第11章　サンプリング法

B4　近松京介
原さん(東北大助教授)
ドクターのときの原さんのパワポ改良ましまし版

11章の内容

- 基本的なサンプリングアルゴリズム
  - 棄却サンプリング・適応的棄却サンプリング
  - 重点サンプリング
  - SIR
  - サンプリングとEMアルゴリズム
  - データ拡大アルゴリズム
- マルコフ連鎖モンテカルロ
  - Metropolis-Hastingsアルゴリズム
- ギブスサンプリング
- スライスサンプリング
- ハイブリッドモンテカルロアルゴリズム

目次

- MCMCを使わないサンプリング
  - 逆関数法
  - 棄却サンプリング
  - 適応的棄却サンプリング
  - 重点サンプリング
  - SIR
- MCMCサンプリング
  - メトロポリス法
  - ギブスサンプラー

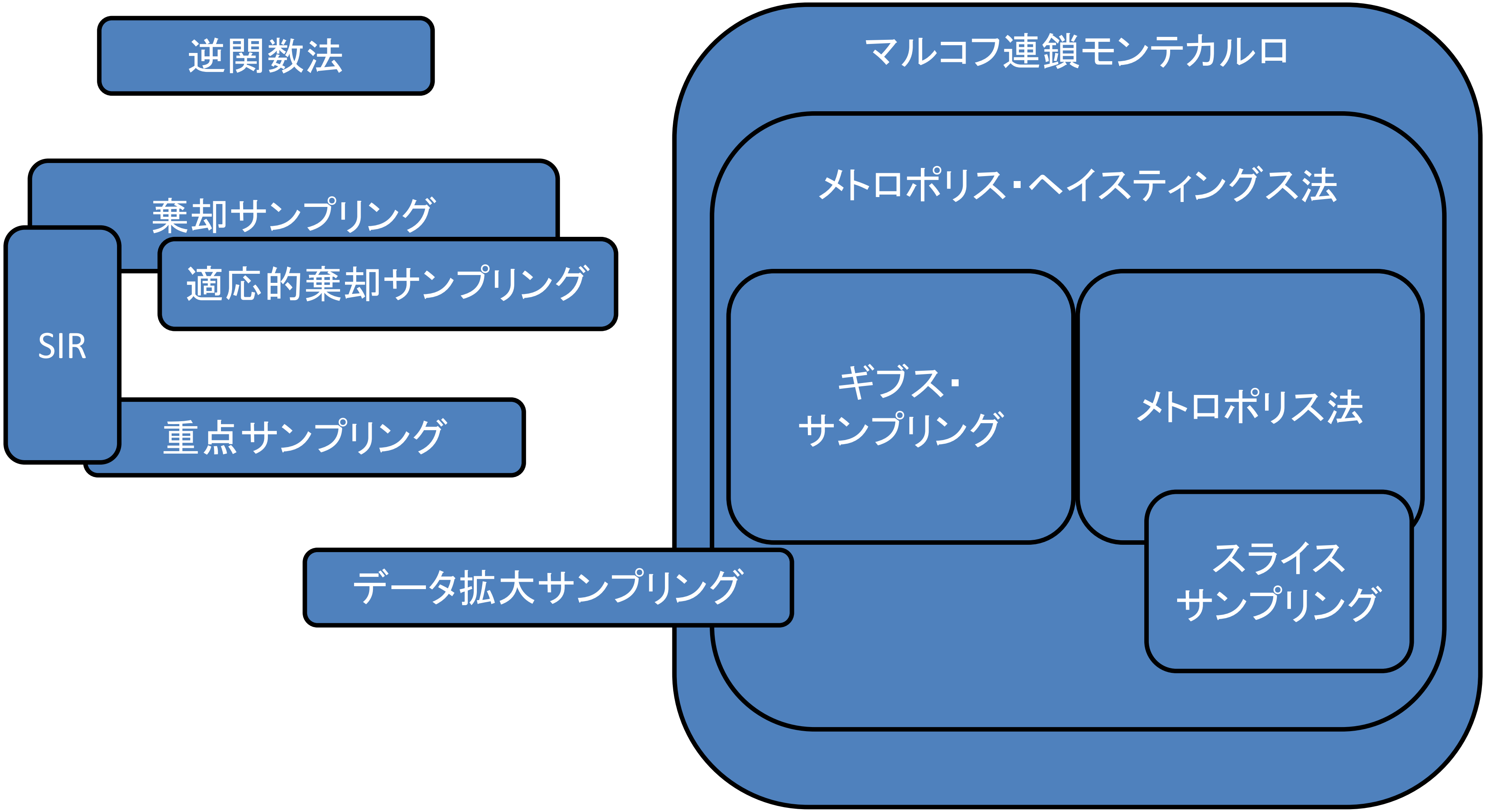
今回の目的

- そもそもサンプリングとはなにか理解する
- サンプリング手法を概観・整理

- **メトロポリス法**
- **ギブスサンプラー**

実装できそうと思うくらい理解する

全体像を俯瞰


逆関数法
棄却サンプリング
適応的棄却サンプリング
SIR
重点サンプリング
データ拡大サンプリング
マルコフ連鎖モンテカルロ
メトロポリス・ヘイスティングス法
ギブス・サンプリング
メトロポリス法
スライスサンプリング

サンプリング法の説明

- 目的は「ある想定した分布」に従う乱数を発生させること（サンプリングすること）
- すなわち分布を再現すること
- たとえばある出発地に1000人いて，目的地として選ばれやすい座標が確率分布として与えられたときにその1000人がどこにいくか
- 正規分布や一様分布ならできるかも？
- より複雑な分布は…？

