

p 値は不要（？）

～ベイズ統計学の可能性について～

2019. 8. 9

古橋 武

1

ベイズ統計学がインターネット上で話題になっています。従来の統計的仮説検定の問題点が指摘され、実験データの評価をベイズ統計学により実行することが推奨され始めています。本チュートリアルに関連する参考文献リストを本稿の末尾に添付します。本講義の骨子は文献[6]～[9]に拠るところ大です。特に文献[6]は、私のベイズ統計学学習の出発点となりました。本講義は、前半 [1](#)～[4](#)では統計的仮説検定に対するベイズ統計学の違いに焦点を当てて解説し、後半 [5](#)～[7](#)ではベイズ統計手法の要点を解説します。正直に述べますと、本チュートリアルを実施することが決まってからベイズ統計学をにわか勉強しました。その過程で筆者自身が理解しづらかった点をできるだけ分かり易くすることを心がけて本講義ノートをまとめました。多くの方に読んでいただけたらと思います。公開します。

目次

1. [統計的仮説検定の流れ](#)
[統計的仮説検定では検定後にデータを追加してはいけない。](#)
2. [ベイズ統計の流れ](#)
[ベイズ統計ではデータを追加してよい。](#)
3. [統計的仮説検定における多重性の問題](#)
4. [ベイズ統計では母数の推定値に対して後から種々検討を加えてよい。](#)
5. [メトロポリス・ヘイスティングス法](#)

6. [ベイズの定理と母数推定](#)
7. [ハミルトニアンモンテカルロ法](#)

[オイラー法](#)

[\(x, p_m\)-空間内の移動と MH 法との関係](#)

[R のスクリプト](#)

[\(x, p_m\)-空間内の移動の様子](#)

[面積保存の性質](#)

[エネルギー保存則](#)

[リープフロッグ法](#)

[R のスクリプト](#)

8. [参考文献](#)

本稿中の R スクリプトを以下からダウンロードできます.

[本稿中の R スクリプトの圧縮ファイル掲載の Web ページ](#)

ご質問・ご意見を歓迎いたします.

furuhashi@nuee.nagoya-u.ac.jp

1. 統計的仮説検定の流れ

帰無仮説 $NH: \mu = 0$
 対立仮説 $AH: \mu \neq 0$

	サンプル	サンプルサイズ
x1	0.302	$(n = 9)$
x2	1.01	
x3	1.598	
x4	0.184	
x5	0.958	
x6	0.436	
x7	0.283	
x8	0.854	
x9	-1.127	

標本平均 $\bar{x} = 0.5$

検定統計量:
$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{v_e^2}{n}}} = 1.969 \quad (1.1)$$

p 値: $p = 0.084 > 0.05$

結論: 帰無仮説を棄却できない。

仮説駆動型

仮説を立てる.



データを採る.



検定統計量を求める.



データの出現確率を求める.



仮説の棄却/保留を決める.

2

まず、従来の統計的仮説検定の流れです。

統計的仮説検定では、最初に仮説を立てます。棄却することを前提に帰無仮説 NH (Null Hypothesis) を立て、帰無仮説を棄却できた場合に採択する仮説として対立仮説 AH (Alternative Hypothesis) を立てます。この例では、帰無仮説は母平均 $\mu = 0$ 、対立仮説は $\mu \neq 0$ です。

その後にデータを採録します。この例ではサンプルサイズ $n = 9$ です。標本平均 $\bar{x} = 0.5$ でした。

次は検定統計量を求めます。この例ではデータが正規分布に従うとして、母標準偏差 σ が未知の場合に、検定統計量 t

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{v_e^2}{n}}} \quad (1.1)$$

が自由度 $n - 1 = 8$ の t 分布に従うという理論を適用します。ここで、 μ は母平均、 v_e^2 は不偏分散です。

$$v_e^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.2)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.3)$$

帰無仮説より $\mu = 0$ です。 $t = 1.969$ でした。 この t 値の絶対値が 1.969 よりも大きな値となる \bar{x} の出現確率が両側検定の場合の p 値です。 t 分布の理論より $p = 0.084$ です。 検定の閾値を 0.05 とすると、帰無仮説 NH を棄却できないという結論が得られます。

統計的仮説検定の流れは、「仮説を立てる」→「データを採る」→「検定統計量を求める」→「データの出現確率を求める」→「仮説の棄却／保留を決める」です。 仮説を立てることが起点となるので、本講義では、統計的仮説検定は「仮説駆動型」として扱います。

統計的仮説検定の考え方

同時確率密度関数（例えば正規分布）

$$p(\mathbf{x}|\mu, \sigma) = \prod_{i=1}^n p(x_i|\mu, \sigma) \quad (1.4) \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$p(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - \mu)^2}{\sigma^2}\right\} \quad (1.5)$$

母平均 μ , 母標準偏差 σ などの**真値が存在する**.
これら真値を持つ同時確率密度関数にしたがって,
データ \mathbf{x} は生成される.

採録されたデータ \mathbf{x} が母集団から生成される確率を求める.

母数 μ, σ が定数, データ \mathbf{x} が確率変数

3

例えば正規分布の場合, サンプルサイズが n のとき, データ $\mathbf{x} = (x_1, x_1, \dots, x_n)^t$ の同時確率密度関数 $p(\mathbf{x}|\mu, \sigma)$ は

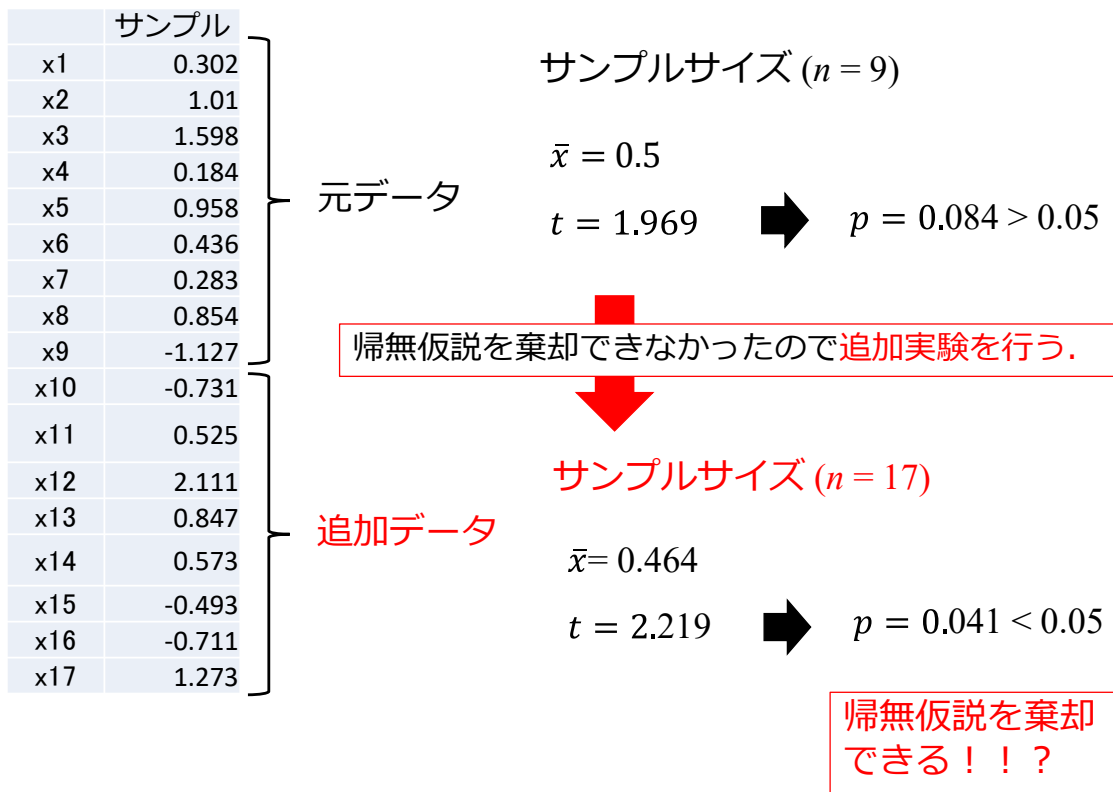
$$p(\mathbf{x}|\mu, \sigma) = \prod_{i=1}^n p(x_i|\mu, \sigma)$$

により与えられます. ここで, μ は母平均, σ は母標準偏差です.

μ , 母標準偏差 σ には真値が存在し, データ \mathbf{x} はこれら真値を持つ同時確率密度関数にしたがって生成されると考えます. そして, 採録されたデータ \mathbf{x} の出現確率を求めます.

統計的仮説検定では, 母数 μ, σ が定数, データ \mathbf{x} が確率変数です.

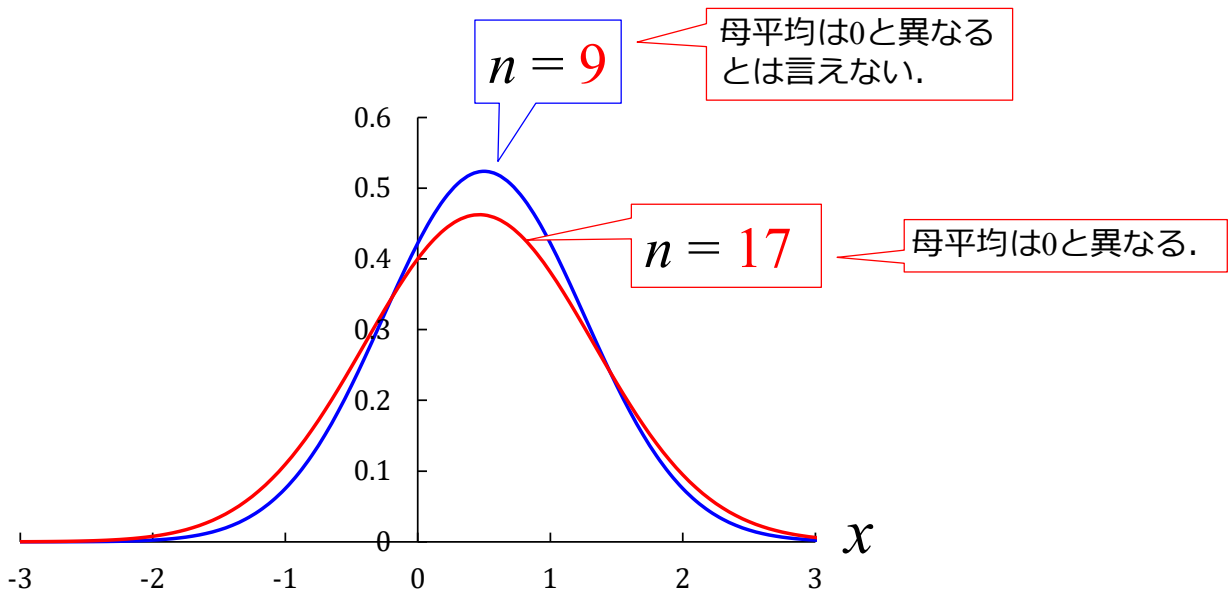
統計的仮説検定でやってしまいがちなこと



4

ここで、統計的仮説検定でやってしまいがちな事例を紹介します。

[スライド番号 2](#) の事例では帰無仮説を棄却することはできませんでした。このままでは論文が書けないので追加実験を実施することにしました。8 個の追加データを収録し、サンプルサイズ $n = 17$ としました。標本平均 $\bar{x} = 0.464$ 、検定統計量 $t = 2.219$ となり、 $p = 0.041$ と得られました。この p 値は閾値の 0.05 より小さくなったので、帰無仮説 NH を棄却できます!!!???



帰無仮説を棄却できなかったときは、サンプルサイズを大きくすれば
良い!!!?

➡ 本当???

5

サンプル平均 \bar{x} , 不偏分散 v_e^2 を用いて

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \bar{x})^2}{v_e^2}\right\} \quad (1.6)$$

のグラフを描きました。青線がスライド番号2のサンプルサイズ $n = 9$ の場合であり、赤線がスライド番号4の $n = 17$ の場合です。両者の形状にはほとんど違いが見られないにもかかわらず、サンプルサイズの違いで、一方は帰無仮説を棄却できず、母平均は 0 と異なるとは言えないという結論となり、もう一方は帰無仮説を棄却でき、母平均は 0 と異なるという結論となりました。

帰無仮説を棄却できなかったときは、サンプルサイズを大きくすればよいのでしょうか？

統計的仮説検定では

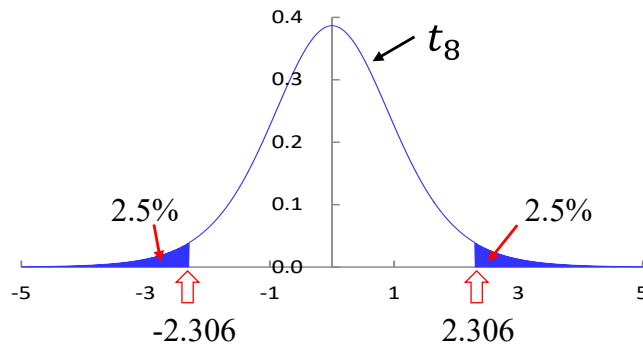
検定後にデータを追加してはいけません。

検定結果が有意でなかった場合、追加実験を実施しても有意な結果は出ない。

6

この問いかけに対する答えを先に述べると、「検定後にデータを追加してはいけません」です。検定結果が有意でなかった場合、追加実験を実施しても有意な結果は得られません。

9個のデータを集めた結果で検定を行い、有意差が出なかった場合に、さらに8個のデータを収集したときの第1種の過誤の確率は、追加実験をしないで17個のデータを収集した場合の第1種の過誤の確率とは異なる。



$|t| \geq 2.306$ であれば実験者は追加実験をやらない。追加実験をやらない場合の第1種の過誤の確率の最大値を P_{max} とすると

$$P_{max} = P(|t| \geq 2.306) = 0.05.$$

$|t| < 2.306$ のとき追加実験をやるので、追加実験をやった場合の第1種の過誤の確率は

$$P(\text{最初の実験} + \text{追加実験}) = P_{max} + P(\text{追加実験}) > 0.05 \quad (1.7)$$

となる。

7

その理由をこのスライドに記します。

9個のデータを集めた結果で検定を行い、有意差が出なかった場合に、さらに8個のデータを収集したときの第1種の過誤の確率は、追加実験をしないで17個のデータを収集した場合の第1種の過誤の確率とは異なります。

どう異なるのでしょうか？

図は自由度8の t 分布のグラフです。両側検定の場合、 $|t| \geq 2.306$ であれば、 p 値は5%以下となり、実験者は追加実験を実施しません。追加実験を実施しない場合の第1種の過誤の確率の最大値を P_{max} とすると、 $P_{max} = P(|t| \geq 2.306) = 0.05$ です。

のときに追加実験を実施するので、追加実験を実施した場合の第1種の過誤の確率は

$$P(\text{最初の実験} + \text{追加実験}) = P_{max} + P(\text{追加実験}) > 0.05$$

となります。

すなわち、追加実験をしない場合の第1種の過誤の確率に、追加実験前に閾値を下回らない場合でかつ追加実験によって閾値を下回る確率が加わるので、追加実験を実施した場合の第1種の過誤の確率は0.05より大きくなります。

統計的仮説検定では、仮説設定から p 値算出までの実験計画を事前に立てておき、データ採録後に仮説や統計モデルを変えることのないよう厳に戒められています。

ベイズ統計学に関する略年表

- 1763年 Thomas Bayesの死後, R. Priceによってベイズの定理が発表された.
- 1925年 R. A. Fisher, “Statistical Methods for Research Workers”^[1]にて
“The theory of inverse probability is founded upon an error, and must be wholly rejected.” (ベイズの定理は誤りの上に立脚しており, 完全に排除されなければならない.)
- 20世紀はベイズ統計学にとって冬の時代
- 1953年 N. Metropolis^[2]が物理化学領域で確率分布の生成法を提案した.
- 1970年 W. K. Hastings^[3]がMetropolisの手法が統計分野に応用できることに気づき, アルゴリズムの形式で整理した.
- 1987年 HMC (Hamiltonian Monte Carlo)法^[4]が提案された.
- 1990年 A. E. Gelfand & A. F. M. Smith^[5]がMCMC法をベイズ統計における困難な積分計算に適用する方法を提案した.
- 2012年 Stanの開発がGitHub上で始められた.
- 2014年 Biometrika (統計学における著名な学術雑誌) の過半数の論文がベイズ統計学を利用している (文献[6]序文) .

21世紀になりベイズ統計学に春がやってきた.

ベイズ統計学を利用したメールフィルタ, ベイズ画像処理などの応用が広がっている.

8

ベイズ統計学に関する略年表です. 文献[2]~[5]はネットから無料でダウンロードできます. なんと, R. Price のオリジナル原稿も無料で入手できます. ありがたい時代です.

ベイズの定理は 1763 年に発表されました. Thomas Bayes の死後に遺稿から, 親族の R. Price によって発見されたとのことです.

ベイズの定理は R. A. Fisher によって, “The theory of inverse probability is founded upon an error, and must be wholly rejected.” と厳しく批判されました. The theory of inverse probability がベイズの定理です. Fisher の“Statistical Methods for Research Workers”は, 参考文献[6]のまえがきによると, 「過去 100 年で, 統計的データ分析にもっとも影響を与えた教科書を敢えて 1 冊だけ選ぶとしたら, 多くの人が R.A.フィッシャーの手による「研究者のための統計的方法」をあげることに賛成するでしょう.’とのことです. 私はこの本の第 13 版(1958 刊)を大学の図書館から借りて当該文を確認しました.

20 世紀はベイズ統計学にとって冬の時代であったと言えます.

文献[6]によると, Fisher による厳しい批判はベイズ統計学が主観確率を利用する点に対してなされました. 事前分布に分析者の主観を反映できることは科学的客観性を損ないかねません. なお, 文献[6]は, 3.7 節にて, 同書をベイズ統計学の啓蒙書と位置づけ, 主観確率の利用には踏み込まず, 「最も拒否されにくいという意味で初学者に適した」無情報的事前分布を使用しています. そして, 無情報的事前分布には簡便な一様分布を使用しています.

ました。1953年にN. Metropolis^[2]が物理化学領域で確率分布の生成法を提案しました。1970年にW. K. Hastings^[3]が、Metropolisの手法が統計分野に応用できることに気づき、アルゴリズムの形式で整理しました。そして、1987年にHMC (Hamiltonian Monte Carlo)法^[4]が提案されました。

ベイズ統計学に対する直接的な貢献は、1990年にA. E. Gelfand & A. F. M. Smith^[5]によるベイズ統計の積分計算に対するMCMC(Markov Chain Monte Carlo)法の適用法の提案です。そして、2012年からStanの開発がGitHub上で始められました。Stanはフリーソフトであり、シミュレーションによりベイズ統計の事後確率分布に従う乱数を生成します。Stanは、R言語用のRStan、Python用のPyStanが開発されていて、いずれもネットから無料でダウンロードできます。これにより初学者でもベイズ統計を利用したデータ解析環境を容易に整えることができます。本稿はRStanを利用しています。

文献[6]によると、2014年におけるBiometrika(統計学における著名な学術雑誌)の過半数の論文がベイズ統計学を利用しているとのことです。

21世紀になりベイズ統計学に春がやってきました。

以下は文献リストです。

References

- [1] R. A. Fisher, “Statistical Method for Research Workers 13th edition,” Oliver and Boyd, 1958
- [2] N. Metropolis, et al., “Equation of State Calculations by Fast Computing Machines,” J. of Chemical Physics, Vol. 21, No. 6, pp.1087-1092, 1953
- [3] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” Biometrika, Vol.57, No.1, pp.97-109, 1970
- [4] S. Duane, et al., “Hybrid Monte Carlo,” Physics Letters B, Vol. 195, No. 2, pp.216-222, 1987
- [5] A. E. Gelfand and A. F. M. Smith, “Sampling-Based Approaches to Calculating Marginal Densities,” J. of American Statistical Association, Vol.85, No. 410, pp.398-409, 1990
- [6] 豊田秀樹「基礎からのベイズ統計学 —ハミルトニアンモンテカルロ法による実践的入門—」朝倉書店, 2015
- [7] 豊田秀樹「はじめての統計データ分析 —ベイズ的<ポスト p 値時代>の統計学—」朝倉書店, 2016
- [8] 豊田秀樹「心理統計法 —有意性検定からの脱却—」放送大学教育振興会, 2017
- [9] 奥村晴彦「Rで楽しむベイズ統計入門 —しくみから理解するベイズ推定の基礎—」技術評論社, 2018
- [10] John K. Kruschke, “Doing Bayesian Data Analysis 2nd edition,” Elsevier, 2015
- [11] A. Gelman, et al., “Bayesian Data Analysis 3rd edition,” CRC Press, 2014

2. ベイズ統計の流れ

	サンプル	サンプルサイズ
x1	0.302	(n = 9)
x2	1.01	
x3	1.598	
x4	0.184	
x5	0.958	
x6	0.436	
x7	0.283	
x8	0.854	
x9	-1.127	

標本平均 $\bar{x} = 0.5$



データを採る.



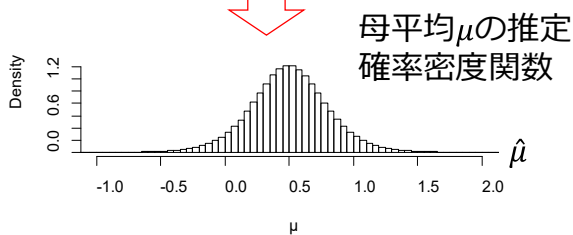
母数の確率分布を求める.



生成量を評価する.

生成量: 母数の乱数から数値計算により求められる量

数値計算



母平均の期待値: $\hat{\mu}_{EAP} = 0.5$
 母平均の標準偏差: $\sigma_{\hat{\mu}} = 0.322$
 母平均の95%確信区間: $[-0.143, 1.141]$

$$P(\hat{\mu} < 0) = 0.054$$

9

ベイズ統計の流れです.

ベイズ統計はデータの採録から始まります. そして, このデータを基に数値計算により, 例えば母平均 μ の推定値 $\hat{\mu}$ の乱数を得ます. 数値計算に際してサンプル x_i ($i = 1, 2, \dots, 9$) は, この例では正規分布に従う母集団から得られたとしています. 母平均推定値 $\hat{\mu}$ は, 母平均の推定確率分布に従う乱数です. 図のヒストグラムは, 面積を1に規格化すれば, 推定確率密度関数の近似です.

得られた母平均推定値の乱数から, 数値計算により母平均の期待値, 母平均の標準偏差, 母平均の95%確信区間, さらには母平均の推定値 $\hat{\mu}$ が0より小さい確率 $P(\hat{\mu} < 0)$ などを求めることができます.

統計的仮説検定では, 帰無仮説 ([スライド番号2](#)では $\mu = 0$) の下で, 標本平均 $|\bar{x}| > 0.5$ の出現確率 P を求めました. 母平均の**評価が回りくどい**です.

$\hat{\mu}$ が0より小さい確率 $P(\hat{\mu} < 0)$ を得られます. 母平均の**評価が直截的**です.

なお, 母数の乱数から数値計算により求められる量は**生成量**と呼ばれます.

ベイズ統計の流れは, 「データを採る.」→「母数の確率分布を求める.」→「生成量を評価する.」です. データの採録が起点となるので, 本講義では, ベイズ統計は「**データ駆動型**」であるとします.

