Stataで学ぶベイズ統計

LightStone Corp

2020年2月

第3回 マルコフチェーンモンテカルロ法(MCMC)

- MCMCの基本的な考え方について理解する
  - 共役事前分布の確認
  - 基礎的なモンテカルロ法(A-R法)
  - モンテカルロ積分
  - マルコフ過程とマルコフチェーン
  - メトロポリス-ヘイスティング(MH)アルゴリズム

事前分布の確認

- 第二回の資料の最後に紹介した「熊の出没」に関する分析例を使って問題意識を整理する
- 事前分布が利用できる状況では、新しい情報を使って簡単に事後分布を求めることができた
- 尤度がポアソン分布で、事前分布がガンマ分布の時、パラメータは次のように更新できる
- ガンマ分布のShapeパラメータは5.6から13.6に、rateパラメータは1から2に更新された

$$\alpha' = n + \alpha, \quad \lambda' = \Sigma x_i + \lambda$$

$$\lambda = 5.6 \rightarrow 13.6$$

$$\alpha = 1 \rightarrow 2$$

事前分布の更新

- 更新の様子を図で確認する
- ガンマ分布のパラメータの使い方には複数あり、第2回で紹介した次式の場合、Stataの関数表現とは異なる部分があるので注意が必要

$$w(x) = \frac{\alpha^{\lambda}}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1}$$

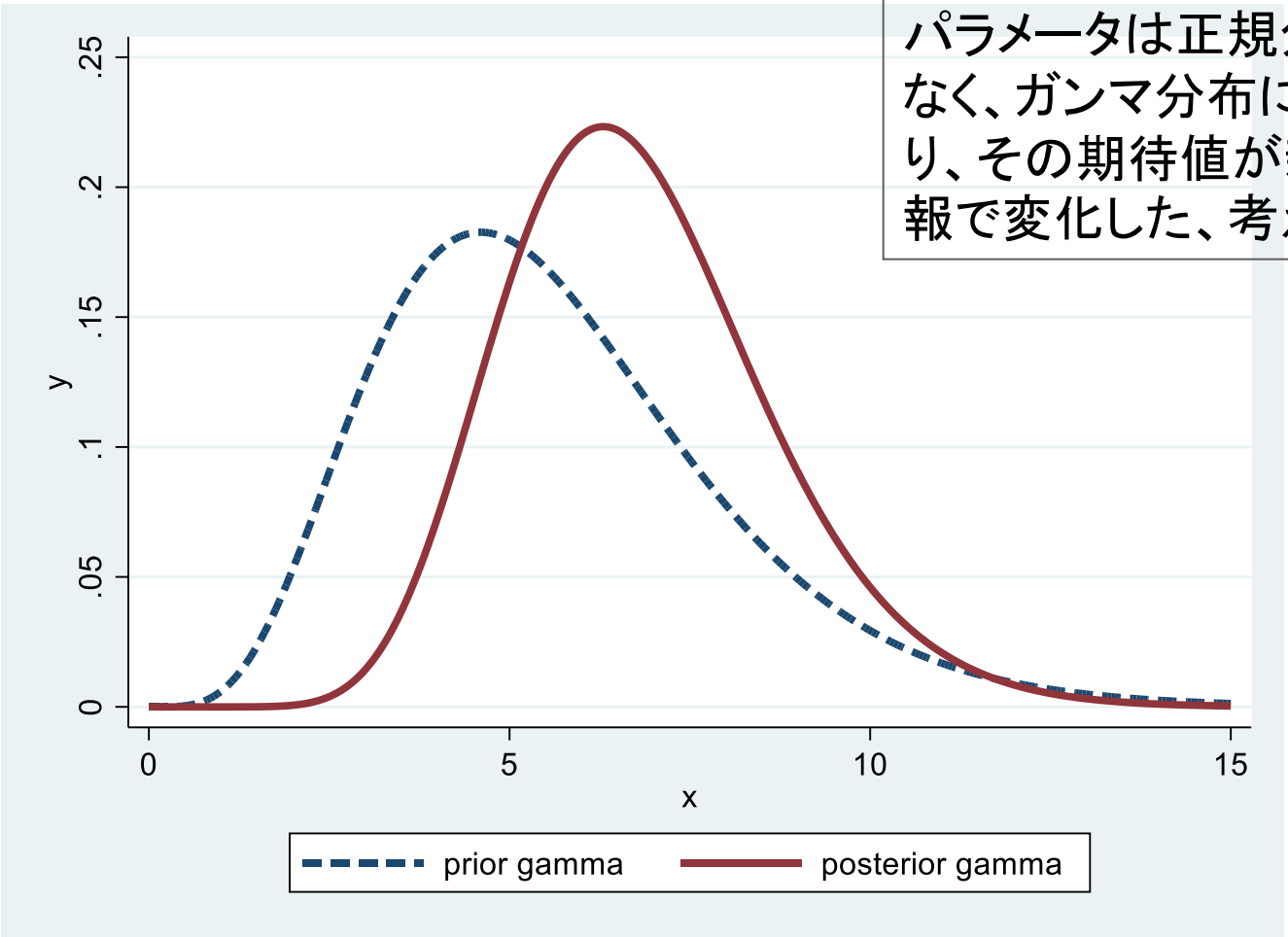
- Stataコマンド(コマンドウィンドウで実行する場合は改行記号 /// は不要)

```
.twoway function y=gammaden(5.6,1,0,x),range(0 15)|| ///
function y=gammaden(13.6,0.5,0,x),range(0 15)
```

