Stataで学ぶベイズ統計

LightStone Corp

2020年3月

修正2020年4月

第4回 ベイズ統計によるモデル推定

- Stata 英文PDFマニュアルの例題を利用してベイズ統計を使ったモデル推定の方法を学ぶ
  - データ:oxygen.dta
  - 目的:酸素摂取量を増やすための運動として1)ステップエアロビクス、2)平地でのランニングのどちらが効果的か?
  - サンプル:24人の健康な男性を1または2のグループに無作為割付する
  - アウトカムはトレーニング実施前後に測定した最大酸素摂取量の差分
  - 持久力がついてくると、最大酸素摂取量は大きくなる

構成

- Example 1:OLSによる推定
- Example 2:無情報事前分布によるベイズ推定
- Example 3:事前分布を利用したベイズ推定
- Example 4:多変量事前分布を利用したベイズ推定
- Example 5:シミュレーションの収束
- Example 6:推定後の統計量
- Example 7:ベイズ予測
- Example 8:モデルの選択
- Example 9:仮説検定
- Example10:シミュレーションに利用したデータセット
- Example11:bayesプリフィックスの利用

Example 1

- ベイズ推定の前に最小二乗法(OLS)による回帰分析を実施する
- OLSにおけるp値と95%信頼区間(95%CI)の意味を確認する(ベイズのそれとは異なるため)

```
. use https://www.stata-press.com/data/r16/oxygen,clear
. des
```

variable name	storage type	display format	value label	variable label
change	float	%9.0g		Change in maximal oxygen uptake (liters/minute)
group	byte	%8.0g	grouplab	Exercise group (0: Running, 1: Aerobic)
age	byte	%8.0g		Age (years)
ageXgr	byte	%9.0g		Interaction between age and group

OLSによるモデル推定

$$\text{change} = \beta_0 + \beta_{\text{group}}\,\text{group} + \beta_{\text{age}}\,\text{age} + \epsilon$$

- OLSでこのモデルを推定する

```
. reg change group age

      Source |       SS           df       MS      Number of obs   =        12
-------------+----------------------------------   F(2, 9)         =     41.42
       Model |  647.874893         2  323.937446   Prob > F        =    0.0000
    Residual |   70.388768         9  7.82097423   R-squared       =    0.9020
-------------+----------------------------------   Adj R-squared   =    0.8802
       Total |  718.263661        11  65.2966964   Root MSE        =    2.7966

------------------------------------------------------------------------------
      change |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   5.442621   1.796453     3.03   0.014     1.378763    9.506479
         age |   1.885892    .295335     6.39   0.000     1.217798    2.553986
       _cons |   -46.4565   6.936531    -6.70   0.000    -62.14803   -30.76498
------------------------------------------------------------------------------
```

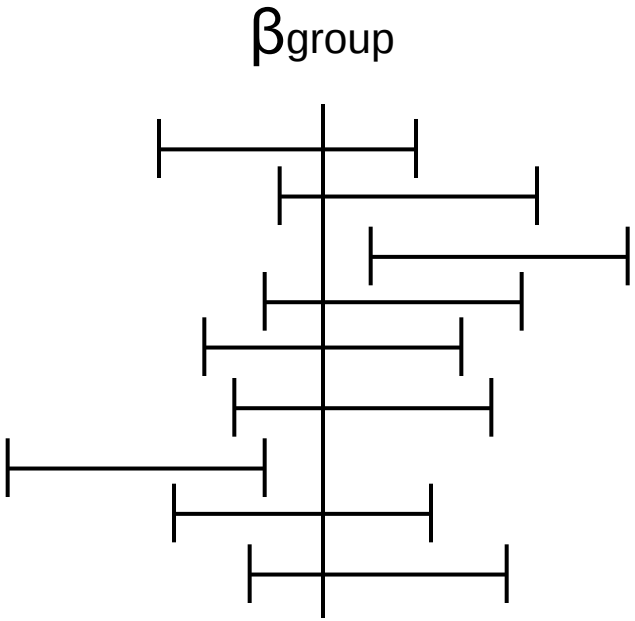
OLSのp値

$$H_0 \ : \ \beta_{\text{group}} \ = \ 0$$

- groupは0または1を取る二値変数で、ランニングは0
- 12週間のステップエアロビクスの効果は5.44で、そのp値は0.014
- p値=0.014とは、同じ研究デザインの実験を、例えば、100回繰り返したとき、帰無仮説であるβ_{group}=0が観測される回数は1.4回ということを示している
- つまり、手元のデータから帰無仮説の主張を受け入れることはできない

OLSの95%CI

- β_{group}の95%CIは[1.38 9.51]
- p値と同じように同じ実験を100回繰り返したとき、β_{group}の95%CIが母数(真の値)を含む回数が95回



Example 2

- Stataにおける簡単なベイズ推定の方法を学ぶ
- 頻度を用いるOLSでは確率変数は被説明変数Yだけ。説明変数Xとパラメータは固定した値と考える
- ベイズ推定ではYだけでなく、パラメータも確率変数と考える

$$\text{change} = \beta_0 + \beta_{\text{group}}\,\text{group} + \beta_{\text{age}}\,\text{age} + \epsilon$$

- ベイズ推定ではパラメータとしてYの分散も推定する。したがって、パラメータは4つ考える

事前分布

- 仮定
  - アウトカムchangeは正規分布に従う
  - 4つのパラメータはJeffreys事前分布に従う

$$\text{change} \sim N(X\beta, \sigma^2)$$

$$(\beta_0, \beta_{\text{group}}, \beta_{\text{age}}, \sigma^2) \sim \frac{1}{\sigma^2}$$