ベイズ統計の考え方

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

手元にあるデータ \boldsymbol{x} は、例えば、母平均 μ 、母標準偏差 σ の正規分布に従う母集団から得られたとする。この μ, σ を推定対象とする。

$$p(\mu, \sigma | \boldsymbol{x}) = ?$$

データ \boldsymbol{x} が定数、母数 μ 、 σ が確率変数

10

ベイズ統計では、手元にあるデータ \boldsymbol{x} は、例えば、母平均 μ 、母標準偏差 σ の正規分布に従う母集団から得られたとします。そして、この μ, σ を推定対象とします。

データ \boldsymbol{x} をもとに μ, σ の条件付き確率密度関数 $p(\mu, \sigma | \boldsymbol{x})$ を求めます。

ベイズ統計では、データ \boldsymbol{x} が定数、母数 μ 、 σ が確率変数です。

ベイズの定理

$$\begin{aligned} p(\mu, \sigma | \mathbf{x}) &= \frac{p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma)}{P(\mathbf{x})} \\ &= \frac{p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma)}{\iint_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma} \end{aligned} \quad (2.1)$$

$p(\mu, \sigma | \mathbf{x})$: 事後確率密度関数
 $p(\mathbf{x} | \mu, \sigma)$: 尤度
 $p(\mu, \sigma)$: 事前確率密度関数

データ \mathbf{x} が得られた条件下での母数 μ, σ の確率密度関数

$$P(\mathbf{x}) = \iint_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma: \text{周辺確率} \quad (2.2)$$

11

ベイズの定理です。

$p(\mu, \sigma | \mathbf{x})$ が事後確率密度関数、 $p(\mathbf{x} | \mu, \sigma)$ が尤度、 $p(\mu, \sigma)$ が事前確率密度関数、 $P(\mathbf{x})$ が周辺確率です。
 $p(\mu, \sigma | \mathbf{x})$ は、データ \mathbf{x} が得られた条件下での母数 μ, σ の確率密度関数です。 $p(\mathbf{x} | \mu, \sigma)$ は、母数 μ, σ の条件下でのデータ \mathbf{x} の確率密度関数です。 $p(\mu, \sigma)$ は、母数 μ, σ の確率密度関数です。 $P(\mathbf{x})$ は、 \mathbf{x} の生起確率です。

ベイズの定理により、尤度、事前確率と周辺確率から事後確率を求めることができます。

コンピュータ技術の進歩により，数値計算により事後確率密度関数に従う乱数を，誰でも簡単に生成できるようになった。

$$p(\mu, \sigma | \mathbf{x}) = \frac{p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma)}{\underbrace{\iint_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma}} \quad (2.3)$$

MCMC

(Markov Chain Monte Carlo method)

MH (Metropolis-Hastings method)

HMC (Hamiltonian MC method)



インターネットより無料で
ツールをダウンロード可能

μ, σ の確率密度関数に従う乱数を，例えば，それぞれ100,000個生成できれば，**近似的に μ, σ の確率密度関数を知ることができる。**

12

ここでは概要だけ述べます。後に詳しく解説します。

コンピュータ技術の進歩により，数値計算により事後確率密度関数に従う乱数を，誰でも簡単に生成できるようになりました。上式の右辺の計算を数値計算により実行するのですが，MH法は分母の積分計算を不要にする巧妙な手法です。そして，HMC法は，MH法の収束速度を飛躍的に高めた手法です。多変数の母数の乱数生成を効率よく行うことができます。MH法，HMC法は，いずれも，MCMC (Markov Chain Monte Carlo)法の一種です。HMC法を実装した Stan はインターネットより無料でダウンロード可能です。

μ, σ の確率密度関数に従う乱数を，例えば，それぞれ 100,000 個生成すれば，近似的に μ, σ の確率密度関数を知ることができます。

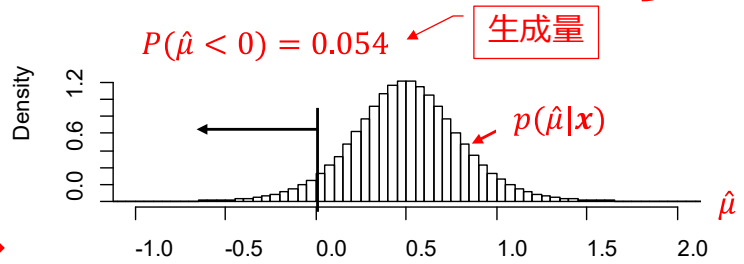
ベイズ統計の具体例

($n = 9$)

	Samples
x1	0.302
x2	1.01
x3	1.598
x4	0.184
x5	0.958
x6	0.436
x7	0.283
x8	0.854
x9	-1.127

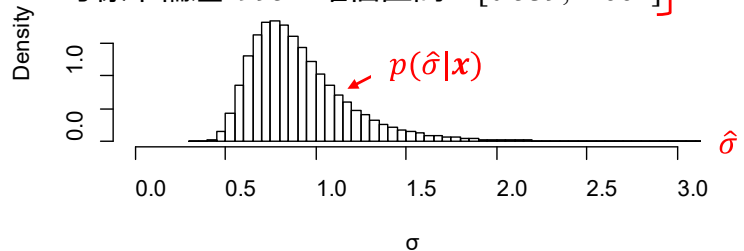
母平均の期待値： $\hat{\mu}_{EAP} = 0.5$
 母平均の標準偏差： $\sigma_{\hat{\mu}} = 0.322$
 母平均の95%確信区間： $[-0.143, 1.141]$

生成量



母標準偏差の期待値： $\hat{\sigma}_{EAP} = 0.917$
 母標準偏差の標準偏差： $\sigma_{\hat{\sigma}} = 0.3$
 母標準偏差の95%確信区間： $[0.539, 1.661]$

生成量



13

ベイズ統計の具体例です。 [スライド番号 2](#) のサンプルと同じデータが得られたとします。このデータを用いて、RStanにより母平均と母標準偏差の推定確率密度関数を求めた結果です。右上が母平均推定値 $\hat{\mu}$ のヒストグラムです。横軸が母平均推定値 $\hat{\mu}$ です。縦軸は密度です。ヒストグラムの面積が1となるように規格化されています。従って、このヒストグラムは確率密度関数の近似です。右下が母標準偏差の推定値 $\hat{\sigma}$ のヒストグラムです。同様にこのヒストグラムは確率密度関数の近似です。

これらの乱数より各種生成量を数値計算により得られます。母平均推定値 $\hat{\mu}$ の期待値 $\bar{\mu}$ は Expected A Priori のイニシャルを取って $\hat{\mu}_{EAP}$ と記してあります。 $\hat{\mu}_{EAP}$ は次式により得られます。すなわち

$$\hat{\mu}_{EAP} = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_i \quad (2.4)$$

です。 N は乱数の数です。平均値の標準偏差 $\sigma_{\hat{\mu}}$ は

$$\sigma_{\hat{\mu}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\mu}_i - \hat{\mu}_{EAP})^2} \quad (2.5)$$

により求められます。ここで、 $\frac{1}{N} \sum_{i=1}^N (\hat{\mu}_i - \hat{\mu}_{EAP})^2$ は標本分散に相当します。分散に真値が存在し、その推定値を求める立場では $\frac{1}{N-1} \sum_{i=1}^N (\hat{\mu}_i - \hat{\mu}_{EAP})^2$ の不偏分散を用います。ベイズ統計では、母平均 μ は確率変数であるとして、真値の存在を仮定していません。そこで、真値の推定値である不偏分散は使いません。

母平均 μ の 95%確信区間は

$$[\hat{\mu}_{low}, \hat{\mu}_{high}]$$

です。ただし、

$$P(\hat{\mu} \leq \hat{\mu}_{low}) = 0.025, \quad P(\hat{\mu} \geq \hat{\mu}_{high}) = 0.025$$

です。

統計的仮説検定には 95%信頼区間があります。95%確信区間とでは、両者はどう違うのでしょうか？

95%確信区間は、上述の求め方の通り、母数の推定乱数の 95%がこの区間にあります。

95%信頼区間について、その求め方をおさらいしてみましょう。データ $y_i (i = 1, 2, \dots, N)$ があり、これらが母平均 μ 、母分散 σ^2 の正規分布 ($N(\mu, \sigma^2)$ と表記します。) に従って生成されたとします。母分散 σ^2 は未知として、[スライド番号 2](#) の(1.1)式より、検定統計量 t を

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{v_e^2}{N}}}$$

とします。自由度 $N-1$ の t 分布において

$$P(|t| \geq t_\alpha) = 0.05$$

となる t_α を用いて、

$$\bar{x} - t_\alpha \sqrt{\frac{v_e^2}{N}} \leq \mu \leq \bar{x} + t_\alpha \sqrt{\frac{v_e^2}{N}}$$

と変形します。95%信頼区間は

$$\left[\bar{x} - t_\alpha \sqrt{\frac{v_e^2}{N}}, \quad \bar{x} + t_\alpha \sqrt{\frac{v_e^2}{N}} \right]$$

と定義されます。すなわち、 N 個の標本抽出実験を M 回実施したときに、この区間が母平均 μ を含む回数が $0.95 \times M$ であることを意味します。

ベイズ統計では母数は確率変数であり、唯一の真値は存在しないという立場です。一方、統計的仮説検定では母数は定数であり、唯一の真値が存在するという立場です。この立場の違いによる区間の意味の違いが混同されないようにするために、ベイズ統計では「確信区間」という新しい言葉が使われています。

ベイズ統計学では

データを追加してよい.

なぜならば、**仮説がなく**、追加前データに対して棄却／保留の判定をしていないので、**場合分けが生じない**。データ収集方法の違い（① 9個を集めた段階で一旦解析し、さらに8個集めて再度解析をする。② 途中段階の解析をしないで、17個を集めた後に解析する。）は μ, σ の確率密度関数に影響しない。

14

さて、[スライド番号6](#)にて、統計的仮説検定では検定後にデータを追加してはいけないと述べ、[スライド番号7](#)にてその理由を解説しました。

では、ベイズ統計学ではどうでしょうか？

ベイズ統計学ではデータを追加してよいです。

なぜならば、**仮説がなく**、追加前データに対して棄却／保留の判定をしていないので、**場合分けが生じない**からです。従って、データ収集方法の違い（① 9個を集めた段階で一旦解析し、さらに8個集めて再度解析をする。② 途中段階の解析をしないで、17個を集めた後に解析する。）は μ, σ の確率密度関数に影響しません。

ベイズ統計の流れ

サンプル	サンプルサイズ
x1	0.302
x2	1.01
x3	1.598
x4	0.184
x5	0.958
x6	0.436
x7	0.283
x8	0.854
x9	-1.127

サンプルサイズ $(n = 9)$

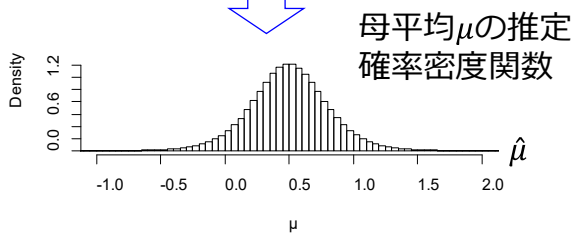
標本平均 $\bar{x} = 0.5$



データを採る.

仮説がない

数値計算



母数の確率分布を求める.

仮説がないので、
棄却/保留の場合
分けがない

母平均の期待値： $\hat{\mu}_{EAP} = 0.5$
母平均の標準偏差： $\sigma_{\hat{\mu}} = 0.322$
母平均の95%確信区間： $[-0.143, 1.141]$

生成量を評価する.

生成量：母数の乱数から数値計算
により求められる量

$$P(\hat{\mu} < 0) = 0.054$$

15

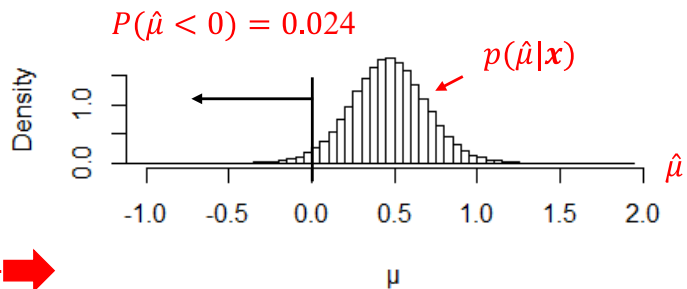
スライド番号 9 を再掲します。ベイズ統計学では、仮説がないこと、仮説がないので仮説の棄却/保留の場合分けがないことを、フローに吹き出しを付して記してあります。

データの追加 $n = 17$ の結果

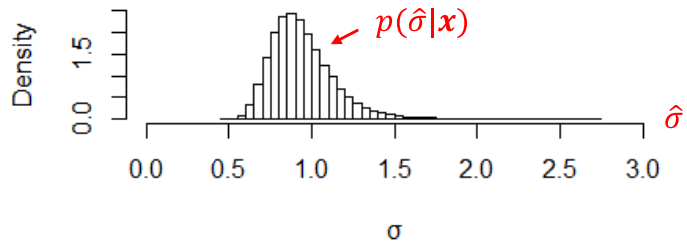
	Samples
x1	0.302
x2	1.01
x3	1.598
x4	0.184
x5	0.958
x6	0.436
x7	0.283
x8	0.854
x9	-1.127
x10	-0.731
x11	0.525
x12	2.111
x13	0.847
x14	0.573
x15	-0.493
x16	-0.711
x17	1.273

追加
データ

母平均の期待値 : $\hat{\mu}_{EAP} = 0.464$
 母平均の標準偏差 : $\sigma_{\hat{\mu}} = 0.232$
 母平均の95%確信区間 : $[-0.004, 0.925]$



母標準偏差の期待値 : $\hat{\sigma}_{EAP} = 0.939$
 母標準偏差の標準偏差 : $\sigma_{\hat{\sigma}} = 0.185$
 母標準偏差の95%確信区間 : $[0.659, 1.378]$

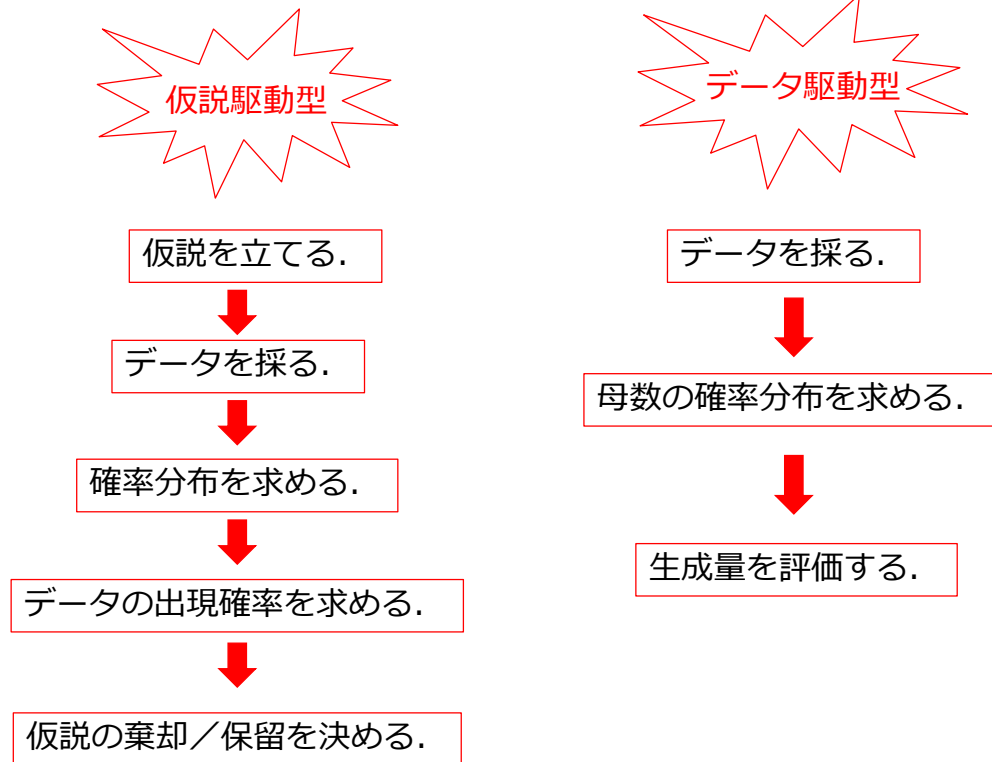


16

データを8個追加して、全体でサンプルサイズを17とした場合です。母平均の推定値 $\hat{\mu}$ 、母標準偏差の推定値 $\hat{\sigma}$ のいずれも、サンプルサイズ9の場合と比較して、それぞれの標準偏差が小さくなっています。そして、各確信区間が狭くなっています。

統計的仮説検定

ベイズ統計



17

以上、統計的仮説検定とベイズ統計学との違いを解説してきました。両者の流れをこのスライドにまとめます。統計的仮説検定は仮説駆動型であること、ベイズ統計学はデータ駆動型であることを述べました。この手法の違いによる両者の特徴を以下にまとめます。

(1) 結論の出し方について

統計的仮説検定では、帰無仮説の棄却／保留の判定をするため、結論が明確です。棄却できれば論文が書けて、保留となれば、実験計画から全てやり直しです。追加実験は厳禁です。

ベイズ統計学では、生成量の評価結果を得ます。この結果から例えば提案法を採用する決断は利用者に委ねられます。仮説がないため、帰無仮説の棄却／保留の判定がありません。追加データを得た場合には、追加データを含めて母数の推定値・生成量を求めて、利用者に再提示できます。

(2) 母数の評価について

統計的仮説検定では、帰無仮説の下で標本平均などの出現確率を求めました。母数の評価は回りくどいです。

一方、ベイズ統計では、例えば、母平均の推定値 $\hat{\mu}$ が 0 より小さい確率 $P(\hat{\mu} < 0)$ を得られます。この母平均の評価は直截的です。母数の推定乱数からは種々の生成量を数値計算により得られます。仮説がないため、帰無仮説の棄却／保留の場合分けがありません。後述の多重比較の例のように母数に対するあの手この手の評価が許容されます。

3. 統計的仮説検定における多重性の問題

検定回数を増やせば、その分
閾値を厳しくしなければなら
ない。

検定のやり方によっては閾値調整の理論を自分で構
築しなければならない。

18

統計的仮説検定においてしばしば多重性の問題に出会います。多重性の問題とは、検定を繰り返すと第1種の過誤の確率が、個々の検定の閾値よりも大きくなってしまう問題です。

このため、検定回数を増やせば、その分、帰無仮説棄却の閾値を厳しくしなければなりません。検定のやり方によっては閾値調整の理論を自分で構築しなければなりません。

統計的仮説検定：多重比較

2群について検定を実施する。

帰無仮説 $NH_i: \mu_i = \mu_{i0}$

対立仮説 $AH_i: \mu_i \neq \mu_{i0} \quad i = 1, 2$

$$\mu_{10} = 3.25, \mu_{20} = 3.4$$

$x1 \leftarrow c(3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.5)$

$x2 \leftarrow c(3.9, 3.5, 3.7, 3.8, 4.5, 3.4, 3.4, 3.5)$

検定統計量: $t_i = \frac{\bar{x}_i - \mu_{i0}}{\sqrt{v_{ei}^2/n_i}} : t$ 検定

$$t_1 = 1.8 \rightarrow p_1 = 0.109$$

$$t_2 = 2.4 \rightarrow p_2 = 0.047 < 0.05$$

NH_2 を棄却できると言えるか？



仮説を立てる。



データを採る。



確率分布を求める。



データの出現確率を
求める。



仮説の棄却/保留を決める。

19

多重比較の簡単な例を紹介します。

2群について検定を実施することとします。

帰無仮説 $NH_i: \mu_i = \mu_{i0}$

対立仮説 $AH_i: \mu_i \neq \mu_{i0} \quad i = 1, 2$

です。ここで、 $\mu_{10} = 3.25, \mu_{20} = 3.4$ とします。そして、

$X_1 = \{3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.5\}$

$X_2 = \{3.9, 3.5, 3.7, 3.8, 4.5, 3.4, 3.4, 3.5\}$

の2群のデータが得られたとします。サンプルサイズはそれぞれ $n_1 = 9, n_2 = 8$ です。

検定統計量は

$$t_i = \frac{\bar{x}_i - \mu_{i0}}{\sqrt{v_{ei}^2/n_i}} \quad i = 1, 2$$

です。 v_{e1}^2, v_{e2}^2 それぞれ1, 2群の不偏分散です。上記の数値例を代入すると

$$t_1 = 1.8 \quad p_1 = 0.109$$

$$t_2 = 2.4 \quad p_2 = 0.047$$

と得られました。ここで、 $p_2 = 0.047 < 0.05$ なので、 NH_2 を棄却できると言えるでしょうか？

多重性の問題

$$NH_1: p_1 = 0.109$$

$$NH_2: p_2 = 0.047$$

NH_1, NH_2 のどちらか、もしくは両方が棄却される確率は？

$$\begin{aligned} P &= 1 - (1 - 0.109)(1 - 0.047) \\ &= 0.151 > 0.05 \end{aligned} \quad (3.1)$$

FWER(Familywise Error Rate)と呼ばれる。

$$FWER = P\{V \geq 1\} \quad (3.2)$$

V : 複数の仮説検定において、第1種の過誤が起きる回数

複数の仮説検定を繰り返したとき、仮説群の中で1回以上有意差が出る確率 $FWER$ が個々の検定の閾値($\alpha = 0.05$)より大きくなってしまう問題を**多重性の問題**という。

20

多重性の問題の確認です。各帰無仮説におけるデータ X_i の p 値は

$$NH_1: p_1 = 0.109$$

$$NH_2: p_2 = 0.047$$

です。そこで、 NH_1, NH_2 のどちらか、もしくは両方が棄却される確率は

$$P = 1 - (1 - 0.109)(1 - 0.047) = 0.151 > 0.05$$

です。この確率 P は FWER (Familywise Error Rate) と呼ばれます。FWER は

$$FWER = P\{V \geq 1\}$$

と表されます。ここで、 V は複数の仮説検定において、第1種の過誤が起きる回数です。すなわち、FWER は、複数の検定を実施したときに、帰無仮説のファミリー（この例では NH_1 と NH_2 ）において1回以上第1種の過誤が起きる確率です。

複数の仮説検定を繰り返したとき、FWER が個々の検定の閾値($\alpha = 0.05$)より大きくなってしまう問題を**多重性の問題**といいます。

個々の検定を有意水準 5%（第1種の過誤が起きる確率 5%）で実施したとして、100回の検定を繰り返せば、全ての帰無仮説が正しかったとしても、100回中5回程は帰無仮説が棄却されてしまいます。

FWER は

$$P = 1 - (1 - 0.05)^{100} = 0.994$$

です。ほぼ確実に100回中のどこかで第1種の過誤を犯してしまいます。

多重比較の方法

$FWER \leq 0.05$ となるように、個別の検定の閾値（名義水準と呼ばれる。）を調整する。名義水準を α' とする。

a : データ群数

$$FWER = 1 - (1 - \alpha')^a \quad (3.3)$$

$$\alpha' = 1 - (1 - FWER)^{\frac{1}{a}} \quad \leftarrow \text{Sidakの方法} \quad (3.4)$$

$$\begin{aligned} \alpha' &= 1 - (1 - 0.05)^{\frac{1}{2}} \\ &= 0.0253 \end{aligned}$$

厳密

$$\alpha' = \frac{FWER}{a} \quad \leftarrow \text{Bonferroniの方法} \quad (3.5)$$

$$\begin{aligned} \alpha' &= \frac{0.05}{2} \\ &= 0.025 < 0.0253 \end{aligned}$$

簡便：身近に計算機がなかった頃に重宝した。
Sidakより厳しい・

21

そこで、多重比較の方法が提案されてきました。 X_1, X_2 の試行が独立の場合は簡単です。 $FWER \leq 0.05$ となるように、個別の検定の閾値（名義水準と呼ばれます。）を調整します。名義水準を α' とすると

$$FWER = 1 - (1 - \alpha')^a$$

です。ここで、 a はデータ群数です。この式を α' について求めると

$$\alpha' = 1 - (1 - FWER)^{\frac{1}{a}}$$

と得られます。この式による名義水準の調整法が Sidak の方法と呼ばれます。2群($a=2$)の場合

$$\alpha' = 1 - (1 - 0.05)^{\frac{1}{2}} = 0.0253$$

と求められます。1/a乗の計算は計算機がないと困難です。

簡便な方法に

$$\alpha' = \frac{FWER}{a}$$

があります。これは Bonferroni の方法と呼ばれます。2群($a=2$)の場合

$$\begin{aligned} \alpha' &= \frac{0.05}{2} \\ &= 0.025 \end{aligned}$$

と求まります。Sidak の方法は厳密な方法です。Bonferroni の方法は Sidak の方法より名義水準を少し厳

しくします。身近に計算機がなかった頃に重宝した方法ですが、今でも広く使われている名義水準調整法です。

先の問題にBonferroniの方法を適用すると

$$\begin{aligned} \text{帰無仮説 } & NH_i: \mu_i = \mu_{i0} \\ \text{対立仮説 } & AH_i: \mu_i \neq \mu_{i0} \end{aligned} \quad i = 1, 2$$
$$\mu_{10} = 3.25, \mu_{20} = 3.4$$

x1 <- c(3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.5)
x2 <- c(3.9, 3.5, 3.7, 3.8, 4.5, 3.4, 3.4, 3.5)

検定統計量: $t_i = \frac{\bar{x}_i - \mu_{i0}}{\sqrt{v_{\hat{\epsilon}}^2/n_i}}$

名義水準

$$t_1 = 1.8 \rightarrow p_1 = 0.109 > \alpha' = 0.025$$

$$t_2 = 2.4 \rightarrow p_2 = 0.047 > \alpha' = 0.025$$

NH₁, NH₂ ともに棄却できない。

22

スライド番号 19 の検定問題に Bonferroni の方法を適用すると、個々の検定の名義水準 $\alpha' = 0.025$ となり、NH₁, NH₂ ともに棄却できません。

ここで、ふと、2群の母平均の差が気になったので、**反則を承知**で追加の検定を行ってみた。

$$\begin{aligned} \text{NH}_3: \mu_1 &= \mu_2 \\ \text{AH}_3: \mu_1 &\neq \mu_2 \end{aligned}$$

$$\text{検定統計量: } t_3 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{v_{e1}^2}{n_1} + \frac{v_{e2}^2}{n_2}\right)}} \quad (3.6)$$

$$t_3 = 2.95$$

$$p_3 = 0.0184 > \alpha'' = \frac{0.05}{3} = 0.0167$$

NH₃ を棄却できない。

Bonferroni の方法
を適用



仮説を立て直す。



確率分布を求める。



データの出現確率を
求める。



仮説の棄却/保留を決
める。

23

ここで、ふと、2群の母平均の差が気になって、追加の検定を行ってみることとしました。(帰無仮説保留の判定後に帰無仮説を追加してはいけないのですが、反則を承知で実施してみたという設定です。)

追加の帰無仮説は

$$\text{NH}_3: \mu_1 = \mu_2$$

$$\text{AH}_3: \mu_1 \neq \mu_2$$

です。ウェルチの t 検定を適用すると、検定統計量は

$$t_3 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{v_{e1}^2}{n_1} + \frac{v_{e2}^2}{n_2}\right)}}$$

となります。スライド番号 22 の具体的数値を代入すると、等価自由度は 8 となりました。この検定統計量 t_3 は自由度 8 の t 分布に従います。

$$t_3 = 2.95$$

$$p_3 = 0.0184 > \alpha'' = \frac{0.05}{3} = 0.0167$$

であり、 p 値 ($p_3 = 0.0184$) は 3 群の場合の Bonferroni の方法の名義水準 $\alpha'' = 0.0167$ よりわずかに大きな値となりました。NH₃ を棄却できない結果となりました。

禁じ手

なまじ, NH_1, NH_2 の検定を一緒にやってしまったばかりに, NH_3 を棄却できなくなりました。

NH_1, NH_2 の検定は無かったことにしよう!!!