- Stataのgamma関数の2番目の引数はαの逆数にする

事前分布の更新

パラメータは正規分布ではなく、ガンマ分布に従っており、その期待値が新しい情報で変化した、考えている



期待値は5.6から6.8に変化した

モンテカルロ法

- MCMCを理解するための基礎知識としてモンテカルロ法について学ぶ
- モンテカルロ法とは確率分布から乱数を発生させる方法
- 問題意識
  - 例えば、Statalにはrnormal()など、乱数を作成する関数が予め用意されている
  - コマンドとして用意されていないものはどうすれば良いか?
- 基礎的なモンテカルロ法の一つである受容-棄却法(accept-rejection法)について解説する

A-R法

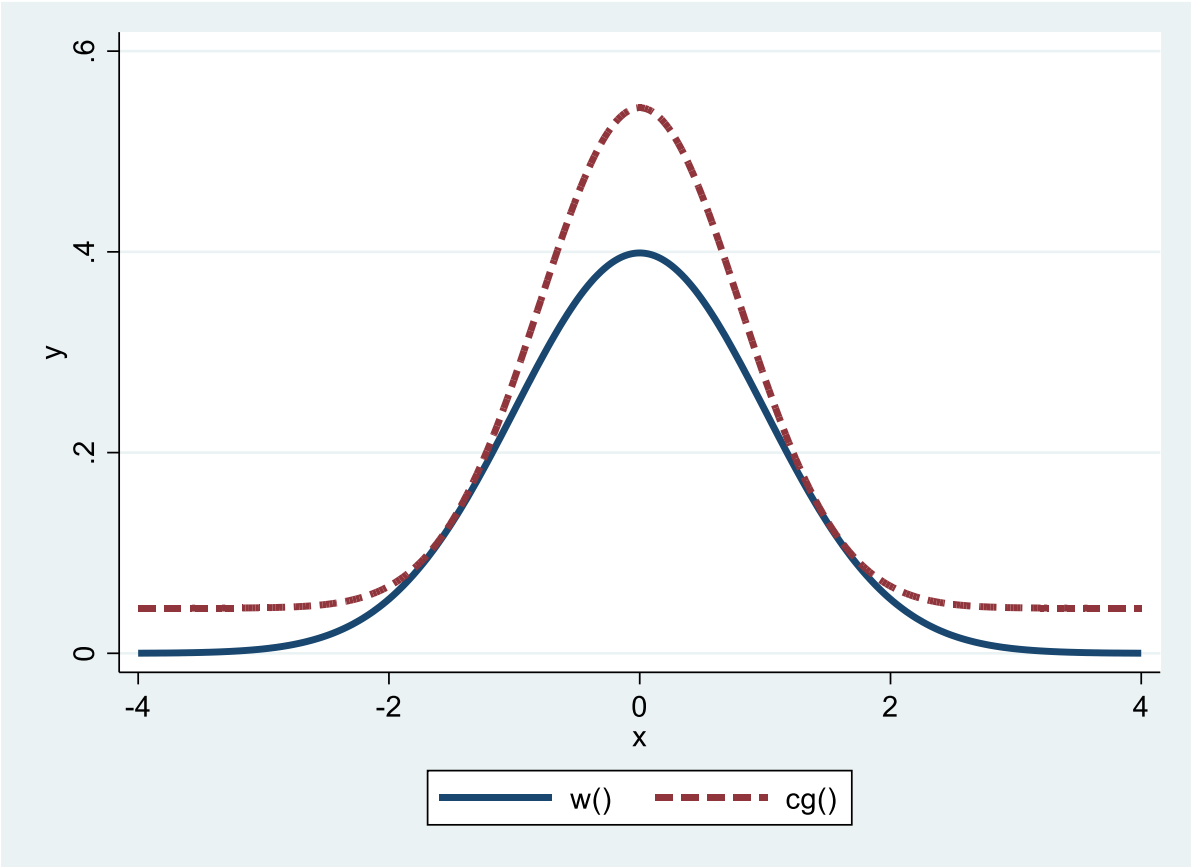
- A-R法は既存の関数g(θ)で作成する乱数の一部を棄却しながら、目的の関数w(θ)の乱数を作成する方法
- この時、次の条件が必要する