事前分布の選択

- 事前分布はすべてのパラメータに対して設定する
  - 事前分布は推定に役立つinformativeなものを選択する
  - 何も情報がない場合にはnoninformative(無情報)な分布を利用する
  - 無情報分布の代表的なものとして一様分布がある
  - 正規分布とガンマ分布も分散が大きくなることを許容すれば、無情報分布として利用できる
  - 分散の逆数を利用する無情報分布としてJeffreys事前分布というものがある(Jeffreys 1946)
  - ただし、無情報分布はMCMCの計算に悪影響を与える事が多い
  - 良い事前分布を見つけることがベイズ推定の基本

ベイズ推定

- MHアルゴリズム

  . bayesmh change group age, likelihood(normal({var}))
  prior({change:},flat) prior({var}, jeffreys)

- 部分ごとに見ていく

  bayesmh change group age

  OLSのモデル表現と同じ。コマンドをbayesmhにする

  likelihood(normal({var}))

  データの尤度の設定。正規分布normalに従い、分散を推定するため大カッコを使って{var}とし、推定値であることを示す

ベイズ推定

. bayesmh change group age, likelihood(normal({var}))
prior({change:},flat) prior({var}, jeffreys)

- prior({change:},flat)はモデルに含まれる3つのβの事前分布を一様分布としている
- 3つのβとは{change:group}, {change:age}, {change:_cons},
- prior({var}, jeffreys)は分散varの事前分布をJeffreysとしている
- オプションprior()は何回でも繰り返し利用できる

バーンイン

- 乱数キーを利用してマニュアルの推定結果を再現する

```
. set seed 14
. bayesmh change group age, likelihood(normal({var}))
                    prior({change:},flat) prior({var}, jeffreys)
```

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =      .1371
Efficiency:  min =     .02687
             avg =     .03765
             max =     .05724
```

- MCMCの繰り返し計算回数のデフォルトは12,500回
- 乱数の偏りの影響を排除するためのバーンイン(稼働検査期間)は2,500回
- 従って10,000個の事後分布を利用して、その平均を求める

受容率

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =      .1371
Efficiency:  min =     .02687
             avg =     .03765
             max =     .05724
```

- MHアルゴリズムにおける今回の推定の受容率(Acceptance rate)は0.14
- 10,000個の提案分布によるパラメータのうち、MHアルゴリズムで受容したのは14%
- 受容率が50%を超えることは稀

- 受容率は一般に30%以下。10%以下の場合、収束計算に問題がある。この例題の14%はやや問題ありと考えるのが妥当
- 効率性(Efficiency)は標本に占める、独立標本の数(推定値)
- 効率性は大きい方が良い
- 効率性が1%未満の場合は好ましくない
- この例の効率性は約5.8%(max)で、やや問題あり

信用区間

事後分布の標準偏差(OLSのSEに相当)

事後分布の中央値(Meanとの比較で対称性を判断)

真の値が信用区間に存在する可能性が95%

事後分布の平均

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	5.429677	2.007889	.083928	5.533821	1.157584	9.249262
age	1.8873	.3514983	.019534	1.887856	1.184714	2.567883
_cons	-46.49866	8.32077	.450432	-46.8483	-62.48236	-30.22105
var	10.27946	5.541467	.338079	9.023905	3.980325	25.43771

Monte Carlo標準誤差…事後分布の平均の精度。MCMCの回数が大きいほど小さくなる

Example 3の前に

- ガンマ分布の密度関数(Stata)

$$\mathrm{gammaden}(a,b,g,x) = \frac{1}{\Gamma(a)b^a}(x-g)^{a-1}e^{-(x-g)/b}$$

- g=0の時(Stata)をG(･)表現すると

$$\mathrm{Ga}(x|a,b) = \frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}$$

Example 3の前に

- bを逆数にして表現するテキストもある

$$\mathrm{Ga}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

- この時、逆ガンマ分布は

$$\mathrm{InvGa}(x|a,b) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{x}\right)^{a+1} e^{-b/x}$$

- ガンマ分布や逆ガンマ分布はパラメータの使い方が2通りあることに注意

Example 3

- 事前分布の変更
- 分散の事前分布だけを変更した推定例

$$(\beta_0, \beta_{\text{group}}, \beta_{\text{age}} | \sigma^2) \sim \text{i.i.d. } N(0, \sigma^2)$$

$$\sigma^2 \sim \text{InvGamma}(2.5, 2.5)$$

3つの{var}は同じ統計量を指す

```
. set seed 14
. bayesmh change group age, likelihood(normal({var}))
        prior({change:}, normal(0, {var})) prior({var}, igamma(2.5, 2.5))
```

- 係数パラメータの事前分布は正規分布(Example 2と同じ)
- データの分散{var}は逆ガンマ分布に従う

分散の事前分布

- 逆ガンマ分布を利用した

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =      .1984
Efficiency:  min =     .03732
             avg =     .04997
             max =     .06264
```

受容率は改善した(14%)

効率性も改善した(5.8%)

事前分布の変更

- 事前分布の設定によって推定結果が変化する

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	6.510807	2.812828	.129931	6.50829	.9605561	12.23164
age	.2710499	.2167863	.009413	.2657002	-.1556194	.7173697
_cons	-6.838302	4.780343	.191005	-6.683556	-16.53356	2.495631
var	28.83438	10.53573	.545382	26.81462	14.75695	54.1965

- ここでは単純に事前分布の影響力を考察するために、逆ガンマ分布を利用した
- 逆ガンマ分布は正規分布の共役事前分布だが、これを利用した特別な理由はない

Example 4

- 線形回帰モデルのパラメータ用によく利用されるZellner (1986)の事前分布gを利用する

$$(\beta_0, \beta_{\text{group}}, \beta_{\text{age}} | \sigma^2) \sim \text{MVN}\left(0, g\sigma^2 (X'X)^{-1}\right)$$

$$\sigma^2 \sim \text{InvGamma}(\nu_0/2, \nu_0 \sigma_0^2/2)$$

自由度

サンプルサイズ

係数間の依存性に影響する

Hoff (2009)

パラメータの個数