逆関数法

- まずは単純な逆関数法から
- 一様分布からサンプリングする（乱数を発生させる）ことができることが前提

$$z \rightarrow h^{-1}(z) \rightarrow p(y)$$

一様乱数

求めたい分布の
不定積分の
逆関数

求めたい分布から
サンプリングした乱数

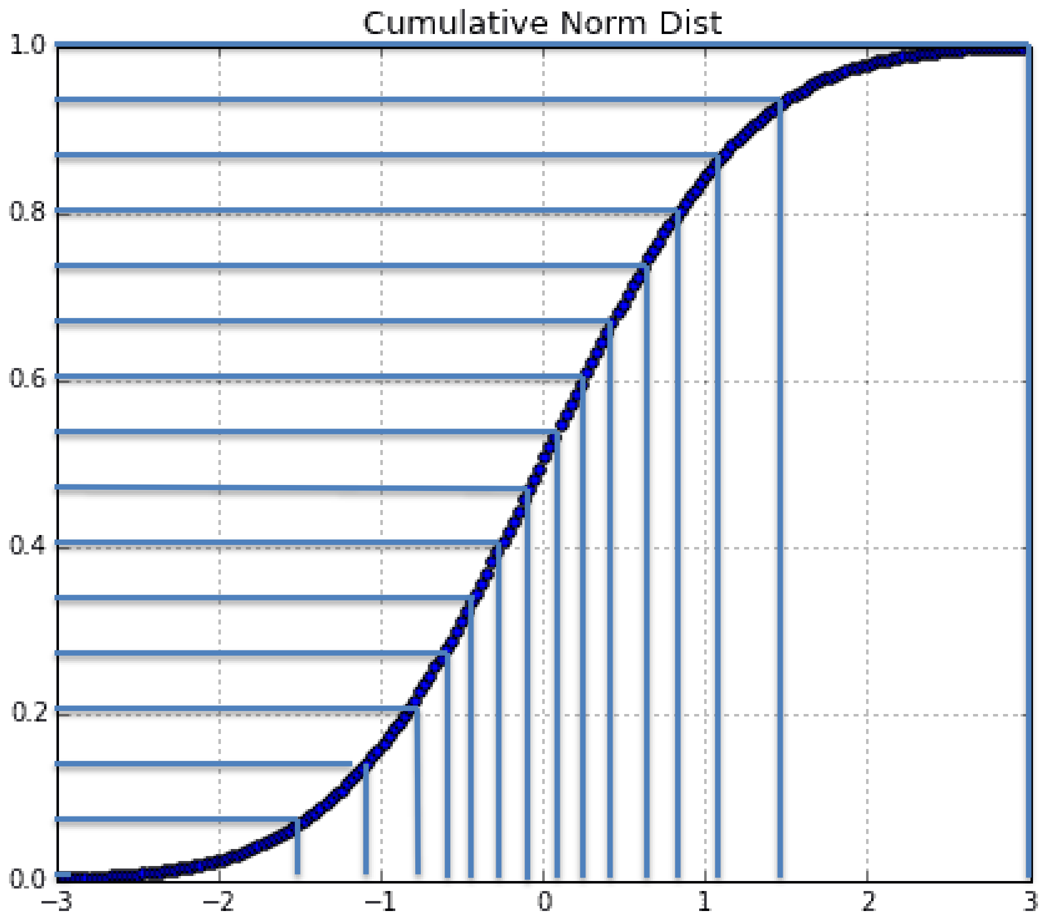
こうならしいが，イメージがわきません

逆関数法

正規分布の累積関数

イメージわきましたか？

一様サンプリング



なんか正規分布っぽい

逆関数法

z → $h^{-1}(z)$ → $p(y)$

一様乱数

求めたい分布の
不定積分の
逆関数

求めたい分布から
サンプリングした乱数

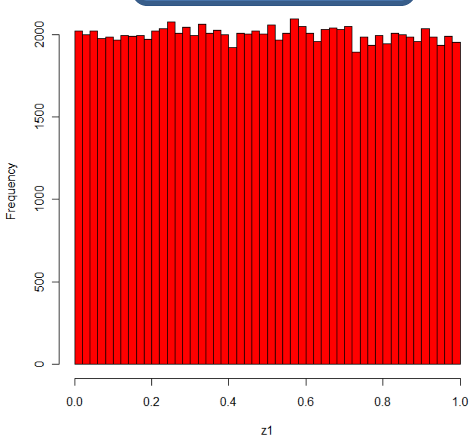
- たとえば指数分布からサンプリングしたいが直接できないとき