$$w(\theta) \leq cg(\theta)$$

- 次のスライドの図を作成するコマンド

```
. twoway function y=normalden(x),range(-4 4) || ///
                function y=normalden(x,0.8)+0.045,range(-4 4)
```

A-R法



$$w(\theta) \leq cg(\theta)$$

A-R法

- Step1:g(θ)から乱数をθ*を発生させる
- Step2:(0 1)区間上で一様乱数uを発生させる
- Step3:

$$u \leq \frac{w(\theta^*)}{cg(\theta^*)}$$ ならθ*をw(θ)の標本とする

$$u > \frac{w(\theta^*)}{cg(\theta^*)}$$ ならθ*を棄却し、Step1へ戻る

A-R法

- 受容したg()の乱数が、w()の乱数となる理由
  - ベイズの定理を利用する

$$w'(\theta|z) = \frac{w(\theta)p(z|\theta)}{\int w(\theta)p(z|\theta)d\theta}$$

  - 関数gを使って, zをu≤w(θ)/cg(θ)で置き換えると

$$g\left(\theta|u \le \frac{w(\theta)}{cg(\theta)}\right) = \frac{g(\theta)p\left(u \le \frac{w(\theta)}{cg(\theta)}|\theta\right)}{\int g(\theta)p\left(u \le \frac{w(\theta)}{cg(\theta)}|\theta\right)d\theta}$$

$$= \frac{g(\theta)\cdot\frac{w(\theta)}{cg(\theta)}}{\int g(\theta)\cdot\frac{w(\theta)}{cg(\theta)}d\theta} = \frac{w(\theta)/c}{\int \frac{w(\theta)}{c}d\theta} = w(\theta)$$

A-R法の注意点

- A-R法においてcg(・)を見つけるのは難しい
- g(・)のw(・)に対する近似が良くなければ、Step3における受容率が悪くなる

モンテカルロ積分

- モンテカルロ積分:ベイズ統計において事後平均や周辺事後確率密度を求めるときに、煩雑な積分計算を避けて、シミュレーションによって計算する方法
- 期待値

$$\hat{\theta} \equiv \widehat{\mathrm{E}(\theta|\mathrm{z})} = \frac{1}{N}\sum_{n=1}^{N}\theta^{(n)}$$

- 標準誤差

$$\widehat{NSE} = \frac{\hat{\sigma}}{\sqrt{N}}$$

モンテカルロ積分

- モンテカルロ積分の標準誤差の数値解は中心極限定理を利用して求める
- 標本はi.i.d.でその分散は有限とする
- 中心極限定理から

$$\sqrt{N}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N(0,\sigma^2)$$

ここで、$\sigma^2 \equiv V(\theta|z)$

- この分散はモンテカルロ積分を使って

$$V(\theta|z) \simeq (1/N)\sum_{n=1}^{N}\left(\theta^{(n)}-\hat{\theta}\right) = \hat{\sigma}^2$$