ダメ!

実験計画からやり直すことにする。その際、閾値を緩められないか？

$NH_1: \mu_1 = \mu_{10}$	←	独立ではない
$NH_2: \mu_1 = \mu_{20}$	←	独立ではない
$NH_3: \mu_1 = \mu_2$	←	独立ではない

各群が互いに独立であることを前提としたBonferroniの方法では閾値が厳しすぎる。

NH_1 と NH_3 が独立でないこと, NH_2 と NH_3 が独立でないことを考慮すると, それぞれの閾値をボンフェローニの方法よりも緩和できる。

しかし, この理論は見当たらないので, 自分で構築しよう!!!

24

なまじ, NH_1, NH_2 の検定を一緒にやってしまったばかりに, NH_3 を棄却できなくなりました。
 NH_1, NH_2 の検定は無かったことにする・・・やってはいけません。

そこで, 実験計画, データ採録からやり直すことにします。 NH_3 を再び棄却できない結果となることが予想されるので, 閾値を緩められないか考えてみましょう。

各群が互いに独立であることを前提としたボンフェローニの方法では閾値が厳し過ぎます。 NH_1 と NH_3 が独立でないこと, NH_2 と NH_3 が独立でないことを考慮すると, それぞれの閾値をボンフェローニの方法よりも緩和できます。しかし, この理論は見当たりません。自分で構築する必要があります。Tukeyの方法の導出過程を理解して, この場合に適用できるように改変して・・・

4. ベイズ統計では

母数の推定値（乱数）に対して種々検討を加えてよい。

結論を採択した場合の過誤の確率等は、いかような組み合わせでも、数値計算により容易に求められる。

25

ベイズ統計学では、母数の推定値（乱数）に対して種々検討を加えてよいです。結論を採択した場合の過誤の確率等は、いかような組み合わせでも、数値計算により容易に求められます。

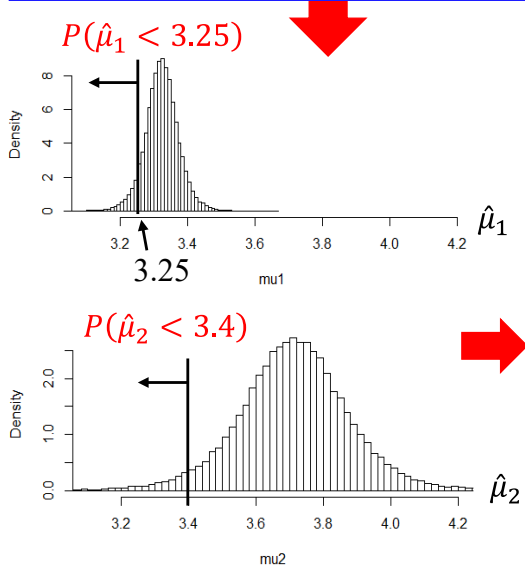
ベイズ統計では

```
x1 <- c(3.3, 3.2, 3.4, 3.3, 3.3, 3.2, 3.5, 3.2, 3.5)
x2 <- c(3.9, 3.5, 3.7, 3.8, 4.5, 3.4, 3.4, 3.5)
```



データを採る.

HMC法により母平均 μ_1, μ_2 の確率密度関数推定値 (乱数) を算出



母平均 $\hat{\mu}_1$ が 3.25 より小さい確率
 $P(\hat{\mu}_1 < 3.25) = 0.067$
母平均 $\hat{\mu}_2$ が 3.4 より小さい確率
 $P(\hat{\mu}_2 < 3.4) = 0.035$
 $P(\hat{\mu}_1 \leq 3.25 \text{ or } \hat{\mu}_2 \leq 3.4) = 0.099$

母数の確率分布を求める.

生成量を評価する.

ベイズ統計では、データを基に HMC 法により、例えば母平均 μ_1, μ_2 の確率密度関数推定値 (乱数) $\hat{\mu}_1, \hat{\mu}_2$ を算出します. 図は、上が $\hat{\mu}_1$ 、下が $\hat{\mu}_2$ のヒストグラムです. それぞれ面積が 1 となるように規格化されているので、これらのヒストグラムは確率密度関数の近似です. 得られた乱数から、数値計算により、母平均 $\hat{\mu}_1$ が 3.25 より小さい確率は

$$P(\hat{\mu}_1 < 3.25) = 0.067$$

であり、母平均 $\hat{\mu}_2$ が 3.4 より小さい確率は

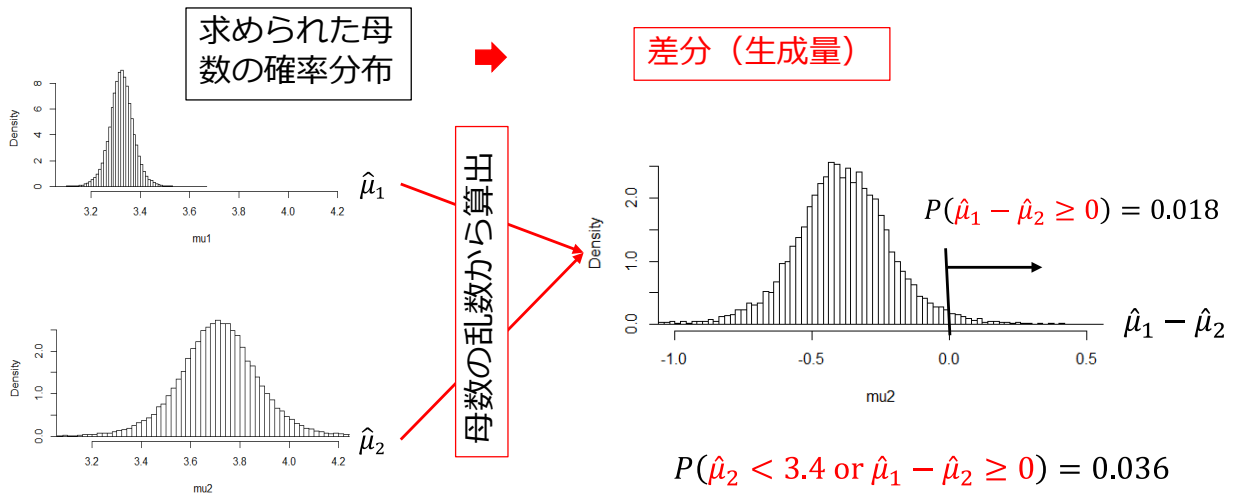
$$P(\hat{\mu}_2 < 3.4) = 0.035$$

と求められます. さらに、 $\hat{\mu}_1 \leq 3.25$ もしくは $\hat{\mu}_2 \leq 3.4$ である確率は

$$P(\hat{\mu}_1 \leq 3.25 \text{ or } \hat{\mu}_2 \leq 3.4) = 0.099$$

と計算されます.

差分は生成量の1つとして求められる。



複雑な組み合わせの確率も母数の乱数から容易に求められる。

μ_2 は3.4より大きいと結論づけると、第1種の過誤の確率は3.5%
 μ_2 は μ_1 より大きいと " " 1.8%
 両結論とも間違える確率は3.6%

27

母平均 μ_1, μ_2 の推定値の差分 $\hat{\mu}_1 - \hat{\mu}_2$ は生成量の1つです。推定値の乱数 $\hat{\mu}_1, \hat{\mu}_2$ から数値計算により、差分 $\hat{\mu}_1 - \hat{\mu}_2$ を求められます。右図が得られたヒストグラムです。面積が1となるように規格化されているので、このヒストグラムは確率密度関数の近似です。 $\hat{\mu}_1 - \hat{\mu}_2 \geq 0$ である確率は

$$P(\hat{\mu}_1 - \hat{\mu}_2 \geq 0) = 0.018$$

です。 $\hat{\mu}_2 < 3.4$ もしくは $\hat{\mu}_1 - \hat{\mu}_2 \geq 0$ である確率は

$$P(\hat{\mu}_2 < 3.4 \text{ or } \hat{\mu}_1 - \hat{\mu}_2 \geq 0) = 0.036$$

です。理論的導出が簡単でない確率も、数値計算により容易に求められます。

以下は、結果の提示例です。

μ_2 は3.4より大きいと結論づけると、第1種の過誤の確率は3.5%です。

μ_2 は μ_1 より大きいと結論づけると、第1種の過誤の確率は1.8%です。

両結論とも間違える確率は3.6%です。

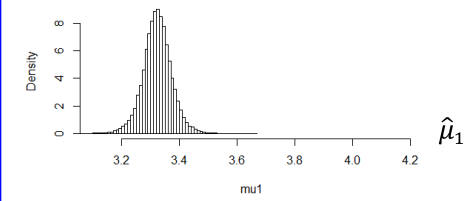
帰無仮説が無いので、仮説の棄却/保留の判定をしていません。複数の帰無仮説の棄却/保留の場合分けがないので、ベイズ統計は多重性の問題を考慮した複雑な確率計算から解放されています。ベイズ統計では種々の生成量を上記のように数値計算により容易に求められます。結論を採用するかどうかの決断は結果の利用者に委ねられます。

さらに,

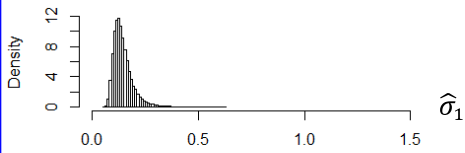
求められた母数の確率分布



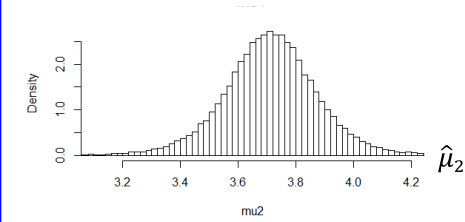
種々の生成量の評価



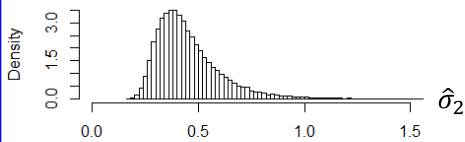
$\hat{\mu}_1$



$\hat{\sigma}_1$



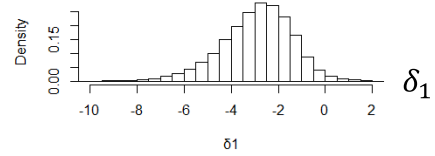
$\hat{\mu}_2$



$\hat{\sigma}_2$

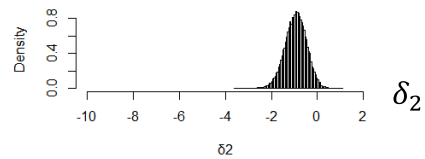
母数の乱数から算出

効果量 $\delta_1 = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_1}$ (4.1)



δ_1

効果量 $\delta_2 = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_2}$ (4.2)

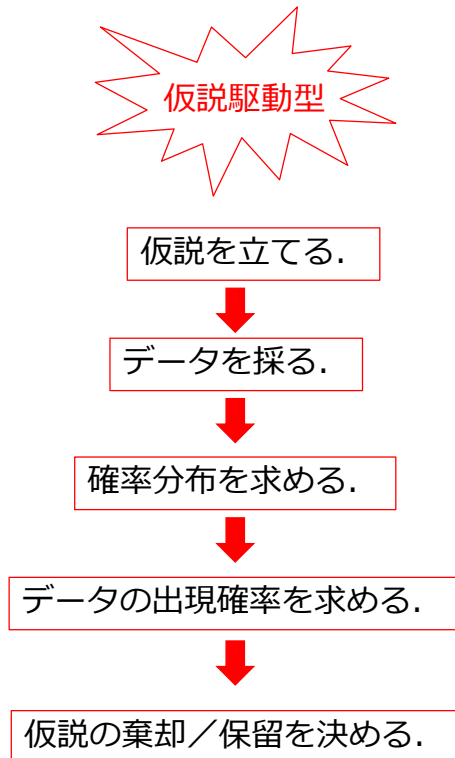


δ_2

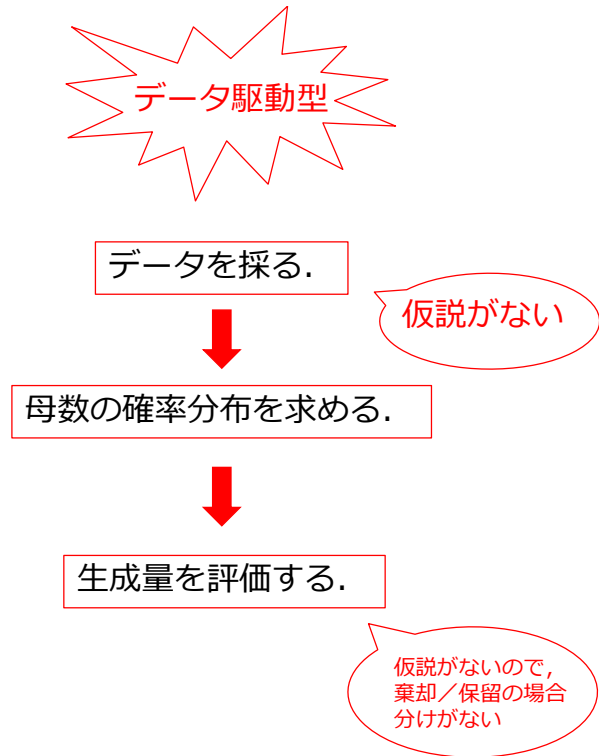
非重複度, 優越率, 闕上率, etc

母数に関する評価指標として, 種々の生成量が提案されています.

統計的仮説検定



ベイズ統計



29

以上、多重性の問題に焦点を当てて、統計的仮説検定とベイズ統計との違いを解説してきました。両者の流れをこのスライドに再掲します。統計的仮説検定は仮説駆動型、ベイズ統計はデータ駆動型です。この手法の違いにより多重性の問題は両者で大きく異なります。

統計的仮説検定では、多重性の問題があります。 $FWER \leq 0.05$ とする多重比較の方法が種々提案されています。仮説の組み合わせによっては、閾値の理論を実験者が導出しなければいけないこともあります。

ベイズ統計は多重性の問題を考慮した複雑な確率計算から解放されています。仮説がないため、複数の帰無仮説の棄却／保留の場合分けがありません。よって、母数の推定値（乱数）に対して種々の検討を、場合分けに悩まされることなく、加えることができます。そして、結論を採択した場合の過誤の確率等は数値計算により容易に求められます。

5. メトロポリス・ヘイスティングス法

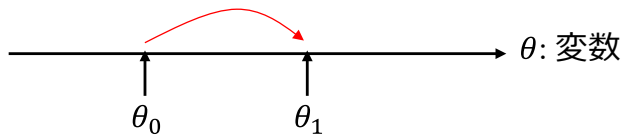
ベイズ統計を実用的にした数値計算手法

30

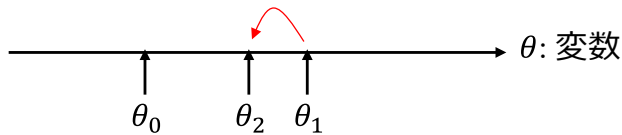
ここからは母数の推定値（乱数）の生成法の解説です。

まず、メトロポリス・ヘイスティングス法の解説です。これはベイズ統計を実用的にした数値計算手法です。

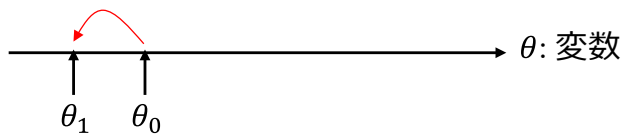
マルコフ過程



乱数候補 θ_1 を生成する. θ_1 と θ_0 を比較して, 所定の条件で θ_1 を受容する. もしくは, θ_1 を棄てて θ_0 を受容する.



θ_1 が受容された場合は, 乱数 θ_2 を生成する. θ_2 と θ_1 を比較して, 同条件で θ_2 を受容する. もしくは, θ_2 を棄てて θ_1 を受容する.



θ_1 が受容されなかった場合は, 乱数 θ_1 を再度生成する. 同条件で θ_1, θ_0 を取捨選択する.

31

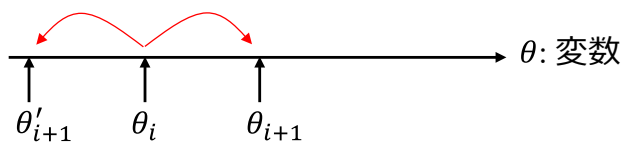
ベイズ統計では母数推定値の乱数を生成します. このスライドではマルコフ過程による乱数候補の取捨選択の仕組みです.

いま, 初期値 θ_0 があるとします. ここで, 乱数候補 θ_1 を生成し, θ_1 と θ_0 を比較して, 所定の条件で θ_1 を受容します. もしくは, θ_1 を棄てて θ_0 を受容します.

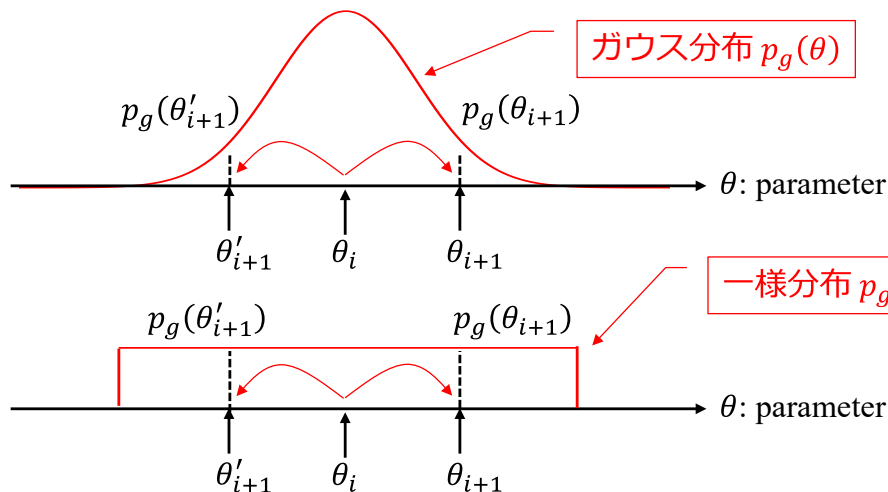
θ_1 が受容された場合は, 乱数候補 θ_2 を生成し, θ_2 と θ_1 を比較して, 同条件で θ_2 を受容します. もしくは, θ_2 を棄てて θ_1 を受容します.

θ_1 が受容されなかった場合は, 乱数候補 θ_1 を再度生成します. 同条件で θ_1, θ_0 を取捨選択します.

