马尔科夫链蒙特卡洛模拟与 Stata 应用

王群勇 (南开大学数量经济研究所, QunyongWang@outlook.com)

Contents

蒙特卡洛模拟

计量经济学中很多模型的估计存在积分问题。比如随机效应面板模型，

$$y_{it} = x_{it}\beta + c_i + u_{it}, u_{it} \sim N(0, \sigma^2)$$

$y_i = \left(y_{i1}, \ldots, y_{i,T_i}\right)$的似然函数为

$$f(y_i|x_{it}, c_i) = \prod_{t=1}^{T_i} f\left(y_{it}|x_{it}, c_i\right)$$

$$f(y_i|x_{it}) = \int f(y_i|x_{it}, c_i)g(c_i)dc_i = E\big(f(y_i|x_{it}, c_i)\big) \leftarrow S^{-1}\sum_{s=1}^{S} f^{(s)}\left(y_i|x_{it}, c_i\right)).$$

蒙特卡洛模拟

在所有随机效应面板模型中都存在相同的问题。比如随机效应 Probit 模型，

$$P(y_{it}|x_i, c_i) = \Phi(x_{it}\beta + c_i)^{y_{it}}\big(1 - \Phi(x_{it}\beta + c_i)\big)^{1-y_{it}}$$

$y_i = \left(y_{i1}, \dots, y_{i,T_i}\right)$的似然函数为

$$f(y_i | x_{it}, c_i) = \prod_{t=1}^{T_i} P\left(y_{it} | x_{it}, c_i\right)$$

$$f(y_i | x_{it}) = \int f(y_i | x_{it}, c_i) g(c_i) dc_i = E\big(f(y_i | x_{it}, c_i)\big) \leftarrow S^{-1} \sum_{s=1}^{S} f^{(s)}\left(y_i | x_{it}, c_i\right)).$$

蒙特卡洛积分

蒙特卡洛积分：设$x \sim g(x)$，求$E\big(h(x)\big) = \int_{-\infty}^{\infty} h\,(x) g(x) dx.$

根据大数定律，$(x_1, x_2, \dots, x_n), \bar{x} \to E(x).$

从分布$g(x)$抽取随机数$(x_1, x_2, \dots, x_S)$,

$$\bar{h}(x) = S^{-1} \sum_{i=1}^{S} h\,(x_i) \to E\big(h(x)\big).$$

例：$\int_0^1 [\cos(50x) + \sin(20x)]^2\, dx.$

令$x \sim Uniform(0,1)$

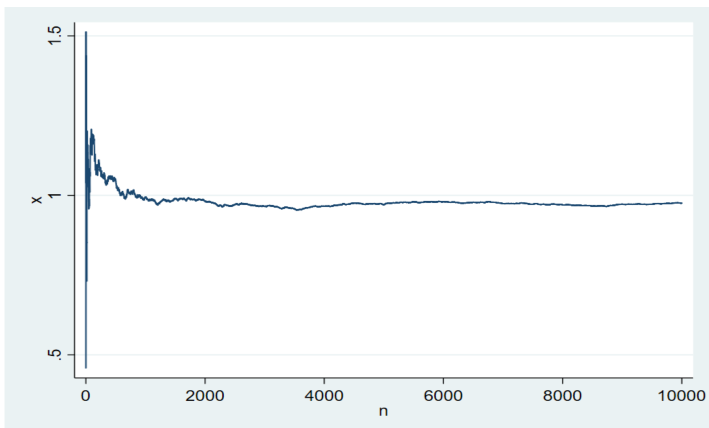
$$\int_0^1 [\cos(50x) + \sin(20x)]^2\, dx = E([\cos(50x) + \sin(20x)]^2).$$

蒙特卡洛积分

```
. set obs 10000
Number of observations (_N) was 114, now 10,000.

. qui gen x = .

. gen n=_n

. mata
------------------------------------------------- mata (type end to exit) -------
---------------------------------------

: n = 10000
```

```
: u = runiform(n,1)

: rmean= runningsum((cos(50*u) + sin(20*u)):^2):/range(1,n,1)

: st_store(., "x", rmean)

: end
```

```
. line x n
```



参数估计量的模拟标准差

模型：$\ln(y_t) = \rho\ln(y_{t-1}) + \ln(x_t)\beta + u_t$

长期弹性：$\eta = \beta/(1-\rho) \to \hat{\beta}/(1-\hat{\rho})$.

如何检验长期弹性的显著性或者置信区间：

- Delta 方法
- 抽取(β,ρ)的S个随机数，直接计算$\eta_s(s=1,\cdots,S)$。

传统的模拟方法

- 逆概率转换
- 合成法
- 接受拒绝法

- 重要性抽样法

Contents

MCMC 与贝叶斯分析

贝叶斯定理：

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

对于计量模型，数据为y，参数为θ。贝叶斯定理表示为

$$\pi(\theta|y)=\frac{f(y|\theta)\pi(\theta)}{f(y)}\propto f(y|\theta)\pi(\theta).$$

即，后验分布 ∝ 似然函数 × 先验分布。 $f(y|\theta)\pi(\theta)$叫做贝叶斯核(kernel)。 其中，$f(y|\theta)$为似然函数(概率密度函数)，$\pi(\theta)$为先验分布，$\pi(\theta|y)$为后验分布。