マルコフチェーンモンテカルロ

- マルコフチェーン:
  - 目的とする事後分布上に存在する値を連続的に作成して出来上がる1セットのグループ
  - 最新の値を作成するときに、依存関係があるのは、直近の値だけ(マルコフ過程)
- マルコフチェーンモンテカルロ:
  - 適切なシミュレーションで作成したデータセットは、特徴としてチェーンが長くなるほど定常過程に近づく

MCMC

- メトロポリスアルゴリズ
  - Metropolis and Ulam (1949)
  - Metropolis et al. (1953)
  - Hasting (1970)
  - 初期のアルゴリズムは提案分布として正規分布のみを利用する
  - Hasting (1970)は任意の提案分布を利用できるアルゴリズムを開発した

Metropolis-Hastingアルゴリズム

- 事前分布のパラメータの初期値をθ_0、データをzとする

$$\pi(\theta_0|z) > 0$$

- 最終的に作成するものはパラメータのチェーン

$$\{\theta_t\}_{t=0}^{T-1}$$

- Metropolis-Hastingアルゴリズムは次のスライドに示す4つのステップを繰り返し計算するもの

MHアルゴリズムのポイント

- 時間軸としてt=1,…,T-1を考える
- Step1:提案分布q()からある値(状態)を得る

$$\theta_* \sim q(\cdot|\theta_{t-1})$$

- Step2:受容確率αを計算する

$$\alpha(\theta_*|\theta_{t-1}) = \min\{r(\theta_*|\theta_{t-1}), 1\}$$

ここで

$$r(\theta_*|\theta_{t-1}) = \frac{\pi(\theta_*|z)q(\theta_{t-1}|\theta_*)}{\pi(\theta_{t-1}|z)q(\theta_*|\theta_{t-1})}$$

MHアルゴリズム

- Step3:一様分布の乱数を用いてuを得る

$$u \sim \mathrm{Uniform}(0, 1)$$

- Step4:uとαの大きさを比較して次の処理を行う

$$\theta_t = \theta_* \quad \text{if} \quad u \leq \alpha(\theta_* | \theta_{t-1}) \quad \text{受容}$$
$$\theta_t = \theta_{t-1} \quad \text{if} \quad \text{otherwise} \quad \text{棄却}$$

- MHアルゴリズムの1-4のステップを利用してパラメータを更新することを、特に「MH更新」と呼ぶ
- Step4の計算が終了したらStep1に戻る

受容確率

- 受容確率αがゼロに近い
  - ほとんどの提案θを棄却してしまう
  - 事後分布の領域を提案できない
- 受容確率αが1に近い
  - チェーンは事後分布の小領域にとどまり、こちらも探索的な調査にならない
- 作成したチェーンが小さな自己相関を持つような状態が好ましい
- Gelman, Gilks, and Roberts (1997)によれば多変量事後分布の場合0.234, 単変量事後分布の場合は0.45程度の受容確率が最適とされている

MCMCの制約と問題点

- チェーンを作成するための初期値の影響が大きい。特にチェーンが短い時は注意を要する
- 事後分布の端に存在する値を利用すると、適切な領域に到達するまで多くの計算回数を要する
- MCMCの繰り返し計算回数に明確な基準はない
- MCMCで作成した標本の自己相関は強くなる傾向があるので、その標本を利用した統計解析には注意が必要

数値例

- 第2回の最後に紹介した熊の出没に関する分析(今回も冒頭でグラフ化した)を、共役分布でなくMHアルゴリズで計算する場合の計算フローを示す
- Step1:尤度
  - ポアソン分布を使って尤度を計算する。過去5年間の平均5.6回からθ=5.6.このパラメータの下、シーズンに8回出没する尤度は

$$L(8|5.6) = \frac{\exp(-5.6)\cdot 5.6^8}{8!} = 0.0887$$

*Stataコマンド: display poissonp(5.6,8)

数値例

- Step2:
  - ガンマ分布を使って事前分布の密度を求める

$$g(5.6,1,0,8) = \frac{1^{5.6} \cdot 8^{5.6-1} \cdot \exp(-1\times 8)}{\Gamma(5.6)} = 0.0777$$

