetregress</b>	<b>[BAYES] bayes:</b> <b>hetregress</b>	Heteroskedastic linear regressions
<b>tobit</b>	<b>[BAYES] bayes:</b> <b>tobit</b>	Tobit regression
<b>intreg</b>	<b>[BAYES] bayes:</b> <b>intreg</b>	Interval regression
<b>truncreg</b>	<b>[BAYES] bayes:</b> <b>truncreg</b>	Truncated regression
<b>mvreg</b>	<b>[BAYES] bayes:</b> <b>mvreg</b>	Multivariate regression

Sample-selection regression models

<b>heckman</b>	<b>[BAYES] bayes:</b> <b>heckman</b>	Heckman selection model
<b>heckprobit</b>	<b>[BAYES] bayes:</b> <b>heckprobit</b>	Probit model with sample selection
<b>heckoprobit</b>	<b>[BAYES] bayes:</b> <b>heckoprobit</b>	Ordered probit model with sample selection

# The **bayes** Prefix

Binary-response regression models

<code>logistic</code>	<code>[BAYES] bayes: logistic</code>
<code>logit</code>	<code>[BAYES] bayes: logit</code>
<code>probit</code>	<code>[BAYES] bayes: probit</code>
<code>cloglog</code>	<code>[BAYES] bayes: cloglog</code>
<code>hetprobit</code>	<code>[BAYES] bayes: hetprobit</code>
<code>binreg</code>	<code>[BAYES] bayes: binreg</code>
<code>biprobit</code>	<code>[BAYES] bayes: biprobit</code>

Logistic regression, reporting odds ratios  
Logistic regression, reporting coefficients  
Probit regression  
Complementary log-log regression  
Heteroskedastic probit regressions  
GLM for the binomial family  
Bivariate probit regression

Ordinal-response regression models

<code>ologit</code>	<code>[BAYES] bayes: ologit</code>
<code>oprobit</code>	<code>[BAYES] bayes: oprobit</code>
<code>zioprobit</code>	<code>[BAYES] bayes: zioprobit</code>

Ordered logistic regression  
Ordered probit regression  
Zero-inflated ordered probit regression

Categorical-response regression models

<code>mlogit</code>	<code>[BAYES] bayes: mlogit</code>
<code>mprobit</code>	<code>[BAYES] bayes: mprobit</code>
<code>clogit</code>	<code>[BAYES] bayes: clogit</code>

Multinomial (polytomous) logistic regression  
Multinomial probit regression  
Conditional logistic regression

# The **bayes** Prefix

Count-response regression models

<b>poisson</b>	<b>[BAYES] bayes:</b> <b>poisson</b>
<b>nbreg</b>	<b>[BAYES] bayes:</b> <b>nbreg</b>
<b>gnbreg</b>	<b>[BAYES] bayes:</b> <b>gnbreg</b>
<b>tpoisson</b>	<b>[BAYES] bayes:</b> <b>tpoisson</b>
<b>tnbreg</b>	<b>[BAYES] bayes:</b> <b>tnbreg</b>
<b>zip</b>	<b>[BAYES] bayes:</b> <b>zip</b>
<b>zinb</b>	<b>[BAYES] bayes:</b> <b>zinb</b>

Poisson regression  
Negative binomial regression  
Generalized negative binomial regression  
Truncated Poisson regression  
Truncated negative binomial regression  
Zero-inflated Poisson regression  
Zero-inflated negative binomial regression

Generalized linear models

<b>glm</b>	<b>[BAYES] bayes:</b> <b>glm</b>
------------	----------------------------------

Generalized linear models

Fractional-response regression models

<b>fracreg</b>	<b>[BAYES] bayes:</b> <b>fracreg</b>
<b>betareg</b>	<b>[BAYES] bayes:</b> <b>betareg</b>

Fractional response regression  
Beta regression

Survival regression models

<b>streg</b>	<b>[BAYES] bayes:</b> <b>streg</b>
--------------	------------------------------------