```
. set seed 14
. bayesmh change group age, likelihood(normal({var}))
          prior({change:}, zellnersg0(3,12,{var}))
          prior({var}, igamma(0.5, 4))
```

Hoff (2009)にならってパラメータは次のように設定した

$$g = 12, \nu_0 = 1, \sigma_0^2 = 8$$

事後分布

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =     .06169
Efficiency:  min =      .0165
             avg =     .02018
             max =     .02159
```

受容率は19%から悪化し、下限である10%も満たしていない

効率性も6.2%から悪化している

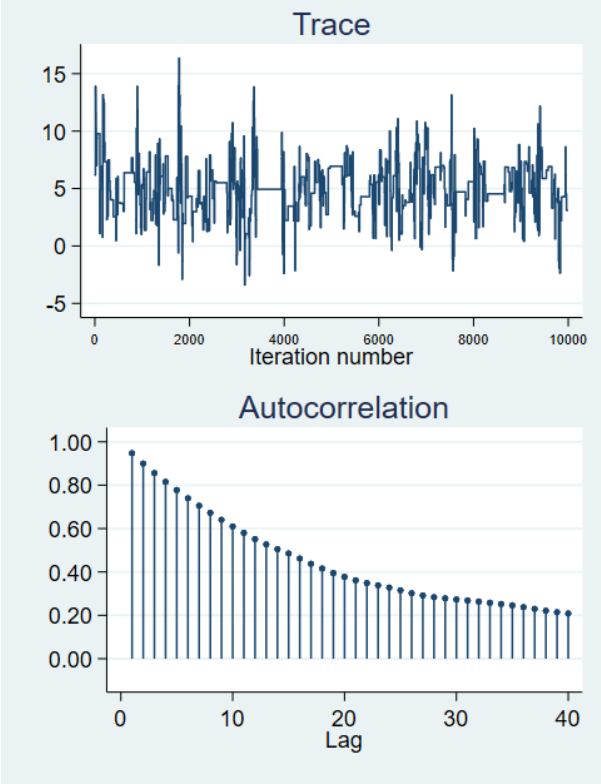
事後分布

- Example 4の推定結果はExample 2の結果に近い
- 受容率はExample 2に比べて0.1371->0.06169
- 効率性は0.06264(max)->0.02159(max)
- Example 4の結果はExample 3だけでなく、Example 2の結果よりも悪い
- 設定を工夫する必要がある
- ここでは分析の流れを理解することを目的として、Example 4の結果を使ってそのまま、MCMCの診断を行う

Example 5

- MCMCの繰り返し計算における係数の収束状態を確認する

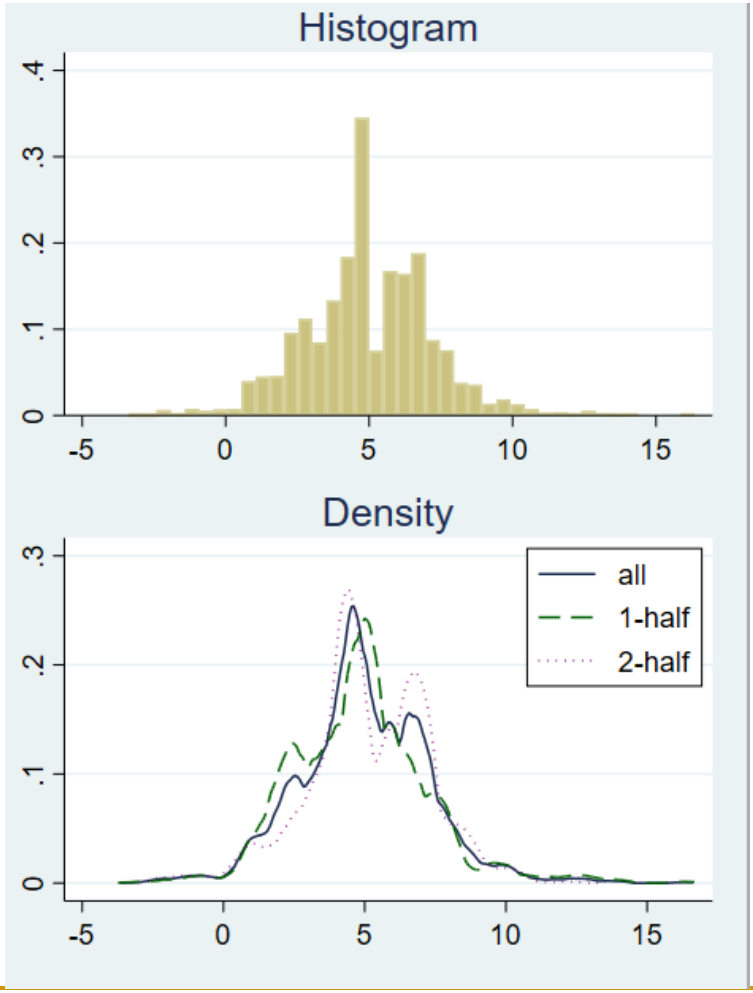
. bayesgraph diagnostics {change:group}



変数groupの係数推定値は平均4.988の周辺に分布している(1万個)