马尔科夫链

传统方法的局限：变量的分布必须具有明确的表达式。

随机过程X_t取不同数值$s=(1,2,\cdots,S)$，转移概率为

$$p_{ij}=P(X_{t+1}=j|X_t=i),i,j\in S.$$

比如，

$$P=\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}=\begin{bmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{bmatrix}.$$

马尔可夫特征：随机变量的t期状态只取决于$t-1$期状态。即下一步去哪里只取决于变量当前所在的位置，而不取决于是怎么到达当前位置的。

马尔科夫链：满足马尔科夫特征的随机过程。

马尔科夫链

如果$\pi_t=(\pi_1,\cdots,\pi_s)$，那么$\pi_{t+1}=\pi_t P$.

$$[\pi_{1,t+1} \quad \pi_{2,t+1}]=[\pi_{1t} \quad \pi_{2t}]\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}$$

$\pi_{t+n}=\pi_{t+n-1}P=\pi_{t+n-2}P^2=\pi_t P^n$.

平稳分布：如果分布π_t不再变化，即$\pi=\pi P$，π称为平稳分布。

$\pi(I-P)=0$，所以π为 P 矩阵特征值 1 对应的特征向量。

例：转移矩阵为

$$P=\begin{bmatrix}0.75 & 0.25\\ 0.125 & 0.875\end{bmatrix}$$

可以求出其对应的平稳分布为(1/3, 2/3).

马尔科夫链

例，从状态$p=(0.1,0.9)$开始，经过 50 次转移，那么分布收敛到(1/3,2/3)。

```
. mata:
------------------------------------------------- mata (type end to exit) -----
-----------------------------------------
: K=(0.75,0.25  \ 0.125, 0.875)

: p=(0.1, 0.9)

: for (i=1; i<=50; i++) {
>  p = p*K
> }

: p
                  1              2
    +-------------------------------+
  1 |  .3333333333    .6666666667   |
    +-------------------------------+
```

```
: p*K
                 1              2
    +-----------------------------+
  1 |  .3333333333   .6666666667  |
    +-----------------------------+

: end
```

马尔科夫链

从状态$p = (0.6,0.4)$开始:

```
. mata:
------------------------------------------------- mata (type end to exit) ------
: K=(0.75,0.25  \ 0.125, 0.875)

: p=(0.6, 0.4)

: for (i=1; i<=50; i++) {
>  p = p*K
> }

: p
                 1              2
    +-----------------------------+
  1 |  .3333333333   .6666666667  |
    +-----------------------------+

: p*K
                 1              2
    +-----------------------------+
  1 |  .3333333333   .6666666667  |
    +-----------------------------+

: end
```

马尔科夫链

对于一个转移矩阵，平稳分布可能不是唯一的。比如

$$P=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

可能存在无穷多的平稳分布。比如，转移矩阵为

$$\begin{bmatrix}1/3 & 2/3 & 0\\1/3 & 2/3 & 0\\0 & 0 & 1\end{bmatrix}$$

马尔科夫链

令$P^{(n)}=\pi_t P^n$，如果$P_{ij}^{(n)}>0$，称j从i可达（accessible）。

如果$P_{ij}^{(n)}>0$，$P_{ji}^{(n)}>0$，称i与j互通(communicate)。一组互通的状态构成互通类(communicate class)。

只包含一个互通类的马尔科夫链称为不可约的(irreducible)。

定理：不可约的马尔科夫链具有唯一的平稳分布。

含义：对于目标分布，如果能找到其对应的马尔科夫链，那么总可以到达该分布。

Contents

MH 抽样

如何找到马尔科夫链？MCMC: Gibbs 抽样(Geman and Geman, 1984), Metropolis-Hastings 抽样(Metropolis et al., 1953; Hastings, 1970)

数值例子：$y=\sum y_i \sim binomial(10,\theta)$, prior: $\theta \sim Beta(1,1)$

设初始值$\theta_1 = 0.517$, 令$\theta_{new} = 0.380$?

$$\rho = \frac{\text{posterior}(\theta_{new})}{\text{posterior}_{\theta_1}} = \frac{Beta(1,1,0.380) \times Binomial(10,4,0.380)}{Beta(1,1,0.517) \times Binomial(10,4,0.517)} = 1.307$$

$\theta_2 = 0.380$, 令$\theta_{new} = 0.286$?

$$\rho = \frac{\text{posterior}(\theta_{new})}{\text{posterior}_{\theta_2}} = \frac{Beta(1,1,0.286) \times Binomial(10,4,0.286)}{Beta(1,1,0.380) \times Binomial(10,4,0.380)} = 0.747$$

θ_3以 0.747 的概率取 0.286，以 0.253 的概率取 0.380.

……

？如何生成θ_{new}： 工具分布(proposal distribution).