メトロポリス・ヘイスティングス法



θ_{i+1} の生成確率を $p_g(\theta_{i+1})$ とする. $p_g(\theta_{i+1})$ は θ_i を中心に対称とする.
 if $|\theta_{i+1} - \theta_i| = |\theta'_{i+1} - \theta_i|$, then $p_g(\theta_{i+1}) = p_g(\theta'_{i+1})$ (5.1)



32

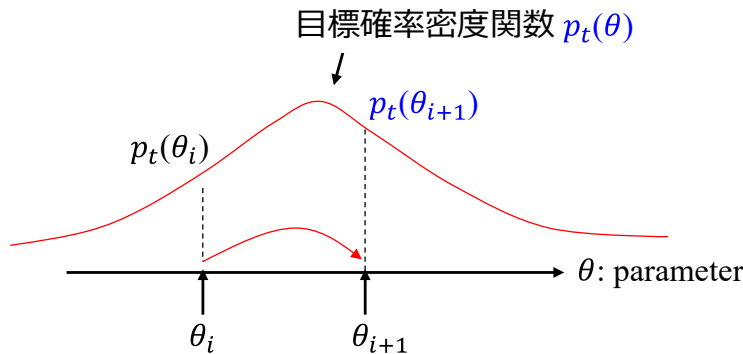
メトロポリス・ヘイスティングス法の記述を簡単にするために、乱数の生成確率を「左右対称」にします。

i 番目に受容された乱数を θ_i とします。 θ_{i+1} の生成確率を $p_g(\theta_{i+1})$ とします。ポイントは $p_g(\theta_{i+1})$ を θ_i を中心に対称とすることです。すなわち

$$\text{if } |\theta_{i+1} - \theta_i| = |\theta'_{i+1} - \theta_i|, \text{ then } p_g(\theta_{i+1}) = p_g(\theta'_{i+1})$$

です。ただし、この関係を満たす関数には θ_i の左右等距離に 1 個ずつの δ 関数を配したのも含まれてしまうので、「ある範囲内ではいたる所で 0 より大きい有限な値を持つ」左右対称分布とします。ここでは、 $p_g(\theta)$ には図のようなガウス分布、一様分布を使用することとします。

メトロポリス・ヘイスティングス法



注意!

$$p_t(\theta) \neq p_g(\theta)$$

$p_t(\theta)$ は生成される乱数が従うべき確率密度関数

乱数候補の受容条件

θ_{i+1} を θ_i の後継値として受容する条件:

$$\text{If } p(\theta_{i+1}) \geq p(\theta_i), \text{ then } \theta_{i+1} \text{ の受容確率} = 1. \quad (5.2)$$

$$\text{If } p(\theta_{i+1}) < p(\theta_i), \text{ then } \theta_{i+1} \text{ の受容確率} = p_t(\theta_{i+1})/p_t(\theta_i). \quad (5.3)$$

$$\theta_{i+1} \text{ の受容確率} = \min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right) \quad (5.4)$$

$$\theta_{i+1} \text{ の非受容確率} = 1 - \min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right) \quad (\text{この場合は } \theta_i \text{ を受容}) \quad (5.5)$$

33

メトロポリス・ヘイスティングス法の乱数候補の受容条件です。

$p_t(\theta)$ を目標確率密度関数とします。 $p_t(\theta) \neq p_g(\theta)$ なので注意してください。 $p_g(\theta)$ は乱数候補 θ の生成確率です。 $p_t(\theta)$ は取捨選択後に残された乱数 θ が従うべき確率密度関数です。残された乱数の(面積が1となるように規格化された)ヒストグラムはこの確率密度関数の近似となることを目標にします。

θ_{i+1} を θ_i の後継値として受容する条件を

$$\text{If } p(\theta_{i+1}) \geq p(\theta_i), \text{ then } \theta_{i+1} \text{ の受容確率} = 1.$$

$$\text{If } p(\theta_{i+1}) < p(\theta_i), \text{ then } \theta_{i+1} \text{ の受容確率} = p(\theta_{i+1})/p(\theta_i).$$

とします。この条件は

$$\theta_{i+1} \text{ の受容確率} = \min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right)$$

$$\theta_{i+1} \text{ の非受容確率} = 1 - \min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right) \quad (\text{この場合は } \theta_i \text{ を受容})$$

と書き換えることができます。

なお、 $p_g(\theta)$ が「左右対称」でない場合には、 θ が θ_x であるときに θ_y の値を候補として生成する条件付き確率を $p_g(\theta_y|\theta_x)$ とし、上記受容確率を

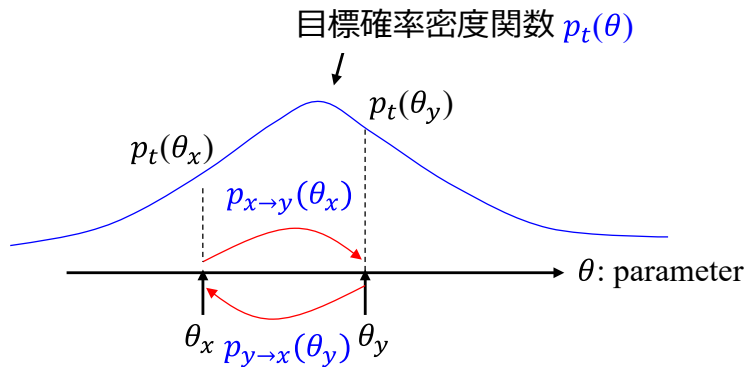
$$\theta_{i+1} \text{ の受容確率} = \min\left(1, \frac{p_g(\theta_i|\theta_{i+1})p_t(\theta_{i+1})}{p_g(\theta_{i+1}|\theta_i)p_t(\theta_i)}\right)$$

とします. $p_g(\theta)$ が「左右対称」の場合は

$$p_g(\theta_i|\theta_{i+1}) = p_g(\theta_{i+1}|\theta_i)$$

となり, [\(5.4\)式](#)が得られます.

遷移確率比



$p_{x \rightarrow y}(\theta_x)$: 遷移確率
 θ_x から θ_y へと遷移する確率

$$\frac{p_{y \rightarrow x}(\theta_y)}{p_{x \rightarrow y}(\theta_x)} = \frac{\min\left(1, \frac{p_t(\theta_x)}{p_t(\theta_y)}\right)}{\min\left(1, \frac{p_t(\theta_y)}{p_t(\theta_x)}\right)} = \begin{cases} \frac{1}{p_t(\theta_y)/p_t(\theta_x)} & (\text{if } p_t(\theta_y) < p_t(\theta_x)) \\ \frac{p_t(\theta_x)/p_t(\theta_y)}{1} & (\text{if } p_t(\theta_y) \geq p_t(\theta_x)) \end{cases}$$

$$= \frac{p_t(\theta_x)}{p_t(\theta_y)} \quad (5.6)$$

$$p_t(\theta_y)p_{y \rightarrow x}(\theta_y) = p_t(\theta_x)p_{x \rightarrow y}(\theta_x) \quad \text{詳細つり合い条件} \quad (5.7)$$

34

遷移確率比を確認します。前スライドのように乱数候補の受容条件を設定することで「詳細つり合い条件」が満たされることを示します。

遷移確率 $p_{x \rightarrow y}(\theta_x)$ を θ_x から θ_y へと遷移する確率と定義します。 θ_y から θ_x へと遷移する確率は $p_{y \rightarrow x}(\theta_y)$ と表記します。前スライドの乱数候補受容条件は、遷移確率と同じなので

$$p_{x \rightarrow y}(\theta_x) = \min\left(1, \frac{p_t(\theta_y)}{p_t(\theta_x)}\right)$$

です。よって、 $p_{x \rightarrow y}(\theta_x)$ と $p_{y \rightarrow x}(\theta_y)$ の比は

$$\frac{p_{y \rightarrow x}(\theta_y)}{p_{x \rightarrow y}(\theta_x)} = \frac{\min\left(1, \frac{p_t(\theta_x)}{p_t(\theta_y)}\right)}{\min\left(1, \frac{p_t(\theta_y)}{p_t(\theta_x)}\right)}$$

$$= \begin{cases} \frac{1}{p_t(\theta_y)/p_t(\theta_x)} & (\text{if } p_t(\theta_y) < p_t(\theta_x)) \\ \frac{p_t(\theta_x)/p_t(\theta_y)}{1} & (\text{if } p_t(\theta_y) \geq p_t(\theta_x)) \end{cases}$$

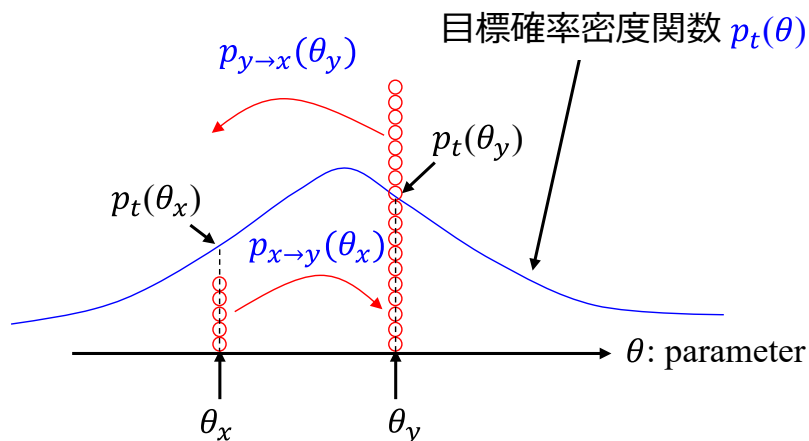
$$= \frac{p_t(\theta_x)}{p_t(\theta_y)}$$

となります。この式を変形すると

$$p_t(\theta_y)p_{y \rightarrow x}(\theta_y) = p_t(\theta_x)p_{x \rightarrow y}(\theta_x)$$

です。これは詳細つり合い条件と呼ばれます。

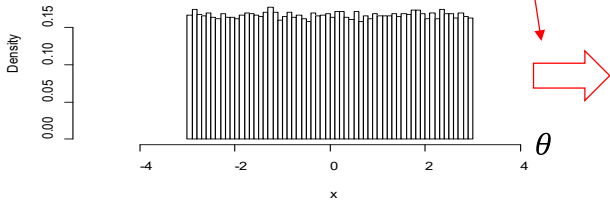
乱数候補の取捨選択に(5.4), (5.5)式を適用します。任意分布（例えば一様分布）しているたくさんの乱数を初期乱数として、各乱数に対して次候補を生成・取捨選択することを繰り返していくと、やがてこれら乱数は目標確率密度関数 $p_t(\theta)$ に沿うように分布します。



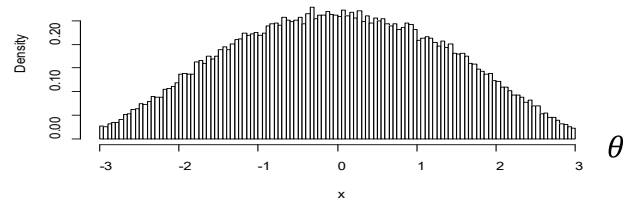
その理由を説明するイメージ図です。いま θ_x には目標確率密度関数 $p_t(\theta_x)$ が規定する数より少ない乱数が配置されていて、 θ_y には規定数より多い乱数が配置されているとします。(5.7)式はそれぞれに規定数ずつ配置されているときに互いへの遷移頻度が同じであることを意味します。従って、上図の例では θ_y から θ_x へと移動する頻度は、 θ_x から θ_y へと移動する頻度よりも高くなります。詳細つり合い条件が θ の任意の対で成立していることで、乱数の生成・取捨選択を繰り返していけば、乱数の分布はやがて目標確率密度関数 $p_t(\theta)$ が規定する数に均されていきます。

例1. 目標確率密度関数： $p_t(\theta) \sim N(0, 1)$ (標準正規分布) .
 区間 $[-3, 3]$ の一様乱数を100000個生成. これら100000個をそれぞれ θ_i として,
 $p_g(\theta_{i+1}) \sim \text{uniform}(-3, 3)$ の候補を生成して取捨選択.

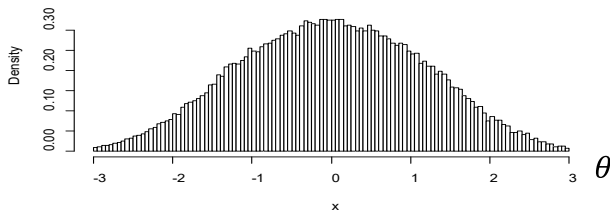
$$\text{次候補の受容確率} = \min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right)$$



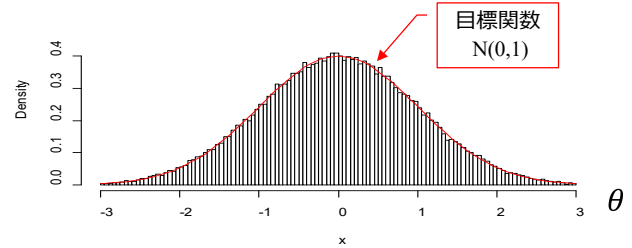
(a) 初期ヒストグラム



(b) 1 回目の遷移後



(c) 2 回目の遷移後

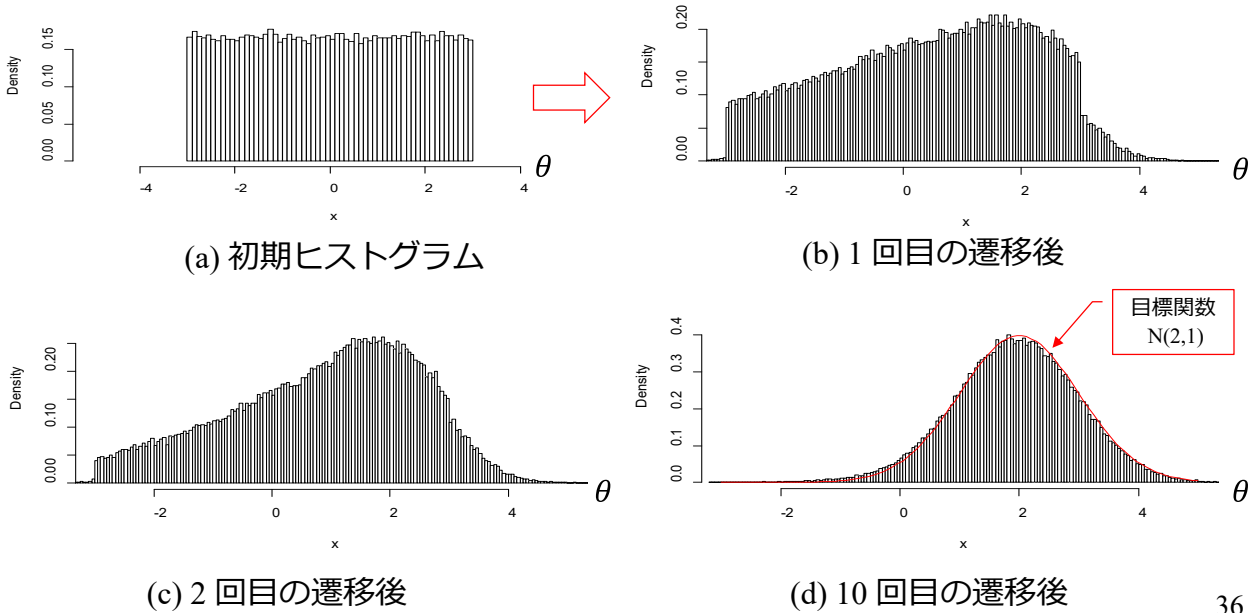


(d) 10 回目の遷移後

35

シミュレーション例の紹介です. 目標確率密度関数を $p_t(\theta) \sim N(0, 1)$ (標準正規分布) とします. 初期乱数として区間 $[-3, 3]$ に一様乱数を 100000 個生成します. 図(a)は初期乱数のヒストグラムです. これら 100000 個をそれぞれ θ_i として, $p_g(\theta_{i+1}) \sim \text{uniform}(-3, 3)$ (区間 $[-3, 3]$ の一様乱数) の候補を生成して, 受容確率を $\min\left(1, \frac{p_t(\theta_{i+1})}{p_t(\theta_i)}\right)$ とします. 図(b)は 100000 個全てに対して 1 回取捨選択を実施した結果のヒストグラムです. 図(c)は 2 回実施後, (d)は 10 回実施後の結果です. 赤線は目標確率密度関数 $N(0, 1)$ です.

例2. 目標確率密度関数 : $p_t(\theta) \sim N(2, 1)$ (平均値が 2 の正規分布) .
 区間 $[-3, 3]$ の一様乱数を 100000 個生成. これら 100000 個をそれぞれ θ_i として,
 $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ の候補を生成して取捨選択.

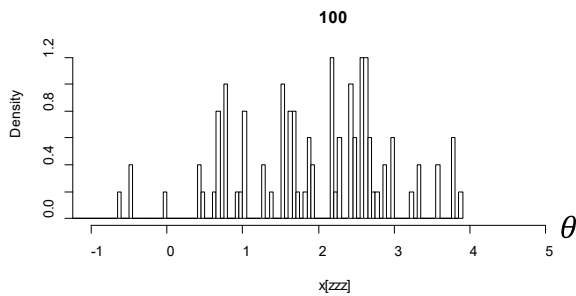


36

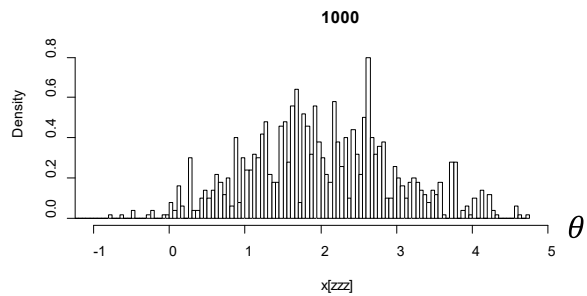
条件を変えたシミュレーション例です. 目標確率密度関数を $p_t(\theta) \sim N(2, 1)$ (平均値が 2 の正規分布) に変えました. 初期乱数は同じです. 区間 $[-3, 3]$ に一様乱数を 100000 個生成しました. これら 100000 個をそれぞれ θ_i として, $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ の候補を生成して取捨選択を実施しました. 分布の中心が $\theta = 2$ に次第に移動して行って, 図(d)の 10 回実施後には目標確率密度関数 $N(2,1)$ に沿う分布が得られています.

例3. 目標確率密度関数： $p_t(\theta) \sim N(2, 1)$ (平均値が2の正規分布) .

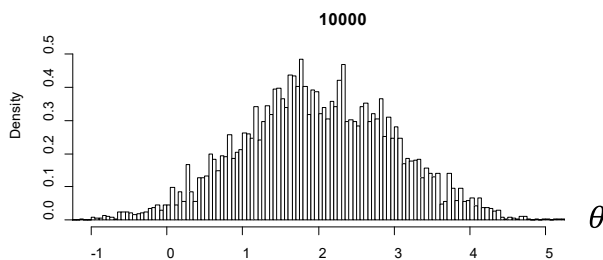
区間 $[-3, 3]$ の一様乱数を1個生成. $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ の候補を生成して取捨選択することを1000000回繰り返した.



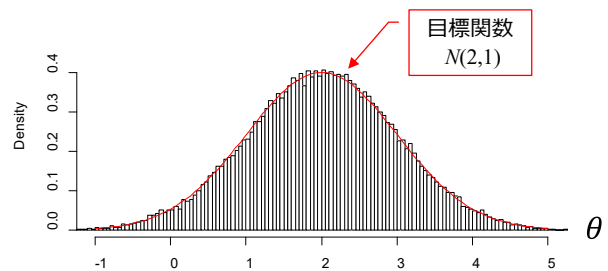
(a) $i = 100$



(b) $i = 1000$



(c) $i = 10000$



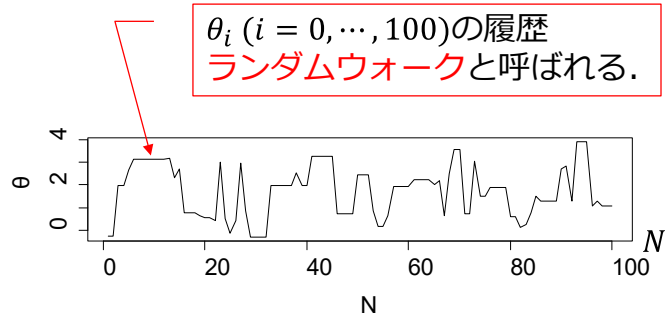
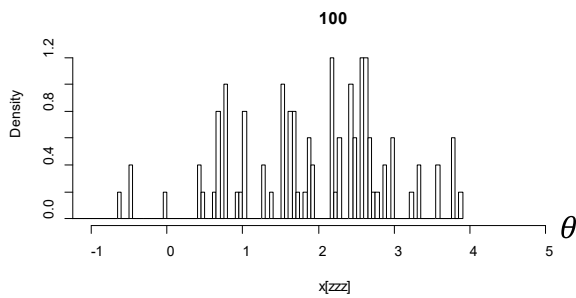
(d) $i = 100000$

37

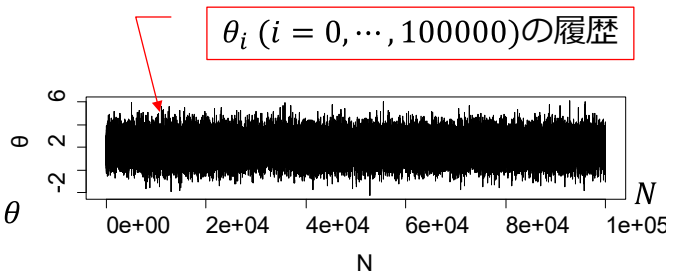
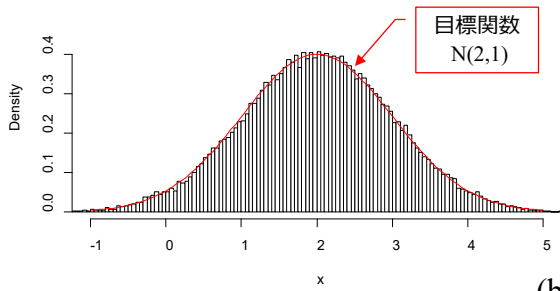
これは、1個の初期乱数を生成して、1個の乱数候補を生成・取捨選択を繰り返した結果です。毎回の取捨選択において受容された乱数を全て記録して、そのヒストグラムを描いてあります。

目標確率密度関数を $p_t(\theta) \sim N(2, 1)$ (平均値が2の正規分布) とし、区間 $[-3, 3]$ の一様乱数を1個生成し、 $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ の候補を生成して取捨選択することを1000000回繰り返しました。図(a)は100回の取捨選択を実施した結果です。100個の乱数のヒストグラムなので、棒がまばらです。図(b)は1000回実施後の結果です。まだ、棒がまばらです。図(c)は10000回実施後の結果です。正規分布の輪郭が見えてきました。図(d)は100000回実施後のヒストグラムです。目標確率密度関数 $N(2,1)$ に沿う乱数の分布が得られました。

例3 (つづき)



(a) $i = 100$



(b) $i = 100000$

38

スライド番号 37 のシミュレーション例において受容された乱数の履歴をグラフにしました。右上は 100 回の取捨選択を実施した際に受容された乱数を示します。横軸の N が実施回数。縦軸が乱数値です。これはランダムウォークと呼ばれます。右下は 100000 回の取捨選択における受容乱数の履歴です。

6. ベイズの定理と母数推定

ベン図

$$P(a_i|b_j)P(b_j) = P(b_j|a_i)P(a_i) \quad (6.1)$$

$a_i \in A (i = 1, \dots, a), b_j \in B (j = 1, \dots, b)$

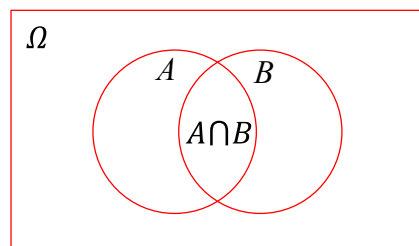
Ω :対象領域

$P(a_i)$: $a_i \in A$ の生起確率

$P(b_j)$: $b_j \in B$ の生起確率

$P(a_i|b_j)$: 条件付き確率: $b_j \in B$ であるときに, $a_i \in A \cap B$ の生起確率

$P(b_j|a_i)$: 条件付き確率: $a_i \in A$ であるときに, $b_j \in A \cap B$ の生起確率



ベイズの定理

$$\begin{aligned} P(a_i|b_j) &= \frac{P(b_j|a_i)P(a_i)}{P(b_j)} \\ &= \frac{P(b_j|a_i)P(a_i)}{\sum_{i=1}^a P(b_j|a_i)P(a_i)} \end{aligned} \quad (6.2)$$

$P(a_i|b_j)$: 事後確率

$P(b_j|a_i)$: 尤度

$P(a_i)$: 事前確率

$\sum_{i=1}^a P(b_j|a_i)P(a_i)$: 周辺確率

39

さて、ベイズの定理はどのようにして利用されるのでしょうか？

図のベン図において次式が成立します。

$$P(a_i|b_j)P(b_j) = P(b_j|a_i)P(a_i)$$

この式を変形すると

$$\begin{aligned} P(a_i|b_j) &= \frac{P(b_j|a_i)P(a_i)}{P(b_j)} \\ &= \frac{P(b_j|a_i)P(a_i)}{\sum_{i=1}^a P(b_j|a_i)P(a_i)} \end{aligned}$$

が得られます。

連続系の場合

ベイズの定理

母変数 μ, σ の下でデータ x が生起する確率密度関数

母変数 μ, σ の確率密度関数

$$p(\mu, \sigma | x) = \frac{p(x | \mu, \sigma) p(\mu, \sigma)}{P(x)} \\ = \frac{p(x | \mu, \sigma) p(\mu, \sigma)}{\int_{-\infty}^{\infty} p(x | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma} \quad (6.3)$$