Parametric survival models

# The **bayes** Prefix

Multilevel regression models

<b>mixed</b>	[BAYES] <b>bayes:</b> <b>mixed</b>
<b>metobit</b>	[BAYES] <b>bayes:</b> <b>metobit</b>
<b>meintreg</b>	[BAYES] <b>bayes:</b> <b>meintreg</b>
<b>melogit</b>	[BAYES] <b>bayes:</b> <b>melogit</b>
<b>meprobit</b>	[BAYES] <b>bayes:</b> <b>meprobit</b>
<b>mecloglog</b>	[BAYES] <b>bayes:</b> <b>mecloglog</b>
<b>meologit</b>	[BAYES] <b>bayes:</b> <b>meologit</b>
<b>meoprobit</b>	[BAYES] <b>bayes:</b> <b>meoprobit</b>
<b>mepoisson</b>	[BAYES] <b>bayes:</b> <b>mepoisson</b>
<b>menbreg</b>	[BAYES] <b>bayes:</b> <b>menbreg</b>
<b>meglm</b>	[BAYES] <b>bayes:</b> <b>meglm</b>
<b>mestreg</b>	[BAYES] <b>bayes:</b> <b>mestreg</b>

Multilevel linear regression
Multilevel tobit regression
Multilevel interval regression
Multilevel logistic regression
Multilevel probit regression
Multilevel complementary log-log regression
Multilevel ordered logistic regression
Multilevel ordered probit regression
Multilevel Poisson regression
Multilevel negative binomial regression
Multilevel generalized linear model
Multilevel parametric survival regression

# Outline

- Introduction to Bayesian Analysis
- Coin Toss Example
- Priors, Likelihoods, and Posteriors
- Markov Chain Monte Carlo (MCMC)
- Bayesian Linear Regression
- **Advantages and Disadvantages of Bayes**

# Advantages of Bayesian Statistics

- Formally incorporate prior information into studies
- Works when maximum likelihood estimation (MLE) fails or is not identified
- Does not rely on asymptotic normality like MLE
- Works with small sample sizes
- Intuitive interpretation of results such as credible intervals

# US Food and Drug Administration (FDA)

## Guidance for Industry and FDA Staff

### Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Document issued on: February 5, 2010

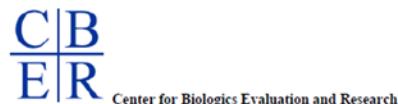
The draft of this document was issued on 5/23/2006

For questions regarding this document, contact Dr. Greg Campbell (CDRH) at 301-796-5750 or [greg.campbell@fda.hhs.gov](mailto:greg.campbell@fda.hhs.gov) or the Office of Communication, Outreach and Development, (CBER) at 1-800-835-4709 or 301-827-1800.



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Devices and Radiological Health

Division of Biostatistics  
Office of Surveillance and Biometrics



#### Incorporating informative prior distributions

A Bayesian analysis of a current study of a new device may include prior information from:

- the new device,
- the control device, or
- both devices.

When incorporating prior information from a previous study, the patients in the previous study are rarely considered exchangeable with the patients in the current study. Instead, a hierarchical model is often used to “borrow strength” from the previous studies. At the first level of the hierarchy, these models assume that patients are exchangeable within a study but not across studies. At a second level of the hierarchy, the previous studies are assumed to be exchangeable with the current study, which acknowledges variation between studies. For more detail on hierarchical models, see Section 4.6.

Quote from page 22

# Disadvantages of Bayesian Statistics

- Subjectivity in the selection of prior distributions
- Computational complexity

# Outline

- Introduction to Bayesian Analysis
- Coin Toss Example
- Priors, Likelihoods, and Posteriors
- Markov Chain Monte Carlo (MCMC)
- Bayesian Linear Regression
- Advantages and Disadvantages of Bayes

# Thank you!

# Questions?

You can download the slides, datasets, and do-files here:

<https://tinyurl.com/IntroToBayes>

You can contact me anytime at [chuber@stata.com](mailto:chuber@stata.com)