MCMC

MH 抽样: 设后验分布为$f(\theta)$（θ代表模型的参数）

(1) 给定θ_t, 定义工具分布$q(y|\theta_t)$ (proposed distribution), 从$q(y|\theta_t)$抽取随机数θ_{new}。其中，工具分布：(a)容易抽样; (b)定义域覆盖后验分布的定义域。

(2) 定义接受概率(acceptance probability): $\rho(\theta_t, \theta_{new}) = \min\left[\frac{f(\theta_{new})}{f(\theta_t)} \frac{q(\theta_t|\theta_{new})}{q(\theta_{new}|\theta_t)}, 1\right]$,

$$\theta_{t+1} = \begin{cases} \theta_{new} & \rho(\theta_t, \theta_{new}) \\ \theta_t & 1 - \rho(\theta_t, \theta_{new}) \end{cases}$$

对称情形：$q(\theta_t|\theta_{new}) = q(\theta_{new}|\theta_t), \rho(\theta_t, \theta_{new}) = \min\left[\frac{f(\theta_{new})}{f(\theta_t)}, 1\right]$,

独立情形：$q(\theta_t|\theta_{new}) = q(\theta_t), q(\theta_{new}|\theta_t) = q(\theta_{new}), \rho(\theta_t, \theta_{new}) = \min\left[\frac{f(\theta_{new})}{f(\theta_t)} \frac{q(\theta_t)}{q(\theta_{new})}, 1\right]$。

1. 如果$q(\theta_t) = q(\theta_{new})$，那么$\rho = \min[f(\theta_{new})/f(\theta_t), 1]$.
2. 计算接受概率时，分布中只有核起作用，其它常数都被略掉。

独立 MH 抽样

1. 与 Accept-Reject 模拟抽样不同， MH 算法中每次迭代都有一个取值，或者新的实验值，或者是上一期的模拟值。

2. MH 模拟样本存在比较高的自相关，自相关滞后阶数可能长达几十，导致有效样本下降。

3. 有效的 MH 抽样需要将接受率控制在一个良好的范围内。

独立 MH 抽样

抽样效率：接受率。接受率太低意味着很多无效的抽样，导致模拟样本存在过高的自相关。但接受率也不是越高越好，过高的接受率可能意味着每次的变化太小，MH 抽样可能陷入一个局部的空间，而无法覆盖完整的定义域。合理的接受率在[0.15, 0.50]之间。

？如何提高抽样效率？(a) 随机游走 MH 抽样；(b) 适应性 MH 抽样; (c) 分块抽样; (d) Gibbs 抽样; ……

MH 算法弹性很大，允许很多不同的算法。不同算法的效率是不一样的。

适应性随机游走 MH 抽样

适应性 MH 抽样是指每隔一定的区间调整ρ_t和Σ_t，以实现最优的接受率(TAR)。

$$\theta_{new} = \theta_{t-1} + e_t, e_t \sim N(0, \rho_t^2 \Sigma_t)$$

根据 Gelman, Gilks, and Roberts (1997)，单个参数的最优接受率为 0.44，多个参数的最优接受率为 0.234。 (ρ_k^2, Σ_k)的更新方程为

$$\begin{aligned} \rho_k &= \rho_{k-1} \exp\left[\beta_k\left(\Phi^{-1}(AR_k/2) - \Phi^{-1}(TAR/2)\right)\right]. \\ \Sigma_k &= (1-\beta_k)\Sigma_{k-1} + \beta_k \hat{\Sigma}_k. \end{aligned}$$

其中，$AR_k = (1-\alpha)AR_{k-1} + \alpha \widehat{AR}_k$.

根据AR_k的公式，如果当前的接受率超过最优水平，ρ_k会提高，即工具密度的方差会增加，这会降低接受率。根据 Gelman, Gilks, and Roberts (1997), Roberts and Rosenthal (2001), and Andrieu and Thoms (2008)，初始值$_0 设定为$2.38/\sqrt{d}$，其中，$d$表示参数的个数。$\widehat{AR}_k$表示第 k 段的接受率，$\hat{\Sigma}_k$表示第 k 段中模拟数值的协方差。$\alpha \in (0,1)$决定了 AR 的平滑水平，$\beta_k = \beta_0/k^{\gamma}$，$\beta_0 \in (0,1)$，$\gamma \in [0,1]$表示降低适应性调整的水平。可以参考的经验设定值，$\alpha = 0.75, \beta_0 = 0.8, \gamma = 0$。

可以设定适应性调整的区间长度、适应性调整的最小或最大次数、停止规则。比如，alen=100；最小调整次数为 5，最大调整次数为 max(25, burnin/alen)；当接受率 AR 与 TAR 接近时则停止调整，等等。

为了防止方差矩阵的退化，Roberts and Rosenthal (2009)建议用固定的方差矩阵，

$$\Sigma_k = (1-\beta_k)\Sigma_{k-1} + \beta_k \Sigma_{fixed}.$$

Stata: bayesmh

bayesmh 可以用于单方程线性模型、单方程非线性模型、多方程线性模型、多方程非线性模型、多水平模型、概率分布等。可以自己设定似然函数或后验分布。

贝叶斯预测(bayespredict)只能用于 bayesmh，而不能用于 bayes。

bayesmh *depvar* [*indepvarspec*] [*if*] [*in*] [*weight*], `likelihood`(*modelspec*) `prior`(*priorspec*) `llevaluator`(*log-likelihood*) `evaluator`(*log-postdist*) [*options*]