観測されたデータ x とそれを生ぜしめた母変数 μ, σ の確率密度関数

データ x の確率密度関数

$p(\mu, \sigma | x)$: 事後確率密度関数

$p(x | \mu, \sigma)$: 尤度

$p(\mu, \sigma)$: 事前確率密度関数

$P(x) = \int_{-\infty}^{\infty} p(x | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma$: 周辺確率密度関数

40

連続系の場合のベイズの定理です。

左辺の $p(\mu, \sigma | x)$ は観測されたデータ x の下で、それを生ぜしめた母数 μ, σ の確率密度関数です。

右辺の $p(x | \mu, \sigma)$ は母数 μ, σ の下でデータ x が生起する確率密度関数です。 $p(\mu, \sigma)$ は母数 μ, σ の確率密度関数です。

右辺の分母の $\int_{-\infty}^{\infty} p(x | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma$ はデータ x の確率密度関数です。

ベイズの定理は、 $p(\mu, \sigma)$ を母数 μ, σ の事前情報とすると、データ x が得られた後では、右辺の計算により左辺の $p(\mu, \sigma | x)$ を事後情報として得られると解釈されています。

$p(\mu, \sigma | x)$ は事後確率密度関数、 $p(x | \mu, \sigma)$ は尤度、 $p(\mu, \sigma)$ は事前確率密度関数と呼ばれます。

メトロポリ・ヘイスティングス(MH)法にベイズの定理を適用すると

$$r \equiv \frac{p_{i \rightarrow j}(\theta_i)}{p_{j \rightarrow i}(\theta_j)} = \frac{p_t(\theta_j)}{p_t(\theta_i)} : \text{遷移確率の比} \quad (6.4)$$

$$\begin{aligned}
 r &= \frac{p_t(\mu_j, \sigma_j | \mathbf{x})}{p_t(\mu_i, \sigma_i | \mathbf{x})} \\
 &= \frac{\frac{p(\mathbf{x} | \mu_j, \sigma_j) p(\mu_j, \sigma_j)}{\int_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma}}{\frac{p(\mathbf{x} | \mu_i, \sigma_i) p(\mu_i, \sigma_i)}{\int_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma}} \\
 &= \frac{p(\mathbf{x} | \mu_j, \sigma_j) p(\mu_j, \sigma_j)}{p(\mathbf{x} | \mu_i, \sigma_i) p(\mu_i, \sigma_i)} \\
 &= \frac{p(\mathbf{x} | \mu_j, \sigma_j)}{p(\mathbf{x} | \mu_i, \sigma_i)} \quad (6.5)
 \end{aligned}$$

積分演算不要

無情報的事前分布（一様分布）を採用

MH法により分布 $p(\mu, \sigma | \mathbf{x})$ を尤度 $p(\mathbf{x} | \mu, \sigma)$ から生成できる。

ベイズの定理をメトロポリス・ヘイスティングス法に適用します。

目標確率密度関数を事後確率密度関数 $p_t(\mu, \sigma | \mathbf{x})$ とします。(6.4)式の遷移確率の比にベイズの定理を代入すると

$$\begin{aligned}
 r &= \frac{p_t(\mu_j, \sigma_j | \mathbf{x})}{p_t(\mu_i, \sigma_i | \mathbf{x})} \\
 &= \frac{\frac{p(\mathbf{x} | \mu_j, \sigma_j) p(\mu_j, \sigma_j)}{\int_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma}}{\frac{p(\mathbf{x} | \mu_i, \sigma_i) p(\mu_i, \sigma_i)}{\int_{-\infty}^{\infty} p(\mathbf{x} | \mu, \sigma) p(\mu, \sigma) d\mu d\sigma}}
 \end{aligned}$$

となり、分母、分子の積分項を相殺できます。複雑な確率密度関数の場合には積分を解析的に求められないこともあり得ますが、その積分の必要がありません。そして、事前情報が何も無いとして、事前確率密度関数 $p(\mu, \sigma)$ に無情報的事前分布（一様分布）を採用すると

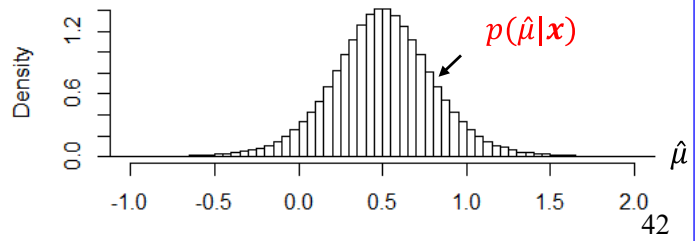
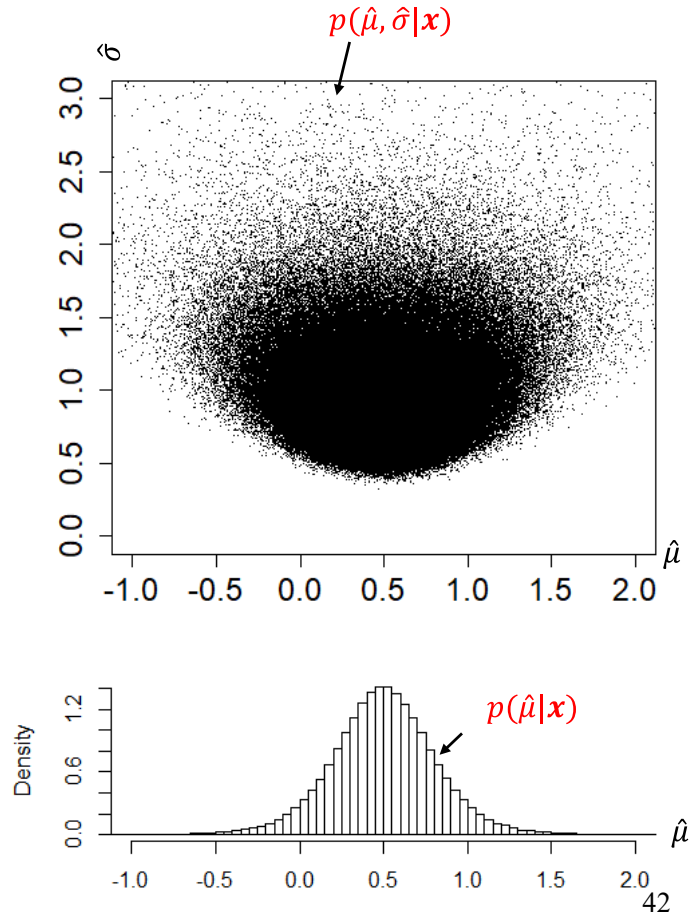
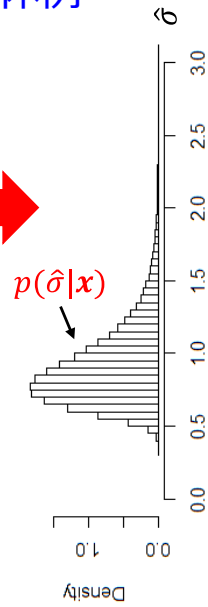
$$\begin{aligned}
 r &= \frac{p(\mathbf{x} | \mu_j, \sigma_j) p(\mu_j, \sigma_j)}{p(\mathbf{x} | \mu_i, \sigma_i) p(\mu_i, \sigma_i)} \\
 &= \frac{p(\mathbf{x} | \mu_j, \sigma_j)}{p(\mathbf{x} | \mu_i, \sigma_i)}
 \end{aligned}$$

となります。これは、MH法により分布 $p_t(\mu, \sigma | \mathbf{x})$ を尤度 $p(\mathbf{x} | \mu, \sigma)$ から生成できることを意味します。

ベイズ統計の具体例

	Samples
x1	0.302
x2	1.01
x3	1.598
x4	0.184
x5	0.958
x6	0.436
x7	0.283
x8	0.854
x9	-1.127

\mathbf{x} →



スライド番号 13 のヒストグラムの再掲です。右上の散布図がデータ \mathbf{x} を基に生成した $p(\hat{\mu}, \hat{\sigma}|\mathbf{x})$ に従う乱数の例です。事前確率密度関数に一様分布を使用し、尤度 $p(\mathbf{x}|\mu, \sigma)$ に

$$p(\mathbf{x}|\mu, \sigma) = \prod_{i=1}^n p(x_i|\mu, \sigma)$$

$$p(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - \mu)^2}{\sigma^2}\right\}$$

の正規分布を用いています。

7. ハミルトニアンモンテカルロ法

なぜHMC(Hamiltonian Monte Carlo)法が使用されているのか？

推定対象が多変数となった場合に、MH法による多次元空間内の探索が遅いことがある。

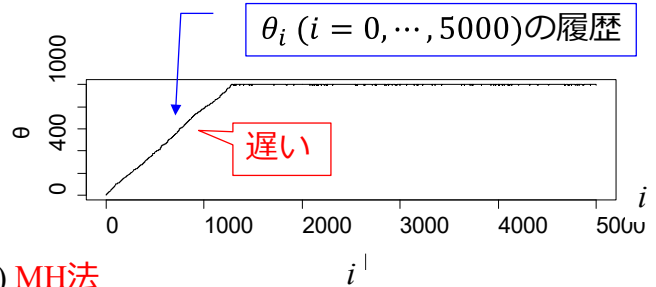
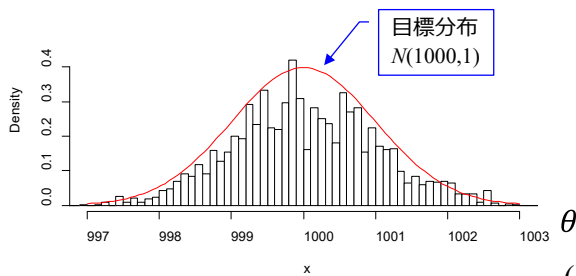
43

なぜハミルトニアンモンテカルロ(HMC: Hamiltonian Monte Carlo)法が使用されているのでしょうか？それは、推定対象が多変数となった場合に、MH法による多次元空間内の探索が遅いことがあるからです。

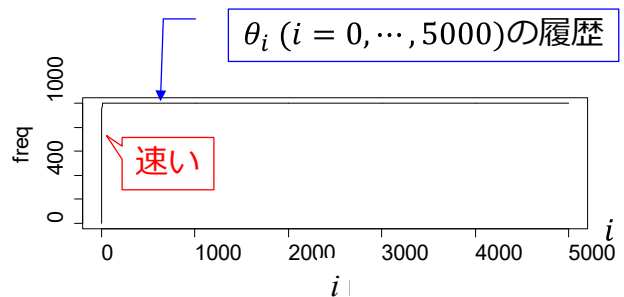
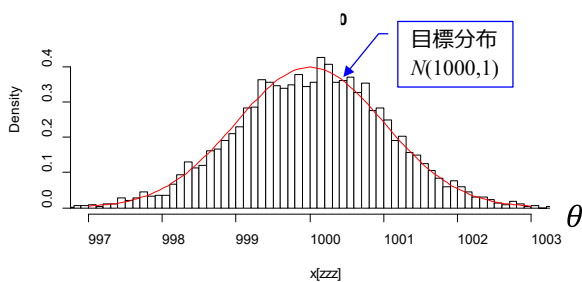
例4. MH法では目標領域への到達が遅い例を強引に作ってみる.

目標確率密度関数 $p_t(\theta) \sim N(1000, 1)$.

MH法において $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ による乱数生成・取捨選択を5000回繰り返した.



(a) MH法



(b) HMC法 ($\epsilon = 0.1, L = 10$)

44

MH法では目標領域への到達が遅い例を強引に作ってみました.

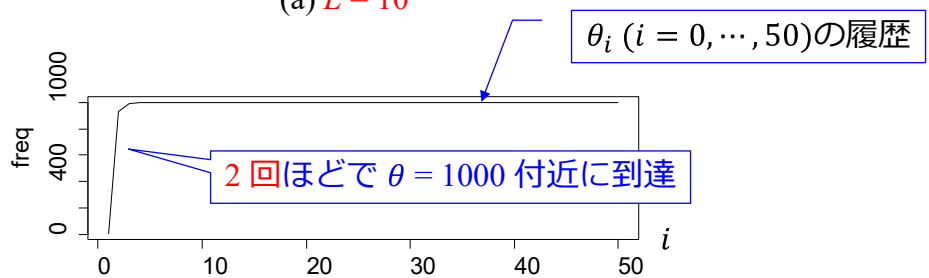
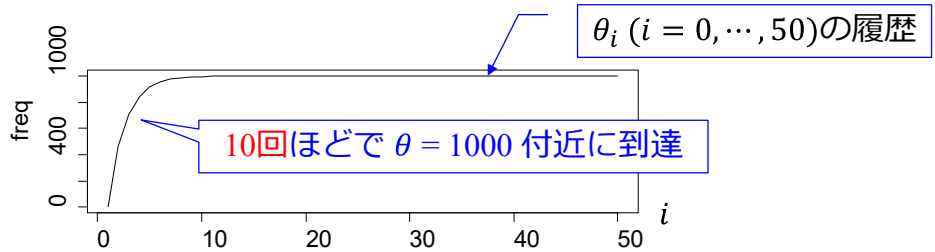
目標確率密度関数を $p_t(\theta) \sim N(1000, 1)$ とします. 目標正規分布の母平均 $\mu = 1000$ です. MH法において, $p_g(\theta_{i+1}) \sim \text{uniform}(\theta_i - 3, \theta_i + 3)$ による乱数生成・取捨選択を5000回繰り返した結果を示します. 図(a)がMH法の結果です. 左がヒストグラム, 右が5000回のランダムウォークの結果です. ヒストグラムは面積が1になるように規格化してあります. 赤線が目標分布です. 右のランダムウォークでは, 生成される乱数は $\theta = 0$ 付近からスタートして, $i = 1300$ 個辺りまでゆっくりと移動して, $\theta = 1000$ 付近に到達しています. この付近に到達以降のランダムウォークにより図(a)左のヒストグラムが得られました. 右のランダムウォークの図は縦軸の1目盛りが200であるため, $\theta = 1000$ 付近の乱数の変化 (標準偏差 = 1程度) はほぼ直線に描かれています. [スライド番号 38](#) のランダムウォークとは縦軸の1目盛りが大きく違うので注意してください.

図(b)はHMC法により得られたヒストグラムとランダムウォークの結果です. MH法との大きな違いは, 生成される乱数が速やかに $\theta = 1000$ 付近に到達している点です.

例5. HMC法におけるパラメータの効果

$\varepsilon = 0.1$ (固定), $L = 10, 15$ を選択

Stanは ε, L を自動調整できる.



前スライド図(b)のランダムウォークのグラフを横方向に拡大して本スライド図(a)に示します。乱数が10個ほど生成された段階で $\theta = 1000$ 付近に到達しています。HMC法には ε と L の2つのパラメータがあります。図(a)は $\varepsilon = 0.1, L = 10$ の場合です。これらのパラメータを調整すると、 $\theta = 1000$ 付近への到達の様子が変わります。図(b)は $\varepsilon = 0.1, L = 15$ とした場合です。乱数2個目で到達しています。

Stanには ε と L を自動調整する機能が備わっています。

HMC法

$$\text{ハミルトニアン: } H(p_m, x) = \frac{p_m^2}{2m} + U(x)$$

\uparrow \uparrow
運動エネルギー 位置エネルギー

m : 質量, $p_m = mv$: 運動量, x : 位置

$$\frac{\partial H}{\partial p_m} = \frac{p_m}{m} = v = \frac{dx}{dt} \quad (7.1)$$

$$\frac{\partial H}{\partial x} = \frac{\partial U}{\partial x} = -ma = -\frac{dp_m}{dt} \quad (7.2)$$

46

さて、本講義の最後の話題、HMC です。

ハミルトニアンは工学部の多くの学科で学ぶ機会があります。いきなり力学系の話になりますが、これが乱数生成の仕組みへと見事につながっていきます。もし、読者がその見事さを感じられないとしたら、単に筆者の説明力不足に責を帰せられます。

ハミルトニアン $H(p_m, x)$ を

$$H(p_m, x) = \frac{p_m^2}{2m} + U(x)$$

と表します。 m は質量, p_m は運動量 ($p_m = mv$, v は速度), $\frac{p_m^2}{2m}$ は運動エネルギー, x は位置, $U(x)$ は位置エネルギーです。 H を p_m および x で偏微分すると

$$\frac{\partial H}{\partial p_m} = \frac{p_m}{m} = v = \frac{dx}{dt}$$

$$\frac{\partial H}{\partial x} = \frac{\partial U}{\partial x} = -ma = -\frac{dp_m}{dt}$$

となります。 a は加速度です。 $\partial U / \partial x = -ma$ はエネルギーの減少方向に力 ma が発生することを意味します。

微分方程式 (7.1), (7.2) を解くことで, (x, p_m) -空間内の x, p_m の軌跡を得られる.

$$\frac{dx}{dt} = \frac{p_m}{m} \quad (7.1)$$

$$\frac{dp_m}{dt} = -\frac{\partial H}{\partial x} \quad (7.2)$$

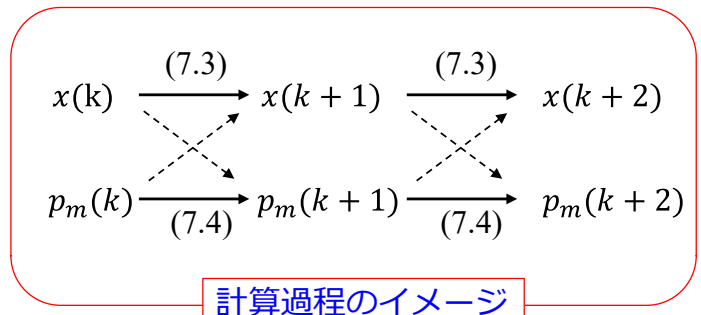
オイラー法 : 最も簡便な近似解法

$$\frac{x(k+1) - x(k)}{\varepsilon} \approx \frac{p_m(k)}{m}$$

$$\frac{p_m(k+1) - p_m(k)}{\varepsilon} \approx -\frac{\partial H(x(k))}{\partial x}$$

$$x(k+1) \approx x(k) + \varepsilon \frac{p_m(k)}{m} \quad (7.3)$$

$$p_m(k+1) \approx p_m(k) - \varepsilon \frac{\partial H(x(k))}{\partial x} \quad (7.4)$$



47

微分方程式 (7.1), (7.2) を解くことで, (x, p_m) -空間内の x, p_m の軌跡を得られます. 最も簡便な近似解法はオイラー法です. 時間微分項をそれぞれ

$$\frac{dx}{dt} \approx \frac{x(k+1) - x(k)}{\varepsilon}$$

$$\frac{dp_m}{dt} \approx \frac{p_m(k+1) - p_m(k)}{\varepsilon}$$

により近似します. $k = 1, 2, \dots$ は離散的な時刻, ε は微小時刻です. 時刻 $t = 0, \varepsilon, 2\varepsilon, \dots$ のとき $k = 1, 2, 3, \dots$ を対応づけます. これにより, (7.1), (7.2)式は

$$\frac{x(k+1) - x(k)}{\varepsilon} \approx \frac{p_m}{m}$$

$$\frac{p_m(k+1) - p_m(k)}{\varepsilon} \approx -\frac{\partial H(x(k))}{\partial x}$$

と変形されます. さらに,

$$x(k+1) \approx x(k) + \varepsilon \frac{p_m(k)}{m}$$

$$p_m(k+1) \approx p_m(k) - \varepsilon \frac{\partial H(x(k))}{\partial x}$$

を得ます.

図のフローはオイラー法による x, p_m の軌跡の計算過程のイメージです. (7.3)式により, 時刻 k の $x(k), p_m(k)$ から $x(k+1)$ が求まり, (7.4)式により $p_m(k+1)$ が求まります. 以降同様にして, $k+2, k+3, \dots$ の x, p_m を求めていきます.

オイラー法による (x, p_m) -空間内の移動とMH法との関係

48

オイラー法により x, p_m を移動させることと MH 法はどう関係するのでしょうか？

位置エネルギー $U(x)$ を $-\log(\text{目標確率密度関数})$ とすると

$$U(x) = -\log p(x) \quad (7.5)$$

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \exp(-H(p_m, x)) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m} - U(x)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m} + \log p(x)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m}\right) p(x) \end{aligned} \quad (7.6)$$

- ① 標準正規乱数を与えて
- ② (x, p_m) -空間内を移動して
- ③ $\exp(-H(p_m, x))$ の値に対して MH法

$$\min\left(1, \frac{\exp(-H(p_m(L), x(L)))}{\exp(-H(p_m(1), x(1)))}\right)$$

を適用すれば

(7.7)

$N(0, 1)$ ($m = 1$ とする)

$N(0, 1)$ と p の積

高速に $p(x)$ の確率の高い領域へと到達でき、効率よく $p(x)$ に従う乱数を生成できる。

49

位置エネルギー $U(x)$ を $-\log(\text{目標確率密度関数})$ とします。すなわち

$$U(x) = -\log p(x)$$

です。ここで、 $\frac{1}{\sqrt{2\pi}} \exp(-H(p_m, x))$ を求めると

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \exp(-H(p_m, x)) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m} - U(x)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m} + \log p(x)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m}\right) p(x) \end{aligned}$$

と得られます。右辺の運動エネルギーの項は、 $m = 1$ とすると

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2}\right) \sim N(0, 1)$$

となり、標準正規分布 $N(0, 1)$ と同じ式です。従って、右辺は独立な標準正規分布の確率密度関数 $N(0, 1)$ と目標確率密度関数 p の積からなる同時確率密度関数となります。左辺の $\exp(-H(p_m, x))$ の比に MH法を適用すれば、この同時確率密度関数に従う乱数を生成することができます！！

標準正規分布に従う乱数の与え方と、 (x, p_m) -空間内の x, p_m の移動のさせ方について、天下り的に

HMC 法のアロリズムを以下に紹介します.

- ① $x(1)$ に適当な初期値を与えます.
- ② $k=1$ として, $p_m(1)$ に標準正規乱数を与えます.
- ③ (7.3), (7.4)式により, x, p_m を $k=L$ となるまで移動させます.
- ④ MH 法

$$\min \left(1, \frac{\exp(-H(p_m(L), x(L)))}{\exp(-H(p_m(1), x(1)))} \right)$$

(7.7)

を適用し, $x(1), x(L)$ の値を取捨選択して $x(1)$ とし, $x(1)$ を記録して, ②にもどります.

$p_m(1)$ に標準正規乱数を与え, (x, p_m) -空間内を移動させることで $x(L)$ に目標確率密度関数に従う乱数を得られます.

なぜこのアルゴリズムで高速に, しかも, 目標確率密度関数に従う乱数が得られるのでしょうか?