変数groupの係数推定値の自己相関は徐々に小さくなっている(自己相関は急激に減衰する方が良い)

分布



変数groupの係数推定値10,000個のヒストグラム。左右対称に分布しているが、単峰型ではないように見える

10,000個、5,000個(前半、後半)の密度プロット。シミュレーションに偏りがなければ、ほぼ一致する

効率性

- ESSがMCMCサンプルサイズに近いほど、MCMCサンプルの相関は弱い。ここでは両者に乖離があるので、逆にMCMCサンプルの自己相関は強い(MHでは一般に強い)

```
Efficiency summaries     MCMC sample size =    10,000
                         Efficiency:  min =     .0165
                                      avg =    .02018
                                      max =    .02159
```

Effective Sample Size

	ESS	Corr. time	Efficiency
change			
group	215.93	46.31	0.0216
age	214.39	46.64	0.0214
_cons	212.01	47.17	0.0212
var	165.04	60.59	0.0165

- ESSがMCMCサンプルサイズの1%以下になる場合は不適切。この値はESS/MCMCのサンプルサイズで計算する

効率性

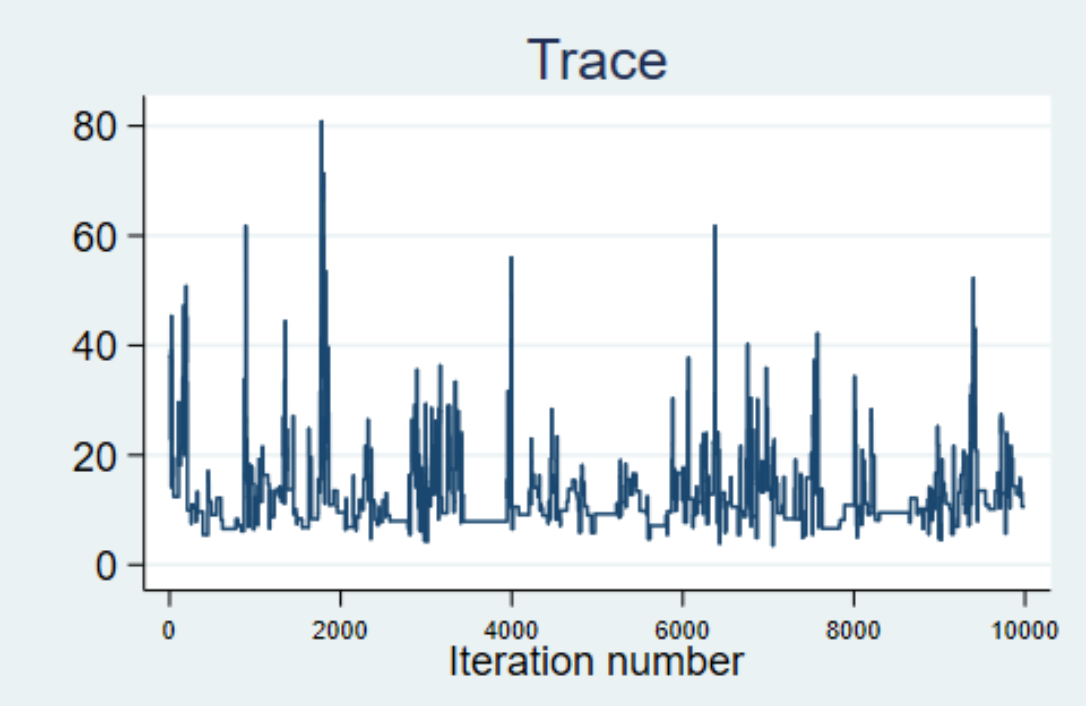
- 分散{var}の効率性が悪い

	ESS	Corr. time	Efficiency
change			
group	215.93	46.31	0.0216
age	214.39	46.64	0.0214
_cons	212.01	47.17	0.0212
var	165.04	60.59	0.0165

- Correlation timeはMCMCサンプルにおける自己相関のラグの推定値
- 値が大きいと効率性が低いことを示す

分散の収束

. bayesgraph diagnostics {var}



自己相関の影響で連続して同じような値をとっているように見受けられる

計算上の工夫

- blockオプションを利用して、係数とは独立した形でMCMCを実行する

```
. set seed 14
. bayesmh change group age, likelihood(normal({var}))
          prior({change:}, zellnersg0(3,12,{var}))
          prior({var}, igamma(0.5, 4)) block({var})
          saving(agegroup_simdata,replace)
. est store agegroup
```