$$p(y) = \lambda \exp(-\lambda y) \qquad z = h(y) \equiv \int_{-\infty}^{y} p(\hat{y}) d\hat{y} \qquad z = h(y) = 1 - \exp(-\lambda y)$$

よって，次の逆関数が得られる

$$h^{-1}(z) = -\lambda^{-1} \ln(1 - z)$$

乱数発生

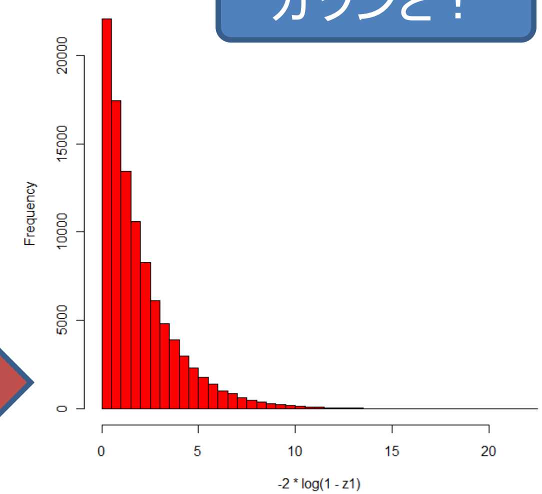


[0,1]の一様乱数zを
100000個発生

$$y = h^{-1}(z) = -2 \ln(1 - z)$$

にぶち込む

ガツンと！



棄却サンプリング

- 逆関数が求められない比較的複雑な分布だってある
- やや複雑な分布p(z)からサンプリングしたいが，直接p(z)からサンプリングするのは困難であるとする
- 任意のzが与えられたとき，正規化定数Zを除いたp(z)を求めることは容易であるとしよう（これはよくあること）

$$p(z) = \frac{1}{Z_p}\tilde{p}(z) \qquad \tilde{p}(z)\text{はわかるが}\ Z_p\text{わからん}$$