三种用法：

- **likelihood() + prior()**: 内置似然函数 + 内置先验分布
- **llevaluator() + prior()**: 自定义的对数似然函数 + 内置先验分布
- **evaluator()**: 自定义的对数后验分布

Stata: bayesmh

似然函数即概率密度函数，不同分布涉及到的参数也不同。比如，正态分布有均值和方差两个参数；泊松分布只有一个参数，既为均值也是方差。

bayesmh 中，分布的均值参数通过模型来设定。分布的形式和其它参数则通过 likelihood()来设定。比如，

- **bayesmh y x1 x2, likelihood(normal({var}))**，表示$y \sim N(b_0 + b_1x_1 + b_2x_2, var)$
- **bayesmh y x1 x2, likelihood(poisson)**，表示$y \sim Poisson\big(\exp(b_0 + b_1x_1 + b_2x_2)\big)$
- **bayesmh y x1 x2, likelihood(probit)**，表示$y \sim Bernoulli\big(\Phi(b_0 + b_1x_1 + b_2x_2)\big)$

Stata: bayesmh

```
. use bintrial, clear

. bayesmh y, likelihood(dbernoulli({p})) prior({p},beta(1,1))

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  y ~ bernoulli({p})
```

```
Prior:
  {p} ~ beta(1,1)
------------------------------------------------------------------------------

Bayesian Bernoulli model                         MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =         10
                                                 Acceptance rate  =      .4517
Log marginal-likelihood = -7.8073834             Efficiency       =      .2739

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
           p |  .4235852   .1358498   .002596   .4210698   .1723984   .6930029
------------------------------------------------------------------------------
```

Contents

MCMC 诊断

接受率(Acceptance Rate, AR): 试验值被接受的概率。Roberts and rosenthal (2001)认为，有效抽样的接受率在 0.15-0.5 之间。

收敛性(是否达到平稳分布)

- 踪迹图（trace plot）： 围绕常数均值波动

- 核密度图（kernel density）：模拟样本分为前后两段，比较其分布。
- 自相关
- Gelman-Rubin (shrink factor)统计量：多条马尔科夫链进行抽样，那么链之间的平均差异与链内的平均差异应该进行相等。二者的比例称作收缩因子（shrink factor），作为经验规则， 收缩因子应低于 1.1。

有效性（Efficiency）是衡量自相关的一个指标。大于 0.1，视作良好。低于 0.01，严重关注。

MCMC 诊断

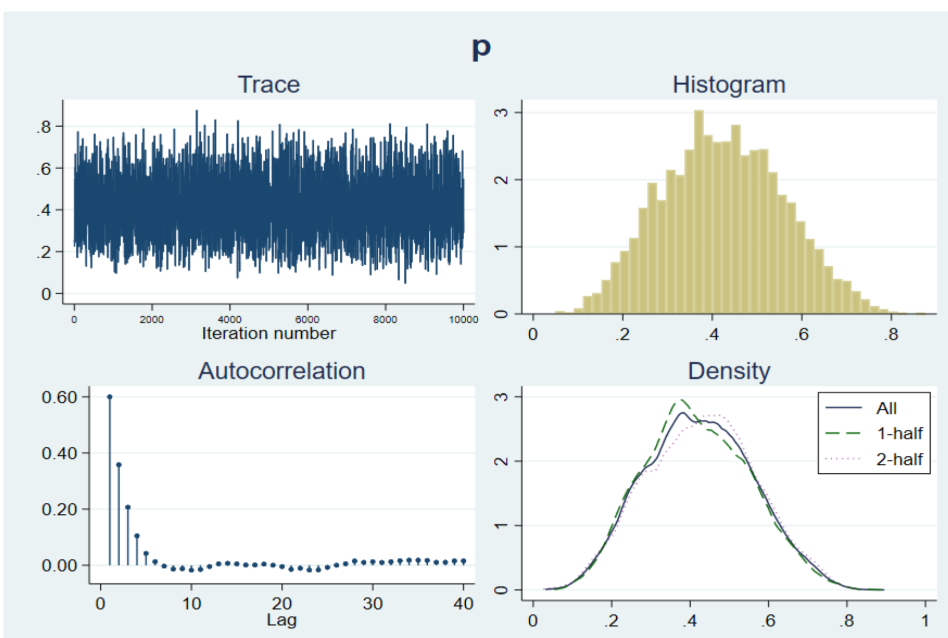
bayesgraph *type which*

其中，*type* 包括:`diag`, `trace`, `histogram`, `kdensity`, `ac`, `cusum`, `matrix`.

which 包括： **_all**, {[*eqname*:]*param*}

MCMC 诊断

```
. bayesgraph diag {p}
```



MC 抽样结果统计

设 MCMC 模拟得到样本量为T，样本均值和标准差为

$$\bar{\theta} = T^{-1}\sum_{t=1}^{T}\theta_t\,; s^2 = (T-1)^{-1}\sum_{t=1}^{T}(\theta_t - \bar{\theta})^2.$$

如果 MCMC 多条链，那么(Gelman et al. 2014)

$$\bar{\theta} = \frac{1}{MT}\sum_{j=1}^{M}\sum_{t=1}^{T}\theta_{jt}.$$

$$s^2 = \frac{T-1}{T}W + \frac{1}{T}B; B = \frac{T}{M-1}\sum_{j=1}^{M}\left(\bar{\theta}_j - \bar{\theta}\right)^2; W = \frac{1}{M}\sum_{j=1}^{M}s_j^2.$$