注意)MCMCによる推定結果をest storeコマンドで保存するためにはsavingオプションでシミュレーション結果をファイル保存する必要がある

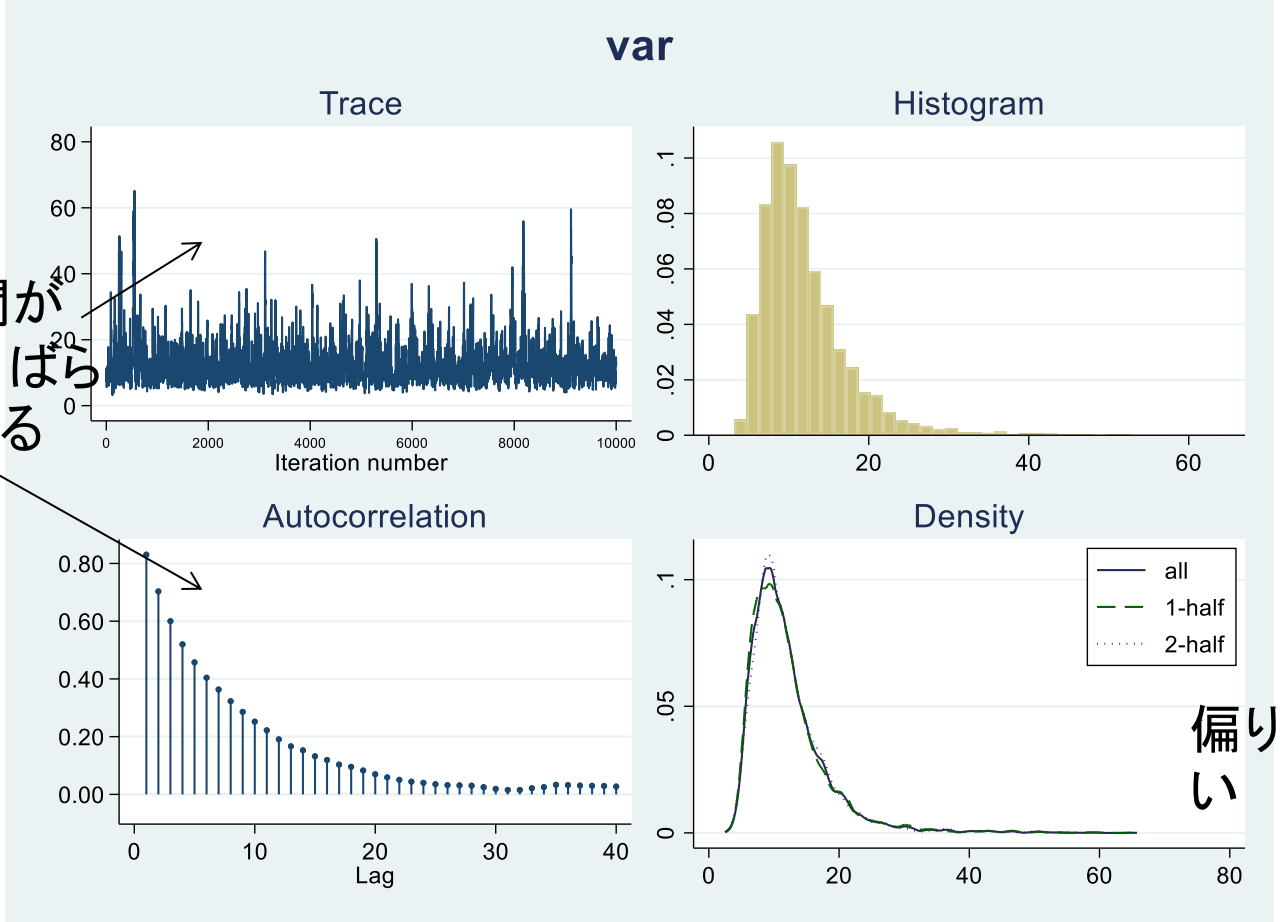
ブロックオプションの利用

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =      .3232
Efficiency:  min =     .06694
             avg =      .1056
             max =      .1443
```

受容率は0.06169から0.3232
効率性(max)は0.06264から
0.1443に改善された

分散の収束

. bayesgraph diagnostics {var}



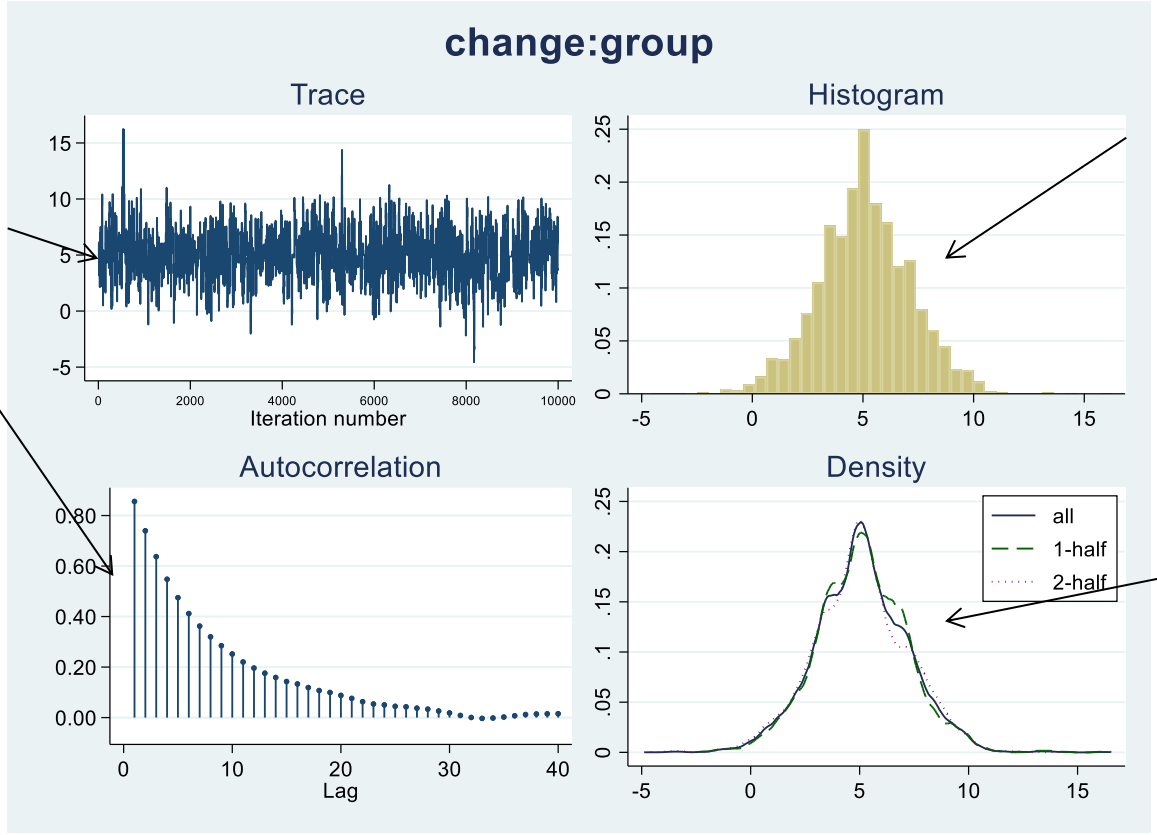
自己相関が弱くなり、ばらついている

偏りが小さい

係数の収束

. bayesgraph diagnostics {change:group}

自己相関が弱くなり、ばらついている



単峰型

偏りが小さい

Example 6

- 推定結果の表示と計算

. bayesstats summary

- bayesstats summaryコマンドはMCMCの計算結果として得られる事後分布の情報だけを表示するコマンド
- 実際、Example 5の操作の後にコマンドを実行すると、次の表だけを出力する

Posterior summary statistics　　MCMC sample size = 10,000

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	5.080653	2.110911	.080507	5.039834	.8564619	9.399672
age	1.748516	.3347172	.008875	1.753897	1.128348	2.400989
_cons	-43.12425	7.865979	.207051	-43.2883	-58.64107	-27.79122
var	12.09916	5.971454	.230798	10.67555	5.375774	27.32451

事後分布の計算

- 事後分布の係数を使って他の計算を行う
- 標準化回帰係数の計算例

$$\hat{\beta}_{\text{group}} \times \sigma_x / \sigma_y$$