  - よって事後分布については

$$\pi(\theta_{current} = 5.6|8) \propto L(8|5.6) \times g(5.6,1,0,8)$$
$$= 0.0887 \times 0.0777 = 0.0068$$

  - ここで利用した5.6を $\theta_{current}$ とする

*Stataコマンド: display gammaden(5.6,1,0,8)

数値例

- Step3:提案分布の利用
  - 中心を5.6、そして標準偏差をσとする正規分布(提案分布)から無作為に新しいθ*proposal*を選択する

$$\theta_{proposal} = 4.6$$

  - 提案分布のσを「チューニングパラメータ」と呼ぶ
  - 新しいθを利用して同じ処理を繰り返す

尤度:poissonp(4.6,8)=0.0499
密度:gammaden(4.6,1,0,8)=0.0446

$$\pi(\theta_{proposal} = 4.6|8) \propto 0.0499 \times 0.0446 = 0.0022$$

数値例

- Step4:受容確率の計算

  最初に、

$$r(\theta_{proposal}|\theta_{current}) = \frac{0.0022}{0.0068} = 0.3235$$

  よって

$$\alpha(\theta_{proposal}|\theta_{current}) = \min\{0.3235, 1\}$$
$$= 0.3235$$

- U(0,1)から乱数を発生させて0.21が出たとすると$\theta_{proposal}$を受容する。そして提案分布から新たなθを得る

- 逆に乱数が0.77だとすると、$\theta_{proposal}$を棄却して$\theta_{current}$を保持して、Step3の提案分布の利用へ戻る

数値例

- 1万回の繰り返し計算の結果、5千個のθを受容したと考える
- 5千個のθの分布がこの例ではガンマ分布に従っており、その分布から計算した期待値や分散が合理的な数字になっていることが大切
- シミュレーション計算の結果、最後に得られた期待値や分散が合理的な値になっていない場合、計算の過程に不具合があると考えられる
- 第4回ではMCMCの計算終了後にチェックすべきポイントについて解説する

第3回のまとめ

- マルコフチェーンモンテカルロ(MCMC)は、ベイズ推定において事後分布をシミュレーションにより求める方法である
- MCMCを理解するために必要なモンテカルロ法の一つであるA-R法について説明した
- 代表的なマルコフチェーンモンテカルロ法であるメトロポリスヘイスティング(MH)アルゴリズムの基本的な考え方と、その仕組みを解説した
- MCMCにより、複雑な積分計算を避けて、事後分布の標本を得ることができる

参考文献

- J. Chan, G. Koop, D.J. Poirier and J. L. Tobias. *Bayesian Econometric Methods* 2nd edition. Cambridge University Press, 2020
- T M. Donovan and R M. Mickey. *Bayesian Statistics for Beginners*. Oxford University Press, 2019
- 小西 貞則 編 (2008). 計算統計学の方法. 朝倉書店

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。

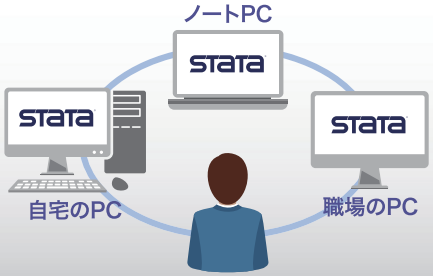
ライセンスはサブスクリプションがおすすめ！

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

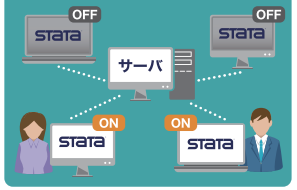
シングルライセンスを複数人でまとめてご購入いただく際のボリュームディスカウント商品です。

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



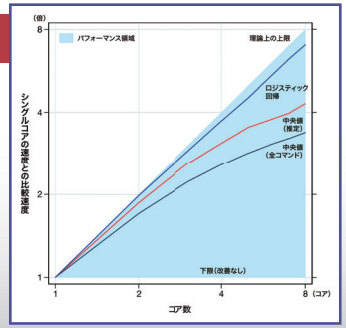
1台の仮想マシンにユーザがアクセスする運用
例:2ライセンスの場合

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/