サンプル

$x = x_1$ ($n = 1$): サンプルサイズ 1

尤度

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (7.8)$$

$-\log(\text{尤度})$

$$U(\mu) = -\log p(x|\mu, \sigma) = \log \sigma + \frac{1}{2\sigma^2} (x - \mu)^2 \quad (7.9)$$

$-\log(\text{尤度})$ の微分値

$$\frac{\partial\{-\log p(x|\mu, \sigma)\}}{\partial\mu} = -\frac{1}{\sigma^2} (x - \mu) \quad (7.10)$$

50

具体例により、HMC法の仕組みを見ていきましょう。

準備として、母平均 μ の推定値 $\hat{\mu}$ の乱数を得る R のスクリプトを紹介します。

参考にしたスクリプトは文献[6]~[9]の出版社の Web ページからダウンロードしたもの、および、ネットから無料でダウンロードしたものです。これらのスクリプトの「劣化版」がこれから紹介するスクリプトです。HMC法では「リープフロッグ法」が使われていますが、ここでは性能の劣る、ただし、簡便なオイラー法を適用しています。

サンプル x のサンプルサイズ $n=1$ とします。尤度は母平均 μ 、母標準偏差 σ の正規分布とします。その $-\log(\text{尤度})$

$$U(\mu) = -\log p(x|\mu, \sigma) = \log \sigma + \frac{1}{2\sigma^2} (x - \mu)^2$$

とその微分値

$$\frac{\partial\{-\log p(x|\mu, \sigma)\}}{\partial\mu} = -\frac{1}{\sigma^2} (x - \mu)$$

を求めておきます。ここでは、推定対象は母平均 μ です。(7.8)~(7.10)式の x はサンプルから決まる定数です。従って、これまでの (x, p_m) -空間は、この具体例では母平均推定値 $\hat{\mu}$ を用いて $(\hat{\mu}, p_m)$ -空間と表記します。

R スクリプト

```
loglk <- function(mu) {x <<- x0; sigma <<- sigma; # negative log likelihood -log lk
  (log(sigma)+((x-mu)^2)/(2*sigma^2))
}
dloglk <- function(mu) { x <<- x0; sigma <<- sigma; # derivative of neg log likelihood -log lk
  -(x - mu)/(sigma^2)
}

set.seed(5315) # setting the seed of random values
x0 <- 10 # mean (sample size = 1)
sigma <- 1 # std

mu0 <- 1 # initial value for mean
N <- 100000 # number of random values to be generated

mu <- numeric(N) # memory for storing the random values of mu
U <- numeric(N) # memory for potential energy
H <- numeric(N) # memory for Hamiltonian

eps <- 0.1 # step size of Euler's method
L <- 15 # number of steps of Euler's method
accepted <- 0 # number of accepted candidates
```

51

loglk が $-\log(\text{尤度})$, dloglk が $-\log(\text{尤度})$ の微分値を定義する関数です。

```
x <<- x0
```

はグローバル変数の代入をしています。

サンプルは $x_0 = 10$ の 1 個です。母標準偏差は既知とし, $\text{sigma} = 1$ を与えています。

推定値 $\hat{\mu}$ の初期値を $\text{mu0} <- 1$ により与えています。生成乱数の数は $N <- 100000$ により, 100000 個を生成させる設定です。

$\text{mu} <- \text{numeric}(N)$ により推定値 $\hat{\mu}$ の格納先を確保しています。位置エネルギー U, ハミルトニアン H の格納先も同様に指定しています。

オイラー法のパラメータ ε と L をそれぞれ $\text{eps} <- 0.1, L <- 15$ としています。

Accepted は MH 法により受容された乱数候補数を数え上げるための変数です。

```

for (i in 1:N) {
  if (i==1){
    mu[i] <- mu0
  }else{
    mu[i] <- mu[i-1]          # initial value for mu at the start of Euler's method
  }
  pm <- rnorm(1)             # new pm at the start of Euler's method
  Hini <- pm^2 / 2 + loglk(mu[i]) # Hamiltonian of new pm and mu
  U[i] <- loglk(mu[i])       # i-th value of potential
  for (k in 2:L) {          # Euler' method
    pm <- pm - eps * dloglk(mu[i])
    mu[i] <- mu[i] + eps * pm
  }
  H[i] <- pm^2 / 2 + loglk(mu[i]) # Hamiltonian after Euler's method
  if (log(runif(1)) < Hini - H[i]){ # MH method
    accepted <- accepted + 1
  }else{
    if(i>1){                # i = 1のときは if文の条件が成立することを
      mu[i] <- mu[i-1]      # 前提としています. 乱数のシードを変える
      U[i] <- U[i-1]        # と不成立となることがあります.
      H[i] <- H[i-1]
    }
  }
}
}

```

52

i 番目の乱数候補生成に際して、 $i = 1$ では mu0 を初期値とし、以降は $i - 1$ 番目に受容された乱数を初期値とします。運動量 pm の初期値は $\text{pm} <- \text{rnorm}(1)$ により標準正規乱数を与えています。1 は生成する乱数が 1 個であることを指定しています。

$\text{Hini} <- \text{pm}^2 / 2 + \text{loglk}(\text{mu}[i])$ で初期値 $\text{mu}[i]$ と pm のハミルトニアンを計算し、 $\text{U}[i] <- \text{loglk}(\text{mu}[i])$ により位置エネルギーを計算しています。

```

for (k in 2:L){
}

```

によりオイラー法を用いて $k = L$ ステップ先に $\text{mu}[i]$ と pm を移動させます。そして、

```

H[i] <- pm^2 / 2 + loglk(mu[i])

```

により移動先のハミルトニアンを求めています。

```

if (log(runif(1)) < Hini - H[i]){

```

```

}else{

```

```

}

```

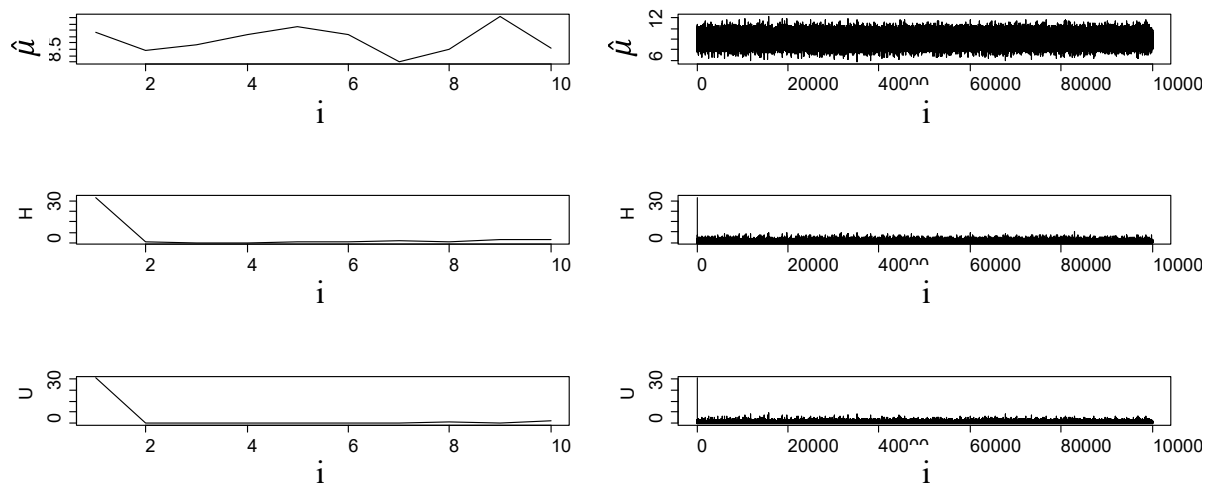
が MH 法です。 $\log(\text{runif}(1)) < \text{Hini} - \text{H}[i]$ により、[\(7.7\)式](#)の

$$\min \left(1, \frac{\exp(-H(p_m(L), x(L)))}{\exp(-H(p_m(1), x(1)))} \right)$$

を実現しています。 $\text{runif}(1)$ は $[0, 1]$ 区間の一様乱数を 1 個生成する関数です。1 は生成する乱数が 1 個であることを指定しています。ハミルトニアンが対数値であるため一様乱数も対数値としています。

生成された母平均の推定値 $\hat{\mu}$, ハミルトニアン H, 位置エネルギー U のグラフ表示

```
par(mfrow=c(3,1))
plot(mu,type="l",cex.axis=1.5,xlab="i",ylab="μ",cex.lab=1.5)
plot(H,type="l",cex.axis=1.5,xlab="i",ylab="H",cex.lab=1.5)
plot(U,type="l",cex.axis=1.5,xlab="i",ylab="U",cex.lab=1.5)
```



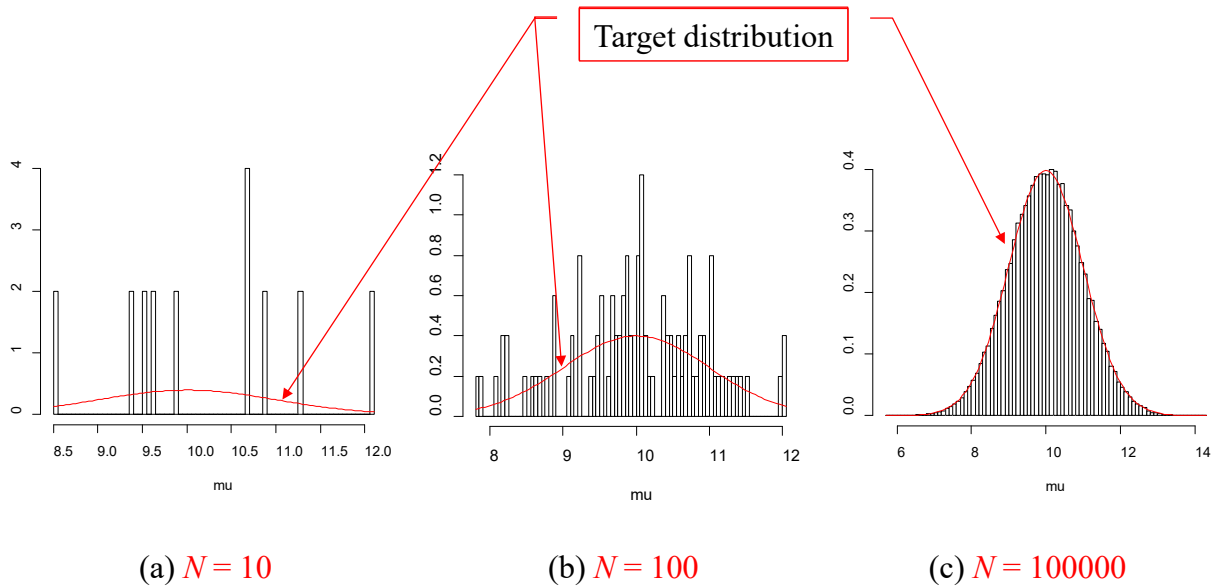
(a) $N = 10$

(b) $N = 100000$

推定値 $\hat{\mu}$ のランダムウォークの様子とハミルトニアン H, 位置エネルギー U の履歴を示します. 図(a)が最初の 10 個の乱数生成までを示し, (b)が 100000 個の乱数生成までを示します. グラフは $i = 1, 2, 3, \dots$ の離散値を直線をつないで示してあります.

生成された母平均の推定値 $\hat{\mu}$ のヒストグラム

```
par(mfrow=c(1,1))  
hist(mu, col="white", xlab="mu",ylab="",main="",freq=FALSE, breaks=100)  
x<-mu  
curve(dnorm(x, x0, sigma), add=TRUE, col="red")  
accepted # Number of accepted candidates
```



54

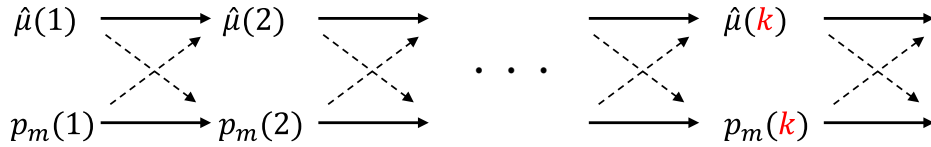
生成された母平均推定値 $\hat{\mu}$ のヒストグラムを示します。図(a)は 10 個、(b)は 100 個、(c)は 100000 個生成した段階です。参考に目標分布を赤線で示してあります。

オイラー法により, $\hat{\mu}, p_m$ は $(\hat{\mu}, p_m)$ - 空間
をどのように移動する？

55

[スライド番号 51, 52](#) のスクリプトにより, 母平均推定値 $\hat{\mu}$ の乱数を生成できることが分りました. そこで, このスクリプトを基に, $\hat{\mu}, p_m$ が $(\hat{\mu}, p_m)$ - 空間をどのように移動しているのか. そして, HMC 法のアルゴリズムでなぜ目標確率密度関数の乱数を生成できるのかを見ていきましょう.

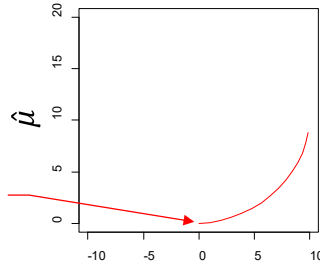
$\hat{\mu}, p_m$ の軌跡



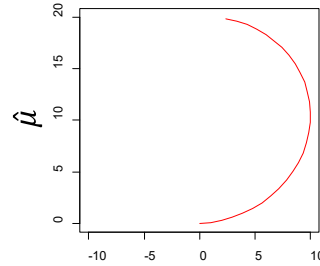
$x = 10, \sigma = 1$
 $\varepsilon = 0.1$
 オイラー法

$p_{m0} = 0, \mu_0 = 0$

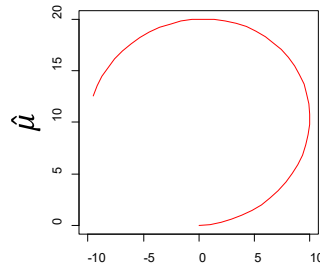
初期位置



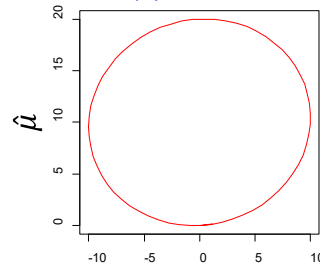
p_m
(a) $k = 15$



p_m
(b) $k = 30$



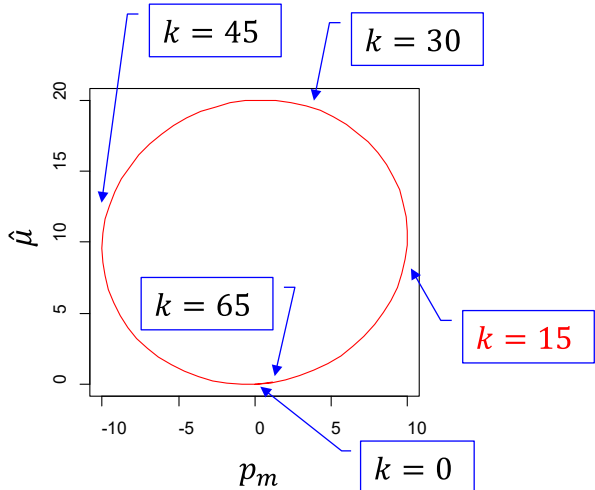
p_m
(c) $k = 45$



p_m
(d) $k = 65$

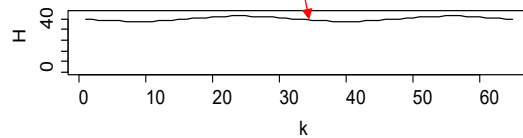
オイラー法の $k=1$ から $k=65$ までのステップにおける, $(\hat{\mu}, p_m)$ -空間内の $\hat{\mu}, p_m$ の軌跡を描きました. [スライド番号 51, 52](#) のスクリプトにおいて $N=1$ として生成する乱数を 1 個だけとし, オイラー法の箇所に k のステップ毎に $\mu[i], p_m$ を格納するスクリプトを追加しました. 図(a)は $k=15$ の場合です. 初期値は $p_m(1) = p_m = 0, \hat{\mu}(1) = \mu_0 = 0$ です. $k=15$ ステップ目では, $p_m(15) \approx 10, \hat{\mu}(15) \approx 10$ です. 図(b)は $k=30$, (c)は $k=45$, (d)は $k=65$ の場合です. k が大きいと, $\hat{\mu}, p_m$ は初期値にもどってきて, それ以降は同じ軌跡を描きます.

$x = 10, \sigma = 1$
 $\varepsilon = 0.1$, オイラー法
 $p_{m0} = 0, \mu_0 = 0$

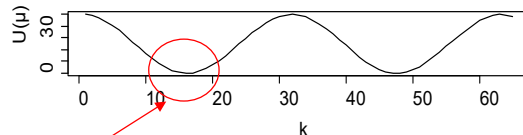


オイラー法はハミルトニアンを一定にできない。

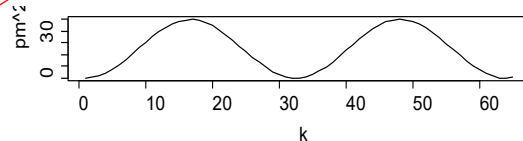
ハミルトニアン H



位置エネルギー U



運動エネルギー $p_m^2/2$



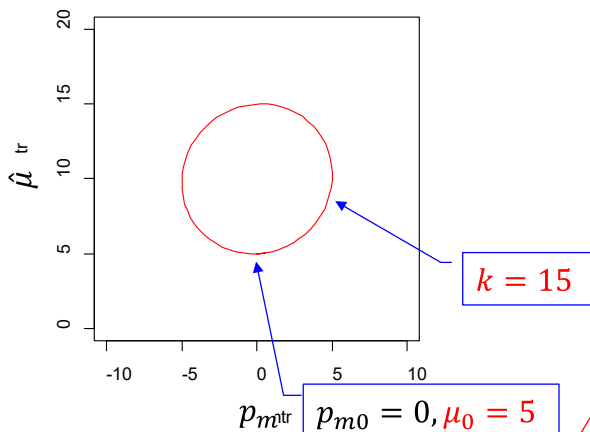
$k = 15$ 辺りで位置エネルギー U 最小。

ステップ k におけるハミルトニアン H , 位置エネルギー U , 運動エネルギー $p_m^2/2$ を記録してグラフに描きました. 左の $\hat{\mu}$, p_m の軌跡は前スライドの図(d)を再掲してあります.

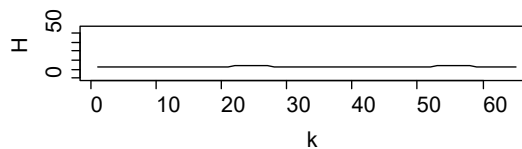
ハミルトニアン H は k とともに振動しています. [スライド番号 71](#) のエネルギー保存則に述べるように, ハミルトニアン H は一定値に保存されるべきなのですが, オイラー法では H が変動してしまいます. このハミルトニアンを一定に保てるのが後述のリープフロッグ法です.

位置エネルギー U と運動エネルギー $p_m^2/2$ は交互に増減を繰り返しています. $k = 15$ 辺りで位置エネルギー U が最小となります.

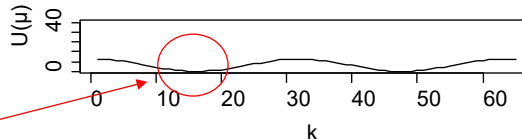
$x = 10, \sigma = 1$
 $\varepsilon = 0.1$, オイラー法
 $p_{m0} = 0, \mu_0 = 5$



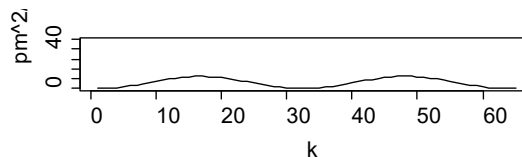
ハミルトニアン H



位置エネルギー U



運動エネルギー $p_m^2/2$



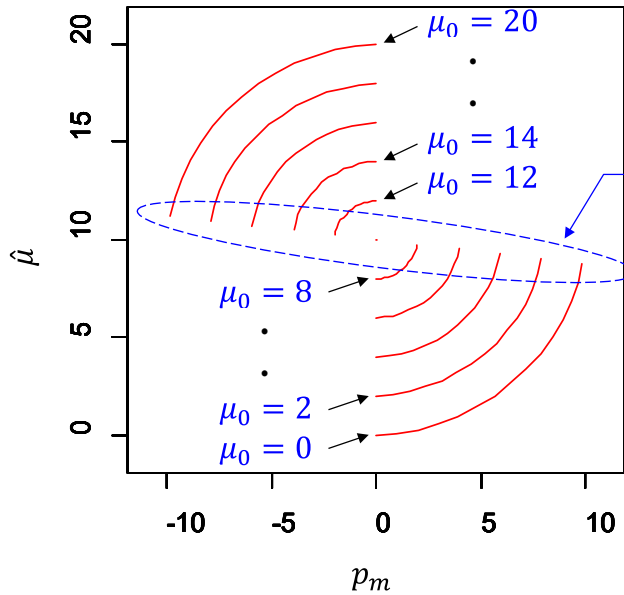
$k = 15$ 辺りで位置エネルギー U 最小.
 $U(x) = -\log p(x|\mu, \sigma)$ なので, この辺りで確率は最大.
 $L = k_{\text{max}} \approx 15$ として, ここで空間内の移動を止めるとよい.

58

初期値を $p_m(1) = p_m = 0, \hat{\mu}(1) = \mu_0 = 5$ に変えて, $\hat{\mu}, p_m$ の軌跡を描きました. 前スライドと比較すると, ハミルトニアン H の値が小さくなって, 位置エネルギー U , 運動エネルギー $p_m^2/2$ の最大値も小さくなっています. 前スライドの場合と同様に, $k = 15$ 辺りで位置エネルギー U が最小になっています. $U(\hat{\mu}) = -\log p(x|\hat{\mu}, \sigma)$ なので, この辺りで目標確率密度関数の値は最大となります.

以上より, $L = k_{\text{max}} \approx 15$ に設定して, この辺りで空間内の移動を止めるとよさそうです. [スライド番号 51](#) のスクリプトでは $L = 15$ に設定しています.

$x = 10, \sigma = 1,$
 $\varepsilon = 0.1, L = 15$
 オイラー法



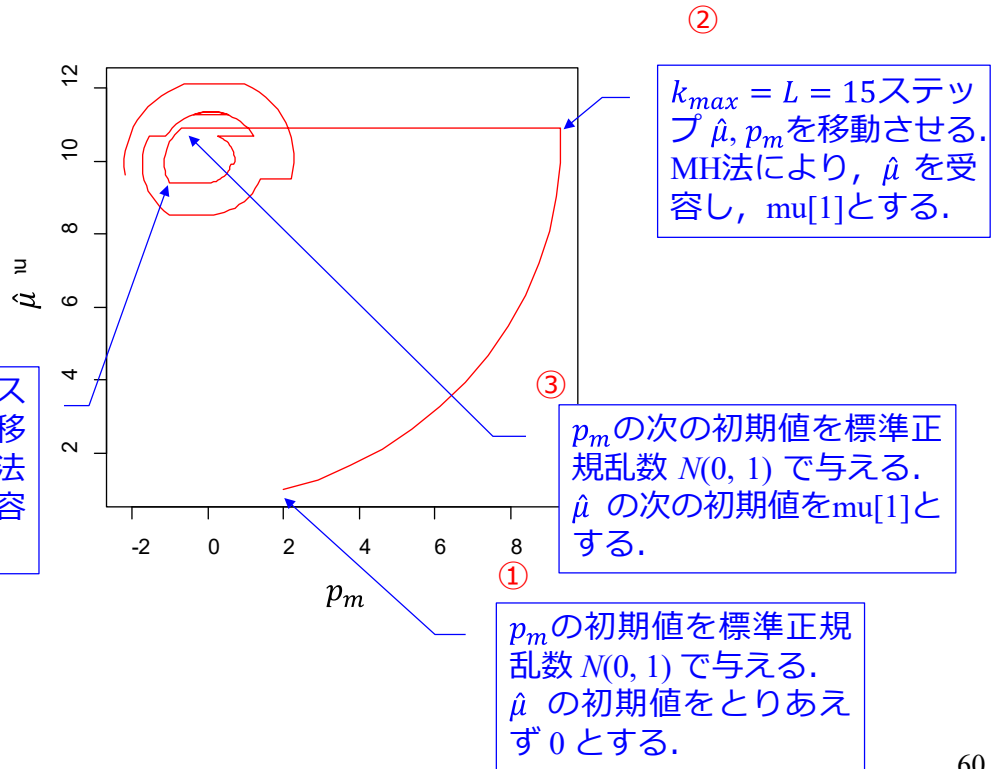
いずれの初期値からも
 $U(x) = -\log p(x|\mu, \sigma)$ を
 小さくするような移動を
 している。

$p_m(1) = p_m = 0$ に固定して, $\hat{\mu}(1) = \mu_0 = 0, 2, \dots, 20$ と与えて $k_{max} = L = 15$ ステップ先の $\hat{\mu}$ の値を
 求めてみました. いずれの初期値からでも, 位置エネルギー $U(\hat{\mu}) = -\log p(x|\hat{\mu}, \sigma)$ を小さくし, 目標確率
 密度関数 $p(x|\hat{\mu}, \sigma)$ の確率が高い辺りへと移動する結果が得られました.