MC 抽样结果统计

由于θ_t存在自相关，后验均值的标准差（也叫做蒙特卡洛标准误差，MCSE）不能通过$s/\sqrt{T}$来估计。两种解决方法：有效样本水平(ESS)和块均值（batch mean）。

$$ESS = \frac{T}{1 + 2\sum_{k=1}^{K}\rho_k}; Efficiency = ESS/T; \text{Correlation time} = T/ESS.$$

$$\rho_k = \frac{\gamma_k}{\gamma_0};\ \ \gamma_k = \frac{1}{T}\sum_{t=1}^{T-k}(\theta_t - \bar{\theta})\,(\theta_{t+k} - \bar{\theta}); \gamma_0 = \frac{1}{T}\sum_{t=1}^{T-k}(\theta_t - \bar{\theta})^2.$$

其中，ρ_k为k阶自相关系数；γ_k为自协方差系数。

最高阶数K可以自行设定（`corrlag()`，Stata 默认值为 min(500,T/2)。）或者根据自相关系数高于某个阈值（`corrtol()`，默认值为 0.01）来判断。

$$MCSE(\bar{\theta}) = \frac{s}{\sqrt{ESS}};\ \text{or}\ MCSE(\bar{\theta}) = \frac{s}{\sqrt{\sum_{j=1}^{M}ESS_j}}.$$

MC 抽样结果统计

有效样本：

```
. use bintrial, clear

. qui bayesmh y, likelihood(dbernoulli({p})) prior({p},beta(1,1))

. bayesstats ess

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

	ESS	Corr. time	Efficiency
p	2298.71	4.35	0.2299

MC 抽样结果统计

```
. bayesstats summary _all, corrlag(100)

Posterior summary statistics                      MCMC sample size =    10,000
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
p	.4189065	.1381554	.002882	.4111325	.168179	.6951997

MC 抽样结果统计

块均值（Jones et al. 2006）：将$(\theta_1,\cdots,\theta_T)$分为 m 块，每块的均值为$\bar{\theta}_j$。块均值（batch mean）为

$$\bar{\theta} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_j.$$

$$s_{batch}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\theta})^2.$$

$$MCSE_{batch}(\bar{\theta}) = s_{batch}/\sqrt{m}.$$

Flegal and Jones(2010)证明，最优的块长度为$O\left(T^{1/3}\right)$。

MC 抽样结果统计

```
. bayesstats summary _all, batch(20)

Posterior summary statistics                      MCMC sample size =     10,000
                                                  Batch size       =         20
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
p	.4189065	.1381554	.002633	.4111325	.168179	.6951997

```
Note: Mean and MCSE are estimated using batch means.
```

Contents

分块 MH 抽样

比如，线性回归模型：

$$y = x\beta + u, u \sim (0, \sigma^2), \sigma^2 \sim \text{Inv-Gamma}(a, b)$$

分块 MH 抽样是将变量分为若干子集。不同组的参数的先验分布是独立的，为不同组设计独立的不同的工具分布，每个子集的变量联合抽样。

分块 MH 抽样

将参数分为两组，(x, y)，每组变量的工具密度为$q_x(x|y), q_y(y|x)$。

设置初始值(x_0, y_0):

- 抽取替代值$x_{new} \sim q_x(x|y)$，接受概率为

$$\rho_x = \frac{f(x_{new}|y_0)}{f(x_0|y_0)} \times \frac{q_x\big(x_0|(x_{new}, y_0)\big)}{q_x\big(x_{new}|(x_0, y_0)\big)}$$

- 抽取替代值$y_{new} \sim q_y(y|x)$，接受概率为

$$\rho_y = \frac{f(y_{new}|x_1)}{f(y_0|x_1)} \times \frac{q_y\big(y_0|(y_{new}, x_1)\big)}{q_y\big(y_{new}|(y_0, x_1)\big)}$$

分块 MH 抽样

在回归模型中，Stata 自动将回归系数和方差参数分组。

```
. use taylor, clear

. bayes,  saving(mhtaylor, replace) : regress ffr L.ffr L.pi L.ygap

Burn-in ...
Simulation ...

file mhtaylor.dta saved.

Model summary
------------------------------------------------------------------------------
Likelihood:
  ffr ~ regress(xb_ffr,{sigma2})

Priors:
  {ffr:L.ffr L.pi L.ygap _cons} ~ normal(0,10000)                          (1)
                       {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_ffr.

Bayesian linear regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
```

```
                                                  Number of obs     =        113
                                                  Acceptance rate   =      .3335
                                                  Efficiency:  min  =     .04356
                                                               avg  =     .08301
Log marginal-likelihood = -186.89491                           max  =      .1671
```

	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred.	interval]
ffr						
ffr L1.	.8064498	.0434149	.001676	.8048175	.7167563	.8942407
pi L1.	.3976928	.087098	.00335	.3980353	.223583	.5684021
ygap L1.	.2578887	.0759919	.003641	.258645	.107127	.4120811
_cons	-.0616117	.200531	.007595	-.0623522	-.4526737	.3494749
sigma2	.9037473	.1259337	.003081	.8893884	.6929364	1.179869

```
Note: Default priors are used for model parameters.