```
. su group
. scalar sd_x=r(sd)
. su change
. scalar sd_y=r(sd)

. bayesstats summary (group_std:{change:group}*sd_x/sd_y)
```

目的とする統計量の名前

計算式

係数の指定

標準化回帰係数の計算例

- bayesstats summaryコマンド

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

   group_std : {change:group}*sd_x/sd_y
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
group_std	.3283509	.1364233	.005203	.3257128	.0553512	.6074792

Example 7

- ベイズ予測
  - changeの予測値(フィット値)を計算する
  - MCMCサンプル(1万個)分の予測値を計算し、別のデータファイルとして保存する

```
. bayespredict {_ysim}, saving(change_pred,replace) rseed(16)
```

  - 元のデータoxygen.dtaは観測値が12行だけのデータ
  - 新たにMCMCの結果を利用した予測値データファイルchange_pred.dtaを作成する
  - rseed(16)は任意の乱数キー

フィット値

- MCMCサンプルのデータ(change_pred.dta)を開いて中身を後で確認してみよう
- 12個のデータについてMCMCサンプルを使った予測値を求める

```
. bayesstats summary {_ysim} using change_pred
```

フィット値

- MeanはMCMCサンプルを利用して求めたアウトカムchangeの平均

Posterior summary statistics　　MCMC sample size = 10,000

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
_ysim1_1	-2.954378	3.763301	.060963	-2.930854	-10.39297	4.528522
_ysim1_2	-4.610688	3.771203	.059014	-4.660554	-11.9289	2.948378
_ysim1_3	-4.620784	3.758543	.057517	-4.645584	-12.03851	2.917013
_ysim1_4	.6417156	3.756645	.063162	.6019013	-6.83463	8.330498
_ysim1_5	4.069868	3.972042	.072874	4.065139	-3.780329	12.06363
_ysim1_6	-8.120147	3.832453	.061674	-8.096888	-15.54334	-.3579446
_ysim1_7	16.18539	4.076738	.072385	16.2033	8.105208	24.23569
_ysim1_8	2.156433	3.921	.072344	2.135557	-5.528265	10.00732
_ysim1_9	9.14268	3.780417	.071241	9.154486	1.571643	16.59816
_ysim1_10	10.91948	3.776916	.068083	10.92263	3.445305	18.59981
_ysim1_11	.3919052	3.969695	.079798	.344616	-7.389234	8.386358
_ysim1_12	3.902787	3.809399	.077872	3.884087	-3.530938	11.49579

適合度

- 観測値とMCMCサンプルを利用した計算値の平均が、どの程度近しい値になっているか、調べる場合にはbayesstats ppvaluesコマンドを利用する

```
. bayesstats ppvalues {_ysim} using change_pred
Posterior predictive summary   MCMC sample size =    10,000
```

T	Mean	Std. Dev.	E(T_obs)	P(T>=T_obs)
_ysim1_1	-2.954378	3.763301	-.87	.2786
_ysim1_2	-4.610688	3.771203	-10.74	.9512
_ysim1_3	-4.620784	3.758543	-3.27	.3479
_ysim1_4	.6417156	3.756645	-1.97	.773
_ysim1_5	4.069868	3.972042	7.5	.1819
_ysim1_6	-8.120147	3.832453	-7.25	.4034
_ysim1_7	16.18539	4.076738	17.05	.4124
_ysim1_8	2.156433	3.921	4.96	.2198
_ysim1_9	9.14268	3.780417	10.4	.3644
_ysim1_10	10.91948	3.776916	11.05	.4858
_ysim1_11	.3919052	3.969695	.26	.5106
_ysim1_12	3.902787	3.809399	2.51	.6498

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

p値が0.05から0.95の間に存在すれば良い。2番目の観測値については不十分。

外挿

- 12行のデータセットに2つのデータを追加してアウトカムchangeの予測値を求める
- group=1はエアロビクス、0はランニング