オイラー法により, $(\hat{\mu}, p_m)$ -空間を, 位置エネルギー $U(\hat{\mu}) = -\log p(x|\hat{\mu}, \sigma)$ を小さくする辺りへと速
 やかに移動できることが分かりました.

ランダムウォーク (10 個の乱数生成)

$x = 10, \sigma = 1$
 $\varepsilon = 0.1, L = 15$
 $N = 10$
 オイラー法



60

スライド番号 51, 52 のスクリプトにおいて、生成乱数の個数を $N = 10$ として、 $(\hat{\mu}, p_m)$ -空間の軌跡を描きました。 $k_{max} = L = 15$ の設定です。スクリプトが実行する母平均推定値 $\hat{\mu}$ の乱数生成手順は以下の通りです。以下の①, ②, ... は図中の番号①, ②, ... と対応しています。

- ① p_m の初期値を標準正規乱数 $N(0, 1)$ で与えます。 $\hat{\mu}$ の初期値をとりあえず 0 とします。
- ② オイラー法により、 $k_{max} = L = 15$ ステップ $\hat{\mu}, p_m$ を移動させます。 MH 法により、 $\hat{\mu}$ を取捨選択します。ここでは、移動後の $\hat{\mu}$ が受容されています。これを $\text{mu}[1]$ とします。
- ③ p_m の次の初期値を標準正規乱数 $N(0, 1)$ で与えます。 $\hat{\mu}$ の次の初期値を $\text{mu}[1]$ とします。
- ④ ②と同様にして、移動・取捨選択を実行します。ここでは、移動後の $\hat{\mu}$ が受容されています。これを $\text{mu}[2]$ とします。
- ⑤ 以下同様です。

(7.6)式を再記します。

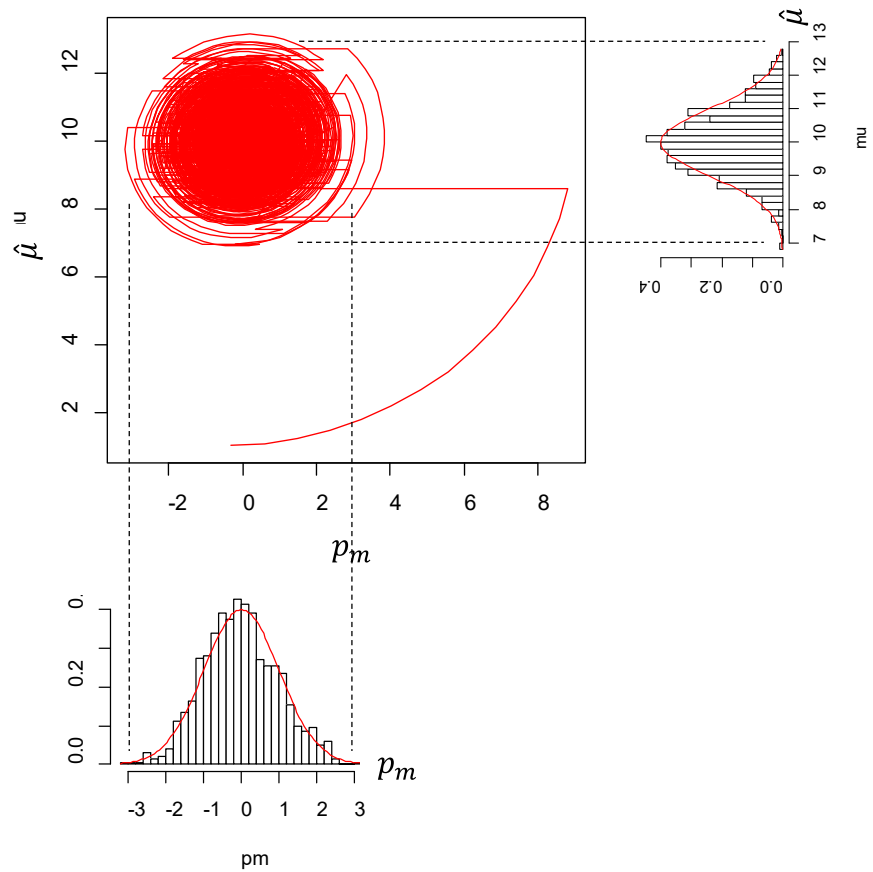
$$\frac{1}{\sqrt{2\pi}} \exp(-H(p_m, \hat{\mu})) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m}\right) p(x|\hat{\mu}, \sigma)$$

上記手順の①, ②では、 $\hat{\mu}, p_m$ の適当な初期値から $(\hat{\mu}, p_m)$ -空間の移動により、 $p(x|\hat{\mu}, \sigma)$ の確率が高い辺りへと移動しています。③以降は、 $p(x|\hat{\mu}, \sigma)$ の値が高い領域に到達後の $\hat{\mu}$ 候補の生成・取捨選択です。上式右辺の p_m に標準正規乱数を与え、 $(\hat{\mu}, p_m)$ -空間を移動させて、 $\exp(-H(p_m, \hat{\mu}))$ の値に対して

MH 法を適用して $\hat{\mu}$ の候補を取捨選択することで、 $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m}\right) p(x|\hat{\mu}, \sigma)$ の同時確率密度関数に従う乱数を得るアルゴリズムです。ここで、 $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_m^2}{2m}\right)$ と $p(x|\hat{\mu}, \sigma)$ の積には自由度があります。 p_m に標準正規乱数を与えて初期位置を振り、 $(\hat{\mu}, p_m)$ - 空間を移動させて、 $p(x|\hat{\mu}, \sigma)$ に沿う乱数 $\hat{\mu}$ を得ています。

ランダムウォーク (1000 個の乱数生成)

$x = 10, \sigma = 1$
 $\varepsilon = 0.1, L = 15$
 $N = 1000$
 オイラー法



61

前スライドの① → ② → … のプロセスを生成乱数が $N=1000$ となるまで実行した結果です. p_m の 1000 個の初期値のヒストグラムを図下に示し, 母平均推定値 $\hat{\mu}$ の 1000 個 ($\text{mu}[1] \sim \text{mu}[1000]$) のヒストグラムを図右に示してあります. $\hat{\mu}$ は, サンプル $x \approx 10$ の値を中心に, 標準偏差 $\sigma \approx 1$ (既知として, `sigma <- 1` により与えてあります) の分布となっています.

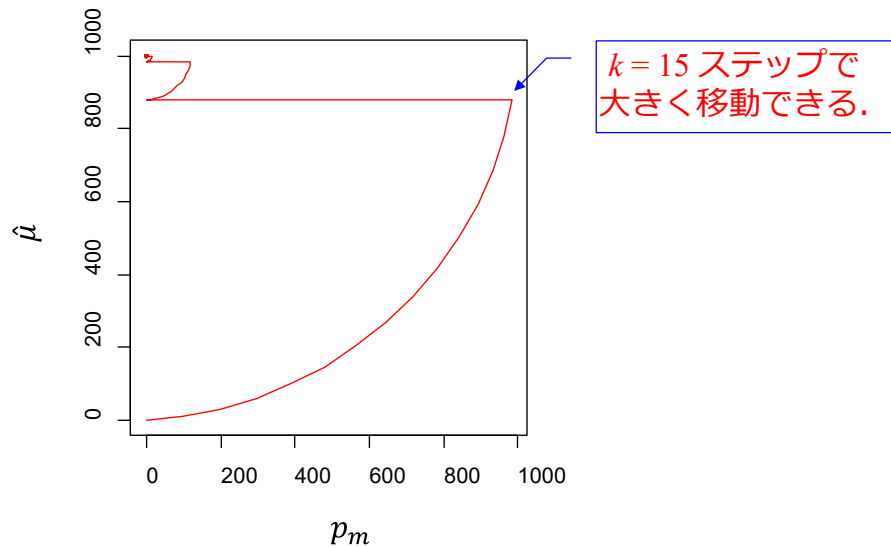
ランダムウォーク (1000 個の乱数生成)

$$x = 1000, \sigma = 1$$

$$\varepsilon = 0.1, L = 15$$

$$N = 1000$$

オイラー法



62

$x=1000$ としたときの $\hat{\mu}$, p_m の軌跡です。 $k_{max} = L = 15$ です。 オイラー法による 15 ステップの移動を 2 回繰り返すことで、 $\hat{\mu} \approx 1000$ の辺りに到達しています。 以降の 998 個の乱数は、 $x \approx 1000$ の値を中心に、標準偏差 $\sigma \approx 1$ の分布であるため、この図ではほとんど 1 点にかたまって描画されています。

ε と L の値を適切に定めることで、速やかに目標確率密度関数の確率の高い領域へと移動できました。その後は、 p_m に標準正規乱数を与えて、移動・取捨選択を繰り返すことで目標確率密度関数に沿う乱数を生成できました。

標準偏差の推定に適用

サンプル

$$X = \{0.302, 1.01, 1.598, 0.184, 0.958, 0.436, 0.283, 0.854, -1.127\}$$

$n = 9$: サンプルサイズ 9

尤度

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (7.8)$$

$-\log(\text{尤度})$

$$U(\sigma) = -\log p(x|\mu, \sigma) = \log \sigma + \frac{1}{2\sigma^2} (x - \mu)^2 \quad (7.11)$$

$-\log(\text{尤度})$ の微分値

$$\frac{\partial\{-\log p(x|\mu, \sigma)\}}{\partial\sigma} = \frac{1}{\sigma} - \frac{1}{\sigma^3} (x - \mu)^2 \quad (7.12)$$

63

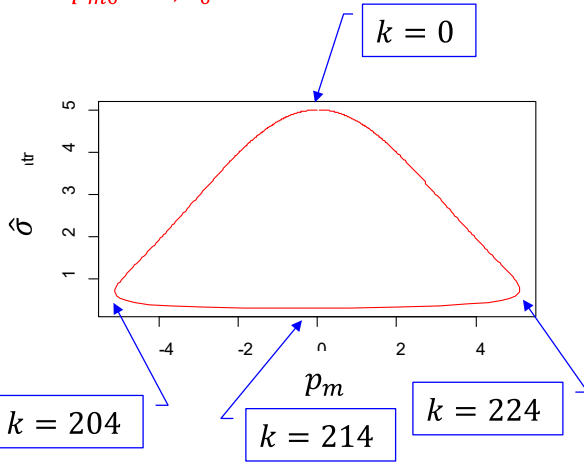
標準偏差 σ の推定に HMC 法を適用してみましょう。 [スライド番号 2](#) のデータを用います。このデータのサンプルサイズは 9 です。尤度は [\(7.8\)式](#) と同じです。位置エネルギーが [\(7.11\)式](#) です。この式は、 [\(7.9\)式](#) と形は同じですが、標準偏差 σ を引数としています。 [\(7.11\)式](#) の σ による微分値が [\(7.12\)式](#) です。

$(\hat{\sigma}, p_m)$ 空間の軌跡とハミルトニアン, 位置エネルギー, 運動エネルギーの変化の様子

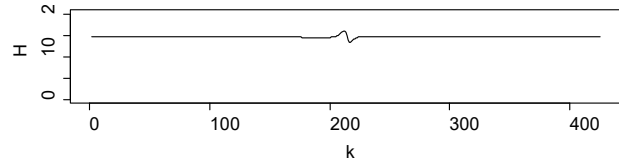
$X = \{0.302, 1.01, 1.598, 0.184, 0.958, 0.436, 0.283, 0.854, -1.127\}$

$\varepsilon = 0.01$, オイラー法

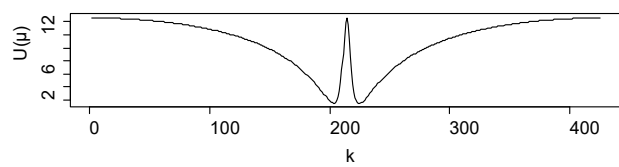
$p_{m0} = 0, \hat{\sigma}_0 = 5$



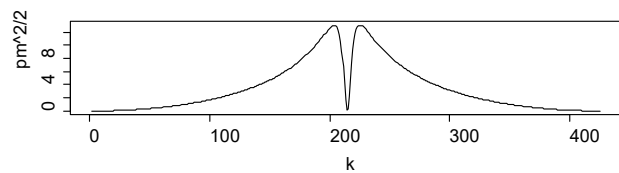
ハミルトニアン H



位置エネルギー U



運動エネルギー $p_m^2/2$

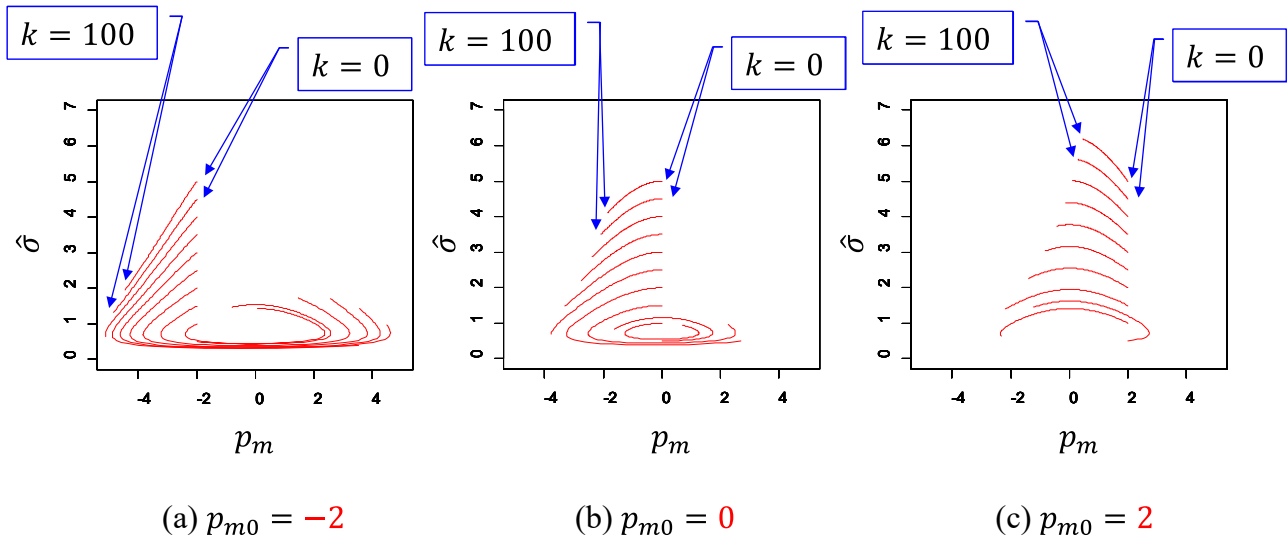


64

$(\hat{\sigma}, p_m)$ -空間の軌跡は $(\hat{\mu}, p_m)$ -空間のそれよりいびつです. ステップ幅 $\varepsilon = 0.01$ と小さくしないとあらぬ方向へと発散してしまいます. 図は初期位置を $p_{m0} = 0, \hat{\sigma}_0 = 5$ として, $k=525$ ステップまでシミュレーションを実施した結果です. $k = 214$ ステップ辺り ($p_m \approx 0$) で位置エネルギー, 運動エネルギーが大きく変化しています. ハミルトニアンにも変動が見られます. (ε をもっと小さくすると, ハミルトニアンの変動は目立たなくなります.)

(a), (b), (c)のいずれも $\hat{\sigma}_0 = 0.5, 1, 1.5, \dots, 5$ の10個の初期値からスタートした場合の $k=100$ ステップまでの軌跡

$X = \{0.302, 1.01, 1.598, 0.184, 0.958, 0.436, 0.283, 0.854, -1.127\}$
 $\varepsilon = 0.01$, オイラー法



65

初期配置を変えて、 $k=100$ ステップまでの軌跡を描きました。

(a) $p_{m0} = -2$ に固定

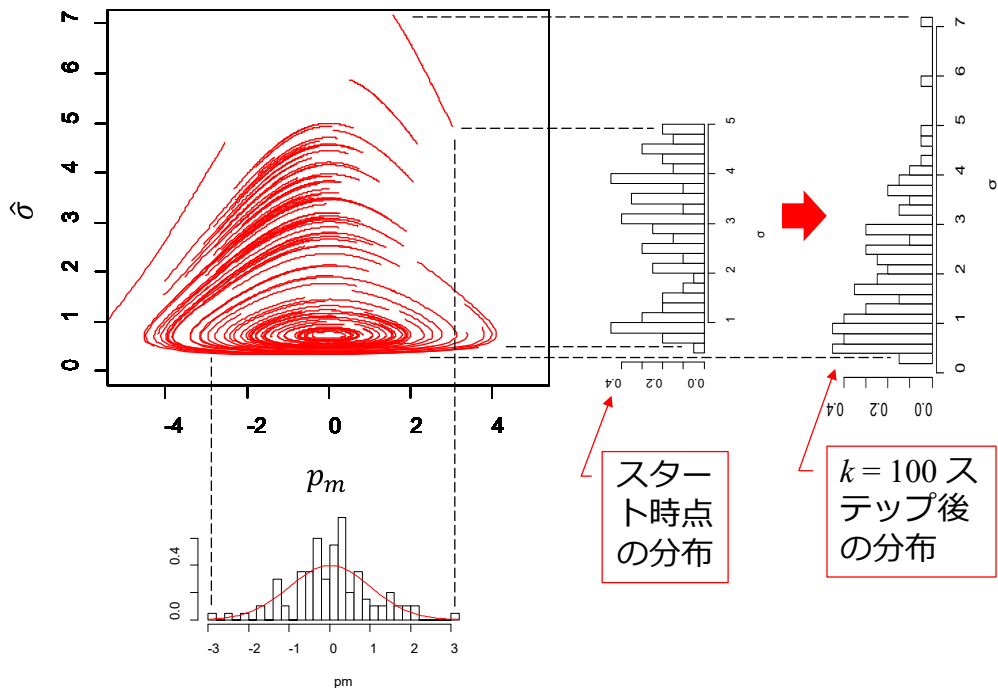
(b) $p_{m0} = 0$ に固定

(c) $p_{m0} = 2$ に固定

とし、いずれの場合も $\hat{\sigma}_0 = 0.5, 1, 1.5, \dots, 5$ の10個の初期値からスタートさせました。(a), (b)では、 $k=100$ ステップ後には $\hat{\sigma} = 1$ 付近の $p(\hat{\sigma})$ 値が高い領域に $\hat{\sigma}$ の値が集まる傾向にあることが見て取れます。一方、(c)では多くの初期値において、 $\hat{\sigma}$ の値が $p(\hat{\sigma})$ 値が高い領域から離れる傾向にあります。

$\hat{\sigma}_0 \sim \text{uniform}(0.5, 5)$, $p_m \sim N(0, 1)$ の100個の初期値からスタートした場合の $k = 100$ ステップまでの軌跡と、スタート時点および $k = 100$ ステップ後の $\hat{\sigma}$ の分布

$X = \{0.302, 1.01, 1.598, 0.184, 0.958, 0.436, 0.283, 0.854, -1.127\}$
 $\varepsilon = 0.01$, オイラー法



そこで、 $\hat{\sigma}_0 \sim \text{uniform}(0.5, 5)$, $p_m \sim N(0, 1)$ の 100 個の初期値を生成して、 $k = 100$ ステップ後の $\hat{\sigma}$ 値の分布を求めてみました。 $\text{uniform}(0.5, 5)$ は区間 $[0.5, 5]$ の一様乱数、 $N(0, 1)$ は標準正規乱数です。100本の軌跡を図中央に示します。図下は p_m 100個のヒストグラムです。参考に標準正規分布を赤線で示してあります。図右に $\hat{\sigma}$ のスタート時点における初期値の分布と $k = 100$ ステップ後の分布を示します。100個と少ないために、スタート時点の分布は一様乱数としてはばらつきが大きいのですが、100ステップの移動による分布の変化を見ることはできます。これより、オイラー法による $\hat{\sigma}$, p_m 点の移動は、 $p(\hat{\sigma})$ 値が高い領域に $\hat{\sigma}$ の値を集める傾向にあることが分ります。

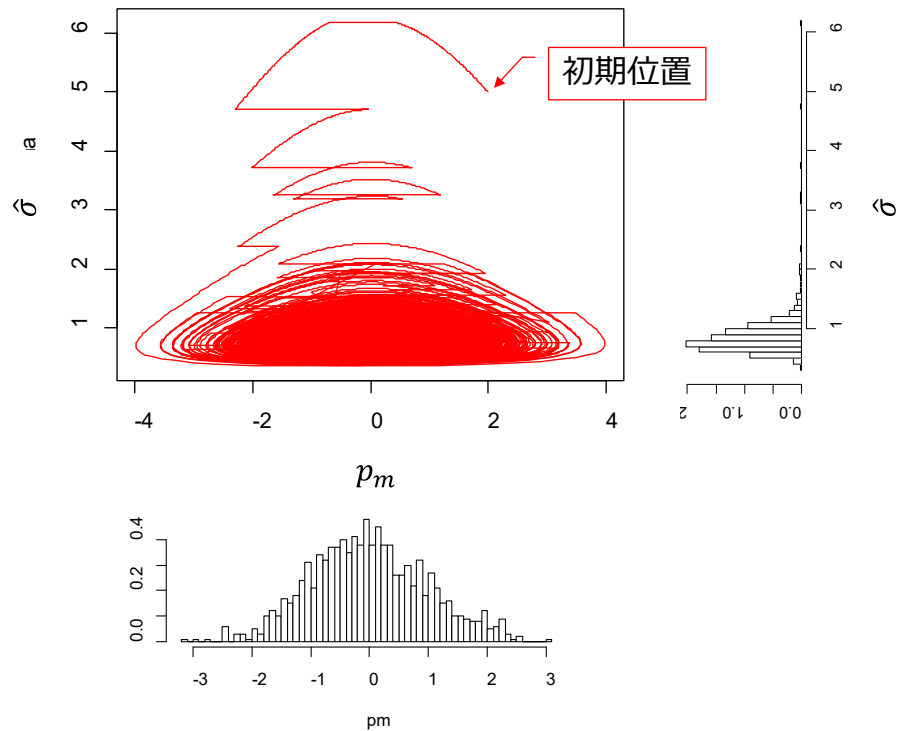
標準偏差の推定 ランダムウォーク (1000 個の乱数生成)

$X = \{0.302, 1.01, 1.598, 0.184, 0.958, 0.436, 0.283, 0.854, -1.127\}$

$\varepsilon = 0.01, L = 100$

$N = 1000$

オイラー法



67

HMC 法によるランダムウォークの軌跡と 1000 個の標準正規乱数 p_m および得られた標準偏差推定値 $\hat{\sigma}$ のヒストグラムを示します。初期値 $\hat{\sigma} = 5$ としました。スライド番号 61, 62 の例と異なり、 $p(\hat{\sigma})$ 値が高い領域への移動に、オイラー法による $k_{max} = 100$ ステップの移動を数回要しています。 $p(\hat{\sigma})$ 値が高い領域へと到達後は、 $p(\hat{\sigma})$ に沿う乱数 $\hat{\sigma}$ を得ています。

右の推定値 $\hat{\sigma}$ のヒストグラムは、面積が 1 となるように規格化されているので、 $\hat{\sigma}$ の確率密度関数です。

面積保存の性質

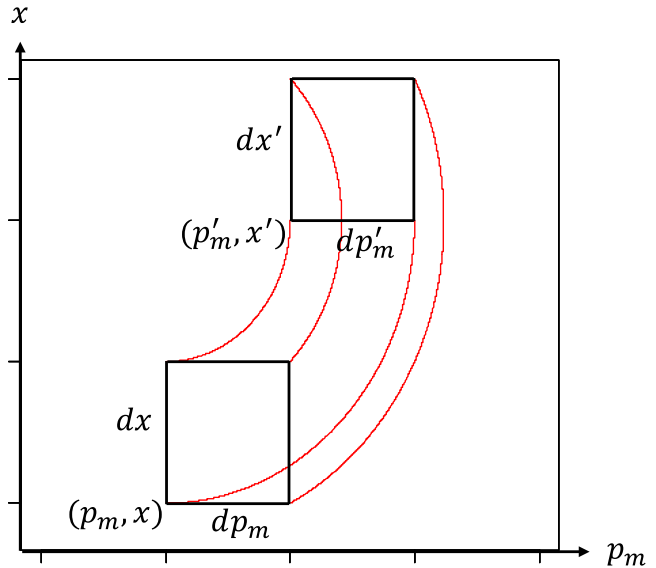
時刻 t にて (p_m, x) に微小面積 $dp_m dx$ があったとする. (7.1), (7.2)式に従って微小面積が移動し, 時刻 $t + \Delta t$ にて (p'_m, x') に到達して, そのときの面積が $dp'_m dx'$ になったとする.

$$\begin{aligned} p'_m &= p_m + \frac{dp_m}{dt} \Delta t \\ &= p_m - \frac{\partial H}{\partial x} \Delta t \\ x' &= x + \frac{dx}{dt} \Delta t \\ &= x + \frac{\partial H}{\partial p_m} \Delta t \end{aligned} \quad (7.13)$$

この式を, 面積分の微小面積素の変数変換と見なせば

$$dp'_m dx' = \left| \frac{\partial(p'_m, x')}{\partial(p_m, x)} \right| dp_m dx \approx dp_m dx \quad (7.14)$$

となる. $\left| \frac{\partial(p'_m, x')}{\partial(p_m, x)} \right|$ はヤコビアンである.