. est store mod1
```

注：在用 est store 保存估计结果之前，bayes 必须用 saving()保存模拟的结果。

Gibbs 抽样

如果工具密度与条件分布完全相同，即

$$q_m(t|x_1,\cdots,x_M)=f(x_m=t|x_1,\cdots,x_{m-1},x_{m+1},\cdots,x_M)$$

那么接受概率为

$$\rho_m=\frac{f\left(x_{m,new}|x_1,\cdots,x_{m-1},x_{m+1},\cdots,x_M\right)}{f\left(x_{m,t}|x_1,\cdots,x_{m-1},x_{m+1},\cdots,x_M\right)}\times\frac{f\left(x_{m,t}|x_1,\cdots,x_{m-1},x_{m+1},\cdots,x_M\right)}{f\left(x_{m,new}|x_1,\cdots,x_{m-1},x_{m+1},\cdots,x_M\right)}=1$$

称之为 Gibbs 抽样。Gibbs 抽样是对每个参数的条件分布进行抽样。

Gibbs 抽样的优势：100%的接受率。

劣势：需要得到条件分布的明确的解析式。因此，只是在某些特殊情况下可以使用。

Gibbs 抽样

将随机变量分为两组，(x_t, y_t)，联合密度为$f(x,y)$。

- 抽取x_0
- $t = 1,2,\cdots$，抽取$y_{t+1} \sim f(y|x_t), x_{t+1} \sim f(x|y_{t+1})$

例：二元正态分布。

抽取$x_0 = 0$。令$t = 1,2,\cdots$，抽取

$$\begin{aligned} y_{t+1}|x_t &\sim N\left(\frac{\sigma_1}{\sigma_2}\rho x_t, (1-\rho^2)\sigma_1^2\right) \\ x_{t+1}|y_{t+1} &\sim N\left(\frac{\sigma_2}{\sigma_1}\rho y_{t+1}, (1-\rho^2)\sigma_2^2\right) \end{aligned}$$

可以直接扩展到随机变量分为多组的情况：Gibbs 依次从$f(x_m|x_{-m})$ $(m = 1,2,\cdots,M)$进行抽样，其中x_{-m}表示除了x_m之外所有其它变量。

Gibbs 抽样: 线性回归

模型: $y = x\beta + u$

似然函数: $y_i \sim Normal(x_i\beta, \sigma^2)$

先验分布:

$$\begin{aligned} \beta &\sim \text{Mumltivariate-Normal}(\beta_0, V_0) \\ \sigma^2 &\sim \text{Inverse-Gamma}(\alpha_0/2, \delta_0/2) \end{aligned}$$

边际后验分布:

$$\beta|\sigma^2, y \sim \text{Mumltivariate-Normal}(\beta_1, V_1)$$

$$V_1 = [\sigma^{-2}X'X + V_0^{-1}]^{-1}, \beta_1 = V_1[\sigma^{-2}X'y + V_0^{-1}\beta_0]$$

$$\sigma^2|\beta, y \sim \text{Inverse-Gamma}\big((\alpha_0+n)/2, [\delta_0 + (y-x\beta)'(y-x\beta)]/2\big).$$

Gibbs 抽样: 线性回归

1. 选择$\sigma^{2(0)}$的初始值.
2. 在第t次迭代,

$$
\begin{aligned}
\beta^{(t)}|\sigma^2, y &\sim \text{MNormal}(\beta_1, V_1) \\
\sigma^2|\beta, y &\sim \text{IG}\Big((\alpha_0 + n)/2, \big[\delta_0 + \big(y - x\beta^{(t)}\big)'\big(y - x\beta^{(t)}\big)\big]/2\Big).
\end{aligned}
$$

Gibbs 抽样：例

```
. use taylor, clear

. bayes,  gibbs : regress ffr L.ffr L.pi L.ygap

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  ffr ~ normal(xb_ffr,{sigma2})

Priors:
  {ffr:L.ffr L.pi L.ygap _cons} ~ normal(0,10000)                          (1)
                       {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_ffr.

Bayesian linear regression                       MCMC iterations  =     12,500
Gibbs sampling                                   Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        113
                                                 Acceptance rate  =          1
                                                 Efficiency:  min =      .9539
                                                              avg =      .9849
Log marginal-likelihood = -186.84401                          max =          1

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
ffr          |
         ffr |
         L1. |  .8072651   .0443575   .000444   .8071702   .7203961   .8947774
             |
          pi |
         L1. |  .3984596   .0892444   .000892    .398595   .2208014   .5715274
             |
        ygap |
         L1. |  .2602214   .0756185   .000756   .2597978   .1120771   .4083866
             |
       _cons | -.0706846    .195486   .001984  -.0697387  -.4545969   .3162705
```

sigma2	.9039901	.1246499	.001276	.8923206	.6947758	1.184623