```
. set obs 14
. replace group = 1 in 13
. replace group = 0 in 14
. replace age = 26 in 13/14
. bayespredict pmean, mean rseed(16)
. list change age group pmean
```

外挿

- 13,14行目の予測計算

	change	age	group	pmean
1.	-.87	23	Running	-2.914124
2.	-10.74	22	Running	-4.613421
3.	-3.27	22	Running	-4.701283
4.	-1.97	25	Running	.545417
5.	7.5	27	Running	4.060798
6.	-7.25	20	Running	-8.111091
7.	17.05	31	Aerobic	16.15393
8.	4.96	23	Aerobic	2.183771
9.	10.4	27	Aerobic	9.155602
10.	11.05	28	Aerobic	10.87576
11.	.26	22	Aerobic	.4234267
12.	2.51	24	Aerobic	3.937901
13.	.	26	Aerobic	7.380203
14.	.	26	Running	2.405744

予測計算

- 12行目までのchangeとpmeanがどの程度近しいかもチェックしよう

準備

- 次のExample 8に進む前に追加してデータを削除し、MCMCサンプルのファイルも削除する

```
*データを消す
. drop in 13/14
. erase change_pred.dta
. erase change_pred.ster
```

est saveコマンドで保存した推定結果だけを含むStataファイル

Example 8

- ベイズ推定におけるモデル選択
- 交互作用項c.age#i.groupを追加したモデルfullを推定する

```
. set seed 14
. bayesmh change group age c.age#i.group, likelihood(normal({var}))
          prior({change:}, zellnersg0(4,12,{var}))
          prior({var}, igamma(0.5, 4)) block({var})
          saving(full_simdata,replace)

. estimates store full
```

モデルの選択

- モデルの比較にはbayesstats icコマンドを利用する

. bayesstats ic full agegroup

限界尤度(marginal likelihood)…大きい方が良い

逸脱度(deviance)…小さい方が良い

	DIC	log(ML)	log(BF)
full	65.03326	-36.73836	.
agegroup	63.5884	-35.46061	1.277756

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

Bayes Factor

- BFは事前分布も考慮する
- log(BF) はベースモデルに対してどの程度優れているか、目安となる統計量である
- AIC,BIC,DICは尤度だけしか利用しない
- BFを利用する場合、サンプルサイズが一致することが条件となる

	DIC	log(ML)	log(BF)
full	65.03326	-36.73836	.
agegroup	63.5884	-35.46061	1.277756

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

Bayes Factor

- Kass and Raftery (1995)

$2\log\left(\mathrm{BF}_{jb}\right)$	BF_{jb}	ベースモデルとの比較
0−2	1−3	少なくとも良い
2−6	3−20	良い
6−10	20−150	明らかに良い
>10	>150	非常に良い

□ logBF*2=2.56なのでagegroupのモデルがfullモデルよりも優れている

Example 9

- データを使って事後分布の確率を計算し、比較する

. bayestest model full agegroup

(仮定)2つのモデルの事後分布は等しい

実際に計算すると差が出た

.

	log(ML)	P(M)	P(M|y)
full	-36.7384	0.5000	0.2179
agegroup	-35.4606	0.5000	0.7821

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

確率が0.9以上になる時、BFでもより強くagegroupが選択される

仮説検定

- agegroupにおけるgroupの係数が4と8の間に存在する確率を求める。推定値は5.08

.

	Mean	Std. Dev.	MCSE
prob1	.6159	0.48641	.0155788

- 4と8の間に存在する確率は約62%

Example 10

- 保存した2つのMCMCのシミュレーションファイルを削除する

```
. erase agegroup_simdata.dta
. erase full_simdata.dta
```

  - Mac/Unixの場合はeraseの代わりにrmコマンドを利用する

Example 11

- ベイズ推定にはベイズプリフィックスbayes:を利用する方法もある(こちらの方が新しい機能)
- Stataは利用するコマンドに応じて事前分布を自動設定

```
. set seed 14
. bayes: regress change group age
```

- regressコマンドの場合、係数の事前分布は平均ゼロの正規分布で、分散は10,000
- 分散の分布は逆ガンマ分布で、shapeとscaleパラメータは0.01とする

ベイズプリフィックス

- Example 3(逆ガンマ分布)と比較する

```
MCMC iterations  =     12,500
Burn-in          =      2,500
MCMC sample size =     10,000
Number of obs    =         12
Acceptance rate  =       .283
Efficiency:  min =     .02715
             avg =     .05779
             max =      .0692
```

19%よりも良い

6.2%よりも良い

ベイズプリフィックス

- デフォルト設定が良い結果をもたらすような保証はない
- したがって事前分布は丁寧に与えることが重要

```
                                                  Efficiency:  min =     .02715
                                                               avg =     .05779
Log marginal-likelihood = -45.562124                           max =      .0692
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	5.425311	2.111038	.080252	5.368975	1.104434	9.425197
age	1.885651	.3255098	.012472	1.887263	1.244666	2.517292
_cons	-46.47537	7.632058	.295505	-46.73244	-60.39245	-30.5054
sigma2	10.28431	7.614468	.462105	8.412747	3.595971	31.47161