(x, p_m) -空間を式(7.1), (7.2)に従い移動する点 (x, p_m) における微小面積 $dp_m dx$ は保存される.

65

一端, 表記を (x, p_m) -空間に戻します. (x, p_m) -空間における x, p_m の移動において, 2つの重要な性質について解説します.

[\(7.1\)](#), [\(7.2\)式](#)に従い移動する質点周りの微小面積には「面積保存の性質」があります.

点 (p_m, x) を左下角の頂点とし, 辺の長さが dp_m, dx の長方形を考えます. この長方形の4つの頂点が [\(7.1\)](#), [\(7.2\)式](#)に従い移動して, Δt 秒後に左下角が頂点 (p'_m, x') に到達し, 各辺の長さが dp'_m, dx' になったとします.

微分方程式を

$$\begin{aligned} \frac{dp_m}{dt} &\approx \frac{p'_m - p_m}{\Delta t} \\ \frac{dx}{dt} &\approx \frac{x' - x}{\Delta t} \end{aligned}$$

により近似し, 変形すると

$$\begin{aligned} p'_m &= p_m + \frac{dp_m}{dt} \Delta t \\ &= p_m - \frac{\partial H}{\partial x} \Delta t \end{aligned}$$

$$\begin{aligned}
 x' &= x + \frac{dx}{dt} \Delta t \\
 &= x + \frac{\partial H}{\partial p_m} \Delta t
 \end{aligned}$$

が得られます. (7.3), (7.4)式 of オイラー法の近似と同じです. これらの式を, 面積分の微小面積素の変数変換と見なせば

$$\begin{aligned}
 dp'_m dx' &= \left| \frac{\partial(p'_m, x')}{\partial(p_m, x)} \right| dp_m dx \\
 &\approx dp_m dx
 \end{aligned} \tag{7.14}$$

となります. $\left| \frac{\partial(p'_m, x')}{\partial(p_m, x)} \right|$ はヤコビアンです.

(7.14)式は, (x, p_m) -空間を(7.1), (7.2)式に従い移動する点 (x, p_m) の微小面積 $dp_m dx$ が保存されることを意味します.

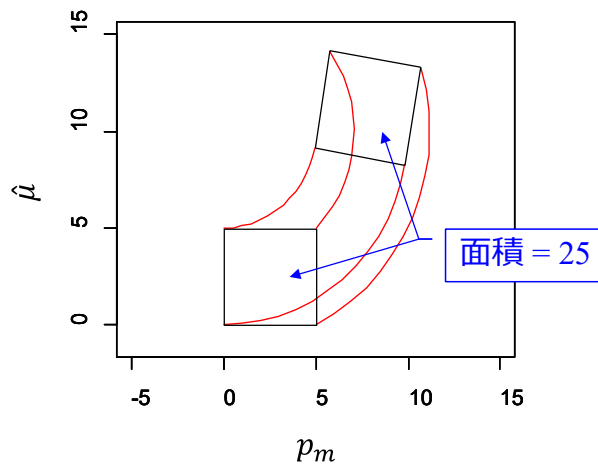
ヤコビアンは

$$\begin{aligned}
 \frac{\partial(p'_m, x')}{\partial(p_m, x)} &= \begin{vmatrix} \frac{\partial p'_m}{\partial p_m} & \frac{\partial x'}{\partial p_m} \\ \frac{\partial p'_m}{\partial x} & \frac{\partial x'}{\partial x} \end{vmatrix} \\
 &= \begin{vmatrix} 1 - \frac{\partial^2 H}{\partial p_m \partial x} \Delta t & \frac{\partial^2 H}{\partial p_m^2} \Delta t \\ -\frac{\partial^2 H}{\partial x^2} \Delta t & 1 + \frac{\partial^2 H}{\partial x \partial p_m} \Delta t \end{vmatrix} \\
 &= 1 + \left(\frac{\partial^2 H}{\partial x \partial p_m} - \frac{\partial^2 H}{\partial p_m \partial x} \right) \Delta t + (\Delta t^2 \text{の項}) \\
 &= 1 + (\Delta t^2 \text{の項}) \\
 &\approx 1
 \end{aligned}$$

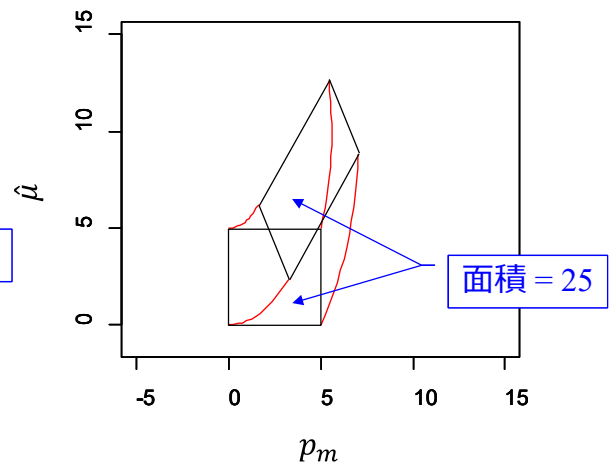
より, Δt が十分に小さければ 1 と見なせます.

面積保存の例

$x = 10, \sigma = 1$
 $\varepsilon = 0.1, L = 15$
オイラー法



$x = 10, \sigma = 2$
 $\varepsilon = 0.1, L = 15$
オイラー法

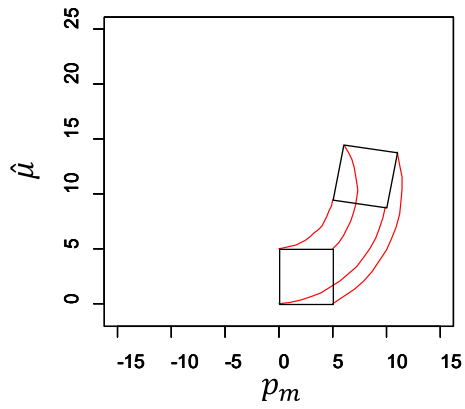


66

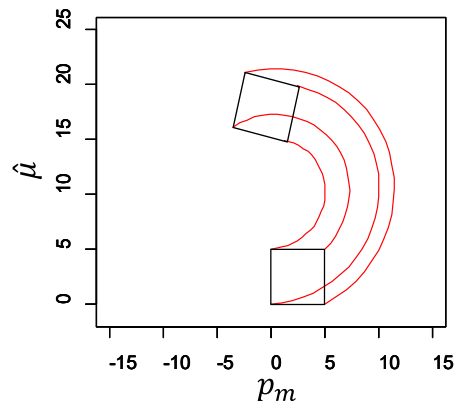
再び、表記を $(\hat{\mu}, p_m)$ -空間として、面積保存の様子をシミュレーションにより示します。 [スライド番号 51, 52](#) のスクリプトにおいて、 $N=4$ とし、 $(0, 0), (5, 0), (0, 5), (5, 5)$ の 4 箇所から出発して $k=15$ ステップ移動させた結果です。左が $\text{sigma} = \sigma = 1$ 、右が $\text{sigma} = \sigma = 2$ の設定です。左右いずれの場合も、4 箇所の初期値が作る正方形の面積が 25 で、15 ステップ先の平行四辺形の面積も 25 です。

この面積保存の性質が意味することは、[\(7.1\), \(7.2\)式](#) による $(\hat{\mu}, p_m)$ -空間内の移動は、正方形内の任意の点（乱数）の移動先を平行四辺形内に留まらせ、**どこかに散逸させたり、消滅させたりしない**ことです。すなわち、初期の四角形内に生成された乱数は全て平行四辺形内へと移動させられます。また、このことから、初期位置の異なる乱数の軌跡が交わることはありません。

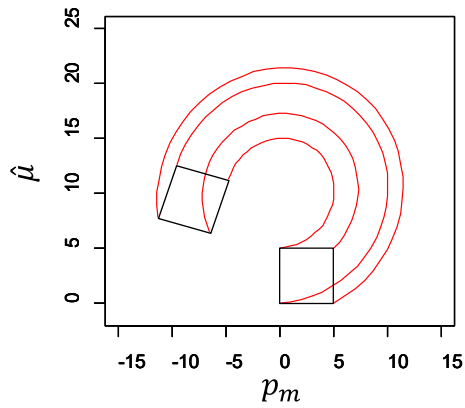
面積保存の例 $x = 10, \sigma = 1, \varepsilon = 0.1$, オイラー法



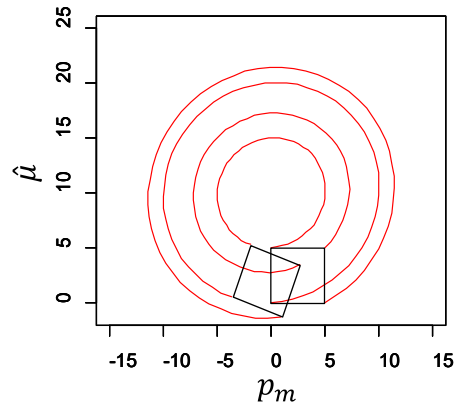
(a) $k = 15$



(b) $k = 30$



(a) $k = 45$



(b) $k = 60$

67

図は、面積保存の別の例示です。4つの頂点を作る4角形の面積は不変です。4つの頂点はそれぞれ交わることのない軌跡を描いています。

エネルギー保存則

$$\begin{aligned}\frac{\partial H}{\partial t} &= \frac{\partial H}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial H}{\partial p_m} \frac{\partial p_m}{\partial t} \\ &= \frac{\partial H}{\partial x} \frac{\partial H}{\partial p_m} - \frac{\partial H}{\partial p_m} \frac{\partial H}{\partial x} \\ &= 0\end{aligned}\tag{7.15}$$

(x, p_m) -空間を式(7.1), (7.2)に従い移動する x, p_m はハミルトニアン H を一定にする. **エネルギー保存則**

68

(x, p_m) -空間内を(7.1), (7.2)式に従い移動する x, p_m に関するもう一つの重要な性質です. この質点のハミルトニアンの時間微分は

$$\frac{\partial H}{\partial t} = \frac{\partial H}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial H}{\partial p_m} \frac{\partial p_m}{\partial t}$$

となります. 右辺の, $\frac{\partial x}{\partial t}, \frac{\partial p_m}{\partial t}$ が(7.1), (7.2)式を満たすので

$$\begin{aligned}\frac{\partial H}{\partial t} &= \frac{\partial H}{\partial x} \frac{\partial H}{\partial p_m} - \frac{\partial H}{\partial p_m} \frac{\partial H}{\partial x} \\ &= 0\end{aligned}$$

と得られます. (x, p_m) -空間を移動する x, p_m はハミルトニアン H を一定にします. これは**エネルギー保存則**です.

リープフロッグ法

エネルギー保存則を満たす (x, p_m) -空間内の移動法

69

[スライド番号 57](#) のハミルトニアン H は振動していました。オイラー法の欠点はハミルトニアン H を一定にできません。これまでは、簡便なオイラー法を用いて HMC 法における (x, p_m) -空間内の移動について解説してきました。しかし、オイラー法はこの欠点のためベイズ統計の有力ツールである Stan では使われていません。

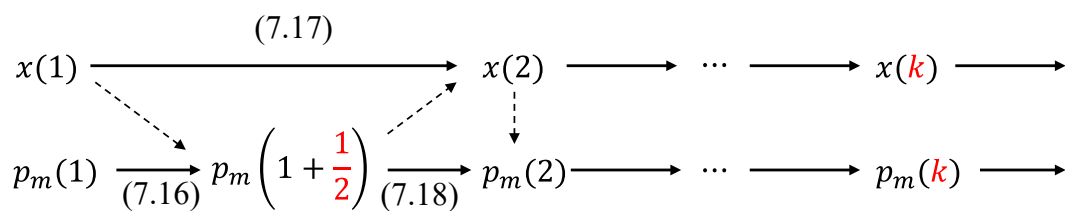
Stan で使われている (x, p_m) -空間内の移動法がリープフロッグ法です。この方法による x, p_m の移動では、ハミルトニアン H が一定に保たれます。

リープフロッグ法

$$p_m \left(k + \frac{1}{2} \right) = p_m(k) - \frac{\varepsilon}{2} \frac{\partial H(x(k))}{\partial x} \quad (7.16)$$

$$x(k+1) = x(k) + \varepsilon p_m \left(k + \frac{1}{2} \right) \quad (7.17)$$

$$p_m(k+1) = p_m \left(k + \frac{1}{2} \right) - \frac{\varepsilon}{2} \frac{\partial H(x(k+1))}{\partial x} \quad (7.18)$$



リープフロッグ法のイメージ

Leapfrog 法を直訳すると「カエル跳び」法です。リープフロッグ法は(7.16)～(7.18)式で表されます。同法による x , p_m 更新のイメージを下図に示します。スライド番号 47 のオイラー法と比較すると、リープフロッグ法では p_m のステップに中間点 $p_m \left(1 + \frac{1}{2} \right)$ が設けられています。

```

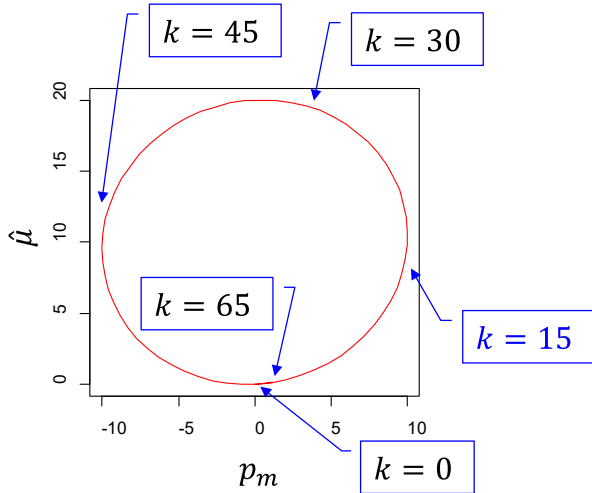
for (i in 1:N) {
  if (i==1){
    mu[i] <- mu0
  }else{
    mu[i] <- mu[i-1]           # initial value for mu at the start of Euler's method
  }
  pm <- rnorm(1)              # new pm at the start of Euler's method
  Hini <- pm^2 / 2 + loglk(mu[i]) # Hamiltonian of new pm and mu
  U[i] <- loglk(mu[i])        # i-th value of potential
  for (k in 2:L) {           # leapfrog method
    pm = pm - eps * dloglk(mu[i])/2
    mu[i] = mu[i] + eps * pm
    pm = pm - eps * dloglk(mu[i])/2
  }
  H[i] <- pm^2 / 2 + loglk(mu[i]) # Hamiltonian after Euler's method
  if (log(runif(1)) < Hini - H[i]){ # MH method
    accepted <- accepted + 1
  }else{
    if(i>1){                 # i = 1のときは if 文の条件が成立することを
      mu[i] <- mu[i-1]       # 前提としています. 乱数のシードを変える
      U[i] <- U[i-1]         # と不成立となることがあるかも知れません.
      H[i] <- H[i-1]
    }
  }
}
}

```

71

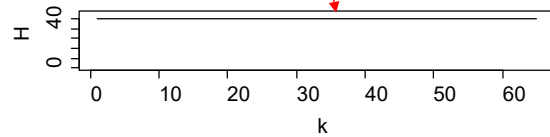
[スライド番号 52](#) のオイラー法のスクリプトをリープフロッグ法に書き換えました.

$x = 10, \sigma = 1$
 $\varepsilon = 0.1$, リープフロッグ法
 $p_{m0} = 0, \mu_0 = 0$

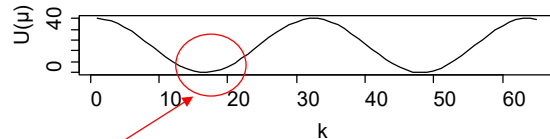


リープフロッグ法はハミルトニアンを一定に保つ.

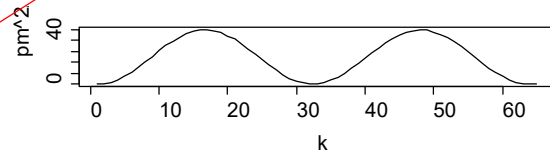
ハミルトニアン H



位置エネルギー U



運動エネルギー $p_m^2/2$



$k = 15$ 辺りで位置エネルギー U 最小.

リープフロッグ法により, $\hat{\mu}, p_m$ の軌跡, ハミルトニアン H , 位置エネルギー U , 運動エネルギー $p_m^2/2$ を求めました. [スライド番号 57](#) のオイラー法の場合と異なり, ハミルトニアン H は k に対して不変です. ハミルトニアン以外はオイラー法の場合と見かけ上変わりません.

MH 法では

```
if (log(runif(1)) < Hini - H[i]){
    ...
} else {
    ...
}
```

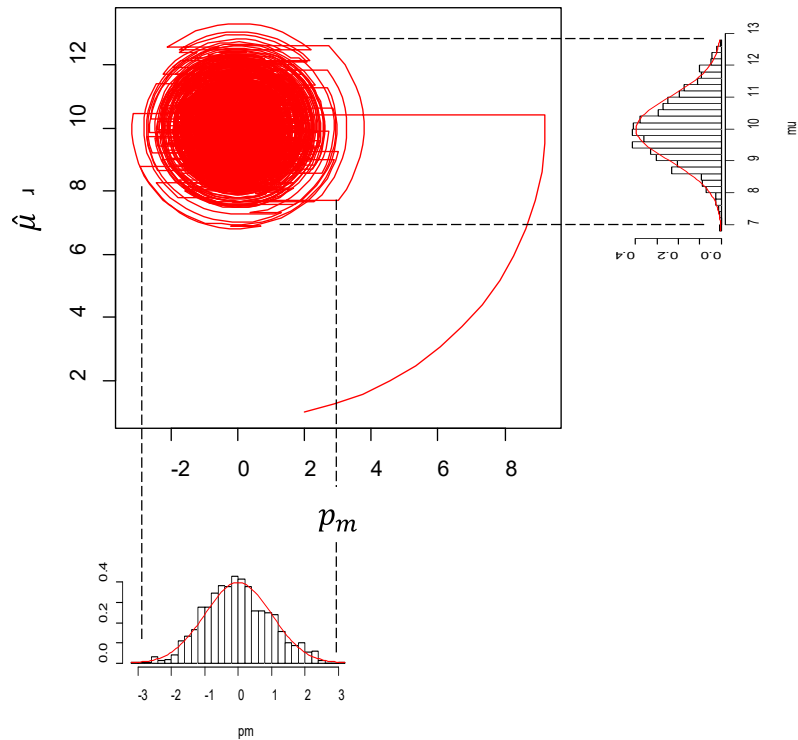
により, 移動前後のハミルトニアン之比 (対数変換後では差分) に応じて $\hat{\mu}$ 候補の取捨選択を行っています. リープフロッグ法では, この移動前後でハミルトニアン値が一定なので, 必ず移動先の $\hat{\mu}$ 候補が選択されます. すると, この if 文は要らないのではと疑問が湧きますが, これは数値計算によりハミルトニアンに誤差が生じた場合の対処用です.

リープフロッグ法, ランダムウォーク (1000個の乱数生成)

$$x = 10, \sigma = 1$$

$$\varepsilon = 0.1, L = 15$$

$$N = 1000$$



73

[スライド番号 71](#) のスクリプトを実行して, 母平均推定値 $\hat{\mu}$ の乱数を生成した結果と, 生成過程における $\hat{\mu}, p_m$ の軌跡です. [スライド番号 61](#) のオイラー法による結果と見かけ上の差はほとんどありません.

Hを一定に保てると何がうれしい?

MH法: L ステップ後の候補を受容する確率

$$\min\left(1, \frac{\exp(-H(p_m(L), x(L)))}{\exp(-H(p_m(1), x(1)))}\right)$$



受容の確率が高くなる.

例. 受容率

オイラー法: .972 (乱数のシードにより変化)

リープフロッグ法: 1.0 $\left(\begin{array}{l} n = 1, x = 10, \sigma = 1 \\ \varepsilon = 0.1, L = 15, N=1000 \end{array} \right)$



$p(\hat{\mu})$ に従う乱数をオイラー法よりも効率よく生成できた.

74

リープフロッグ法とオイラー法との差は $\hat{\mu}$ 候補の受容率にあります. 受容率を新規に生成された乱数候補の中で受容された候補の比率とします. [スライド番号 61](#) と [73](#) の 1000 個の乱数生成において, オイラー法では 972 個が受容され, リープフロッグ法では 1000 個全てが受容されました. したがって, オイラー法での受容率は 0.972, リープフロッグ法でのそれは 1.0 でした. このシミュレーション例では, リープフロッグ法がハミルトニアン H を一定に保つことで, $\hat{\mu}$ 候補が全て受容されました.

リープフロッグ法はオイラー法よりも効率よく乱数生成ができます.

RStan --R interface for Stan--

ネットから無料でダウンロード可

<https://mc-stan.org/users/interfaces/rstan>

R + Rstudio + package “rstan”

PyStan --Python interface for Stan--

<https://pystan.readthedocs.io/en/latest/>

78

References

- [1] R. A. Fisher, “Statistical Method for Research Workers 13th edition,” Oliver and Boyd, 1958
- [2] N. Metropolis, et al., “Equation of State Calculations by Fast Computing Machines,” J. of Chemical Physics, Vol. 21, No. 6, pp.1087-1092, 1953
- [3] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” Biometrika, Vol.57, No.1, pp.97-109, 1970
- [4] S. Duane, et al., “Hybrid Monte Carlo,” Physics Letters B, Vol. 195, No. 2, pp.216-222, 1987
- [5] A. E. Gelfand and A. F. M. Smith, “Sampling-Based Approaches to Calculating Marginal Densities,” J. of American Statistical Association, Vol.85, No. 410, pp.398-409, 1990
- [6] 豊田秀樹「基礎からのベイズ統計学 ―ハミルトニアンモンテカルロ法による実践的入門―」朝倉書店, 2015
- [7] 豊田秀樹「はじめての統計データ分析 ―ベイズ的<ポストp値時代>の統計学―」朝倉書店, 2016
- [8] 豊田秀樹「心理統計法 ―有意性検定からの脱却―」放送大学教育振興会, 2017
- [9] 奥村晴彦「Rで楽しむベイズ統計入門 ―しくみから理解するベイズ推定の基礎」技術評論社, 2018
- [10] John K. Kruschke, “Doing Bayesian Data Analysis 2nd edition,” Elsevier, 2015
- [11] A. Gelman, et al., “Bayesian Data Analysis 3rd edition,” CRC Press, 2014

79

2019年8月

著者：古橋武

所属：名古屋大学工学研究科情報・通信工学専攻

連絡先：furuhashi@nuee.nagoya-u.ac.jp