```
Note: Default priors are used for model parameters.
```

Gibbs 抽样

半共轭先验：条件后验分布与先验分布服从同一个分布族。共轭先验是指所有参数的后验分布与联合先验同属一个分布族。

线性回归：$y = x\beta + u, u \sim (0, \sigma^2). \beta \sim$ Multivariate-Normal, $\sigma^2 \sim$ Inverse-Gamma.

Probit 回归：$P(y = 1) = \Phi(x\beta). \beta \sim$ Multivariate-Normal.

提高 MH 抽样有效性

是否有 Gibbs 算法

提高模拟样本量(`mcmcsize()`)；燃烧样本量(`burnin()`)；间隔取值(`thinning()`)

修改先验分布(`prior()`)：为受限分布的参数设定合理的先验，比如参数取离散值、取正数、取负数、取特定区间（比如 0-1）等。

设定模拟的初始值

观察参数的相关性，进行分块抽样(`block()`)

Stata

bayesopts	note
`nchain(#)`	链的个数，默认值为 `nchain(1)`
`mcmcsize(#)`	模拟样本量，默认值为 `mcmcsize(10000)`
`burnin(#)`	默认值为 `burnin(2500)`
`thinning(#)`	间隔数，默认值为 `thining(1)`
`rseed(#)`	

Contents

蒙特卡洛模拟

马尔科夫链蒙特卡洛模拟

- 马尔科夫链
- MCMC（MH 抽样）
- 模拟结果的质量诊断与统计

Gibbs 抽样

贝叶斯计量分析与预测

Stata: bayes

bayes [, *bayesopts*] : *estimation_command* [, *estopts*]

bayesopts	note
`normalprior(#)`	回归系数的正态先验分布的标准差，默认值为 100（方差为 10,000）
`igammaprior(# #)`	回归模型的方差的先验分布：inverse-gamma 分布(shape, scale)，默认值为 `igammaprior(0.01 0.01)`
`prior(`*priorspec*`)`	自己设定先验分布
`gibbs`	只适用于 regress, mvreg 等模型

其中，*priorspec* 的形式为：{[*eqname*:]*param*[, matrix]}, *priordist*。比如，

```
prior({lwage:educ}, normal(0, 9)) prior({lwage:age}, uniform(-3,3))
```

```
mvec = (1,0,2)
vmat = I(3)*100
prior({lwage:}, mvnormal(3, mvec, vmat))
```

Stata: bayes

Stata 对不同模型的不同类型的参数设置了不同的先验分布。

回归系数的先验分布默认为独立的正态分布$N\left(0,\sigma_{prior}^2\right)$，其中$\sigma_{prior}^2 = 10{,}000$。

- 在多数情况下，$N(0,10{,}000)$是无信息的（uninformative），但如果估计量的数值非常大（比如 2000），那么标准差 100 会仍然含有较强的信息（informative）。这时，需要自己通过 **normalprior()**设定先验分布的方差。
- 一般情况下，可以通过更改变量的测量单位以得到大小适宜的估计量。

正值的参数（比如方差）默认先验分布为 Inverse-Gamma(α,β)。默认值为$\alpha = 0.01, \beta = 0.01$。

贝叶斯计量模型

```
. use taylor, clear

. bayes,  gibbs : regress ffr L.ffr L.pi L.ygap

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  ffr ~ normal(xb_ffr,{sigma2})

Priors:
  {ffr:L.ffr L.pi L.ygap _cons} ~ normal(0,10000)                          (1)
                       {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_ffr.

Bayesian linear regression                       MCMC iterations  =     12,500
Gibbs sampling                                   Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =        113
                                                 Acceptance rate  =          1
                                                 Efficiency:  min =      .9491
                                                              avg =      .9898
Log marginal-likelihood = -186.86974                          max =          1

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
ffr          |
         ffr |
         L1. |  .8061786   .0438886   .000439    .806048   .7197456   .8916158
             |
          pi |
```

```
          L1. |  .4004203   .0878048   .000878   .3996643   .2285947   .5742032
              |
         ygap |
          L1. |   .261852   .0738843   .000739   .2616932   .1124488   .4088988
              |
        _cons | -.0689398   .1941942   .001922  -.0680123  -.4544787   .3127534
--------------+----------------------------------------------------------------
       sigma2 |   .903169   .1246849    .00128   .8932442   .6902638   1.177074
-------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

贝叶斯预测

由于参数是随机的，MCMC 模拟得到T个值。因此，其对单个观测值存在T个预测值。样本量为N的样本，预测值为$T \times N$矩阵。

$$\begin{aligned} y_i &= x_i\beta + u_i, i = 1,2,\cdots,N \\ \beta &= \left(\beta^{(1)}, \beta^{(2)}, \cdots, \beta^{(T)}\right) \end{aligned}$$

即，对每个观测值的预测都对应一个变量。如果对所有观测值进行贝叶斯预测，那么会得到 N 个预测变量，每个预测变量有 T 个贝叶斯预测值。

由于预测的数据量较大，一般只对部分感兴趣的观测值或者观测的描述指标进行贝叶斯预测，比如均值、中位数、最小值、最大值、四分间距等。

也可以只利用T个模拟值的一部分进行贝叶斯预测。

贝叶斯预测: Stata

bayespredict {**_ysim**#[*numlist*]}

numlist 设定对哪些观测值进行预测。得到的预测结果：

- 因变量的 N 个预测变量: _ysim1_1, _ysim1_2, …， _ysim1_*N*，表示第 1 个方程中第 1、2、…、N 个观测值的预测变量。
- 残差为：_resid1_1, _resid1_2, …， _ysim1_*N*，表示第 1 个方程中第 1、2、…、N 个残差的预测变量。
- 均值为：_mu1_1, _mu1_2, …， _mu1_*N*，表示第 1 个方程中第 1、2、…、N 个期望值的预测变量。