ギブス・サンプリング

- MHは効率性で劣るのでギブス・サンプリング(gibbs)を利用する
- Gibbsはregressとmvregだけで利用可能

```
. set seed 14
. bayes, gibbs: regress change group age
```

```
Efficiency:  min =       .556
             avg =       .889
             max =          1

Log marginal-likelihood =  -45.83666
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	5.452439	2.062795	.020628	5.460372	1.360104	9.512987
age	1.875606	.330127	.003301	1.877129	1.228647	2.543129

事前分布

- 事前分布を変更する場合はpriorオプションを利用する
- Example 4と同じ推定を実行する

```
. set seed 14
. bayes, prior({change:}, zellnersg0(3,12,{sigma2}))
         prior({sigma2}, igamma(0.5, 4)): regress change group age
```

change						
group	4.944955	2.184113	.086181	5.052278	.7065487	9.35098
age	1.747984	.3390581	.011132	1.747477	1.045677	2.416091
_cons	-43.09605	7.904334	.263186	-43.01961	-58.57942	-27.11278
sigma2	12.17932	5.87997	.220888	10.72651	5.511202	28.1211

- bayes:は自動的に係数と分散をブロッキングするため、推定結果はExample 4とはやや異なる

ブロッキング解除

```
. set seed 14
. bayes, prior({change:}, zellnersg0(3,12,{sigma2})) prior({sigma2},
        igamma(0.5, 4)) noblocking: regress change group age
```

.

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
change						
group	4.988881	2.260571	.153837	4.919351	.7793098	9.775568
age	1.713159	.3545698	.024216	1.695671	1.053206	2.458556
_cons	-42.31891	8.239571	.565879	-41.45385	-59.30145	-27.83421
sigma2	12.29575	6.570879	.511475	10.3609	5.636195	30.93576

第4回のまとめ

- 実際にStataのbayesmhとbayes:を利用して、ベイズ統計によりモデル推定と診断の方法を解説した
- MCMCに特有な統計用語を理解しよう
- ベイズ統計におけるモデルの選択は、尤度だけでなく、事前分布の仕様も考慮した統計量がある
- MCMCの基本的な手法としてMH法があるが、ギブス・サンプリングなど、それ以外の方法もある
- サンプルサイズの小さい今回のようなケースでは、標本に推定結果が大きく依存してしまう。ベイズ推定を用いて2つの立場から推定結果を考察してみよう

参考文献

- STATA BAYESIAN ANALYSIS REFERENCE MANUAL, RELEASE 16

研究者向けの統計解析ソフトウェア Stata® 評価版のお申込み

機能制限なし！30日間無料で使える評価版でまずはお試しください！ https://www.lightstone.co.jp/stata/evaluate.html

※学生の方への評価版の提供はございません。大学の講義で使用するなど単位取得に関わる形でのご利用となる場合は学生版のご購入をご検討ください。

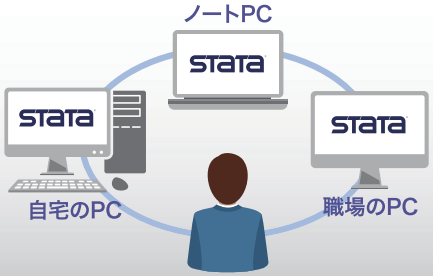
ライセンスはサブスクリプションがおすすめ！

サブスクリプションライセンスを選ぶメリット

- 常にStataの最新バージョンが利用できます
- 技術サポートの対象となります
- 毎年の経費としてStataを導入できます
- 利用者の増減に柔軟に対応可能
- 初期導入費を抑えたい場合にもおすすめ

シングルライセンス（サブスクリプション）

職場のPC・ノートPC・自宅のPC等、個人所有のPCにインストール可能です。
同時に使用できるのは1台のみとなります。



マルチユーザライセンス（サブスクリプション） 2ユーザ以上

ボリュームライセンス

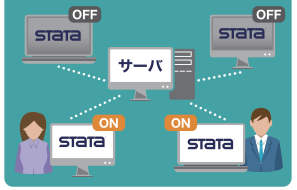
シングルライセンスを複数人でまとめてご購入いただく際のボリュームディスカウント商品です。

同時起動ライセンス

ご利用環境により下記3通りの運用方法からご選択いただけます。

PC毎にスタンドアロンインストールをして運用
例:2ライセンスの場合

クライアント＆サーバ環境によるネットワーク運用
例:2ライセンスの場合



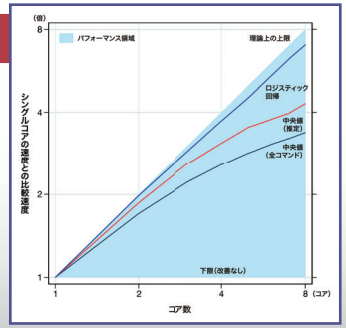
1台の仮想マシンにユーザがアクセスする運用
例:2ライセンスの場合

Stata／MPの演算能力

データ処理時間の削減ならMP

Stata MPはPCの持つマルチコアの特性を活かして、処理を分散・並列する機能を備えます。およそ85%以上のコマンドで処理速度が向上し、コア数に応じた計算時間の短縮が期待できます。
コマンドごとの処理速度の向上の度合いについては以下のページや資料をご覧ください。

コア数	全てのコマンド	推定コマンド	ロジスティック回帰
2	1.7倍	1.8倍	1.9倍
4	2.6倍	3.1倍	3.8倍
8	3.3倍	4.2倍	6.8倍



弊社Webページ内の『Stata／MP』
https://www.lightstone.co.jp/stata/statamp.html

株式会社 ライトストーン
〒101-0031
東京都千代田区東神田 2-5-12 龍角散ビル 7F
TEL：03-3864-5211
FAX：03-3865-0050
e-mail： sales@lightstone.co.jp
WEB： https://www.lightstone.co.jp/