Stata 根据 bayesmh 中 likelihood()或 bayesmh 所设定的模型自动计算期望值。比如，第 i 个观测的第 t 次模拟：

- 线性回归: $\mu_i^{(t)} = x_i\beta^{(t)}$
- Probit: $\mu_i^{(t)} = \exp\left(x_i\beta^{(t)}\right)$

bayespredict 只能用于 bayesmh，而不能用于 bayes。

贝叶斯预测

```
. use taylor, clear

. qui bayesmh ffr l.(ffr pi ygap) in 1/-6, likelihood(normal({var}))  prior({ff
r:}, normal(10)) prior({var}, igamma(0.01
> , 0.01)) saving(taylormh, replace) rseed(123)

. bayespredict {_ysim} in 110/114,  saving(taylor_pred, replace) rseed(16)

Computing predictions ...

file taylor_pred.dta saved.
file taylor_pred.ster saved.

. describe _ysim1_110 _ysim1_111 _ysim1_114 using taylor_pred

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
---------------------------------------
_ysim1_110     double   %10.0g                Simulated ffr, obs #110
_ysim1_111     double   %10.0g                Simulated ffr, obs #111
_ysim1_114     double   %10.0g                Simulated ffr, obs #114
```

贝叶斯预测

模拟的后 5 个观测值的统计指标。

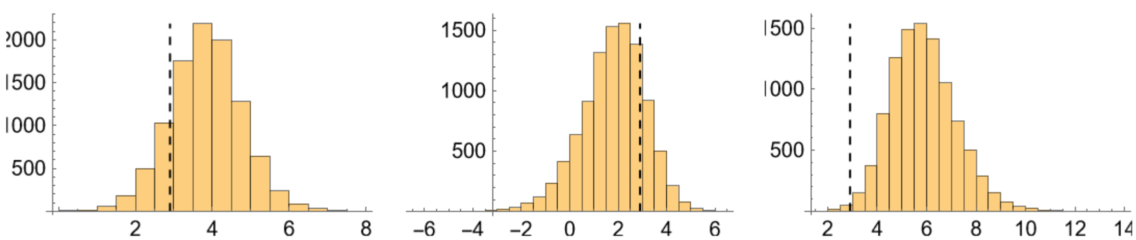
```
. bayesstats summary {_ysim[110/114]} using taylor_pred

Posterior summary statistics                      MCMC sample size =     10,000

-------------------------------------------------------------------------------
```

```
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
  _ysim1_110 |  5.447771   .9888649   .011341   5.450288   3.480159   7.397394
  _ysim1_111 |  5.510331   .9842391   .009942   5.505292   3.574745   7.416676
  _ysim1_112 |  5.352498   .9961558   .010243   5.353572   3.380797    7.32326
  _ysim1_113 |  5.290157   .9758961   .009759   5.300427   3.386443   7.218715
  _ysim1_114 |  5.181558   .9752148   .010116   5.192345   3.261653   7.076171
------------------------------------------------------------------------------
```

贝叶斯预测

贝叶斯预测也是检查模型设定的一种方法，通过比较预测值的统计量(比如，均值、最小值、最大值、分位数、偏度、峰度、对称性)分布与实际观测值的统计量进行比较检验（bayestest interval)。

观测值	y_{46}	y_{47}	y_{48}	y_{49}	y_{50}		mean	min	max
observed	2.9	2.9	2.9	2.9	2.9	⇒	2.9	2.9	2.9
simulated	5.01724	1.68042	3.93164	5.10653	6.3661		4.42038	1.68042	6.3661
	3.21664	2.61576	6.89439	4.5356	-1.49672		3.15313	-1.49672	6.89439
	2.64348	5.82413	3.15353	0.339442	1.79671	⇒	2.75146	0.339442	5.82413
	8.97096	3.68248	4.03834	4.89328	5.27929		5.37287	3.68248	8.97096
	7.24218	4.91776	4.12351	4.69013	3.0464		4.804	3.0464	7.24218
	...	...	...	...	...		...	...	...



贝叶斯预测

```
. bayesstats ppvalues (mean:@mean({_ysim})) (min:@min({_ysim})) (max:@max({_ysi
m})) using taylor_pred

Posterior predictive summary   MCMC sample size =    10,000

-----------------------------------------------------------
           T |      Mean   Std. dev.  E(T_obs)  P(T>=T_obs)
```

mean	5.356463	.4496118	5.214667	.6276
min	4.211609	.6729031	5.073333	.0896
max	6.504275	.6754822	5.256667	.979

Note: P(T>=T_obs) close to 0 or 1 indicates lack of fit.

这些统计量的概率值越接近 0.5，表明模拟值与实际值越吻合。

贝叶斯诊断主要考察抽样的收敛性，而贝叶斯预测检查是考察模型设定的合理性。不合理的模型也往往容易影响到 MCMC 抽样的有效性。

谢 谢！