- ここで，容易にサンプリングできる提案分布q(z)を考える
- これがミソ

棄却サンプリング

乱数発生

z_0

提案分布

p(z)を提案分布で覆う

分布p(z)をkq(z)が覆うような定数kを考える

乱数発生

u_0

区間[0,kq(z0)]で一様分布からu0発生

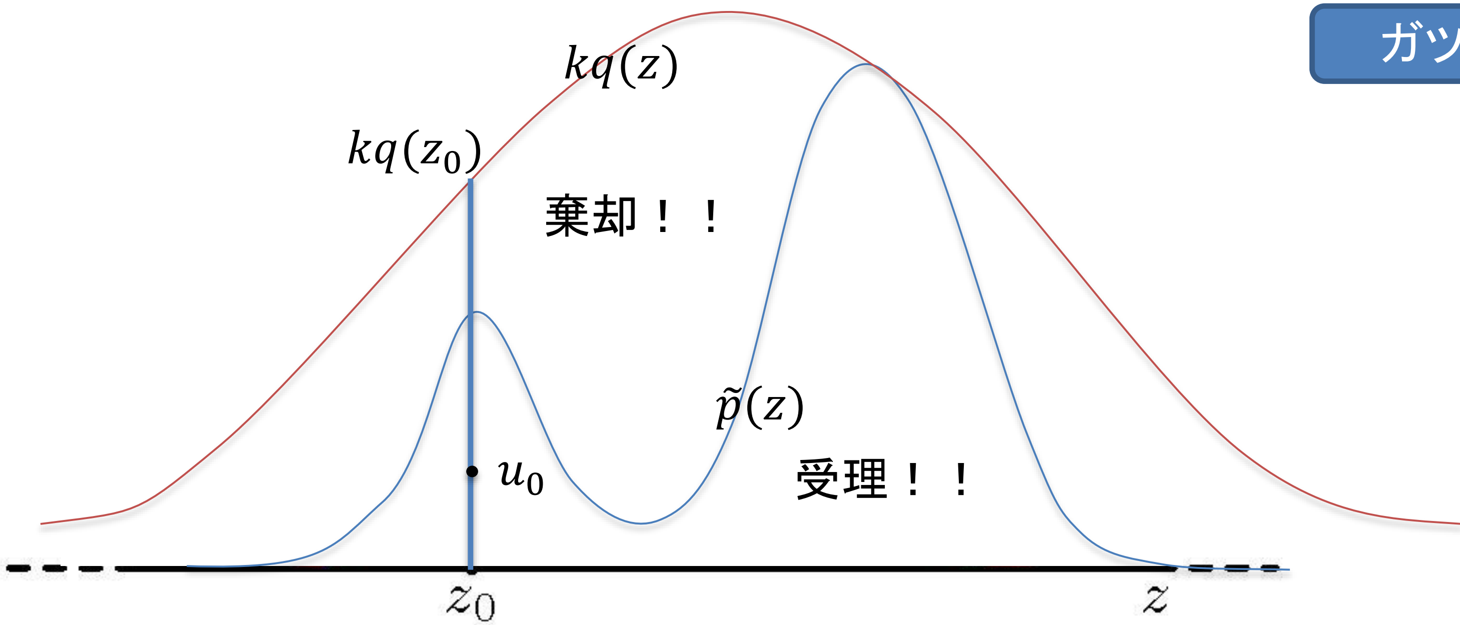
$u_0 > \tilde{p}(z_0)$

ならz0は 棄却

$u_0 \le \tilde{p}(z_0)$

ならz0はp(z)に従うサンプリングとして

ガツンと得られる



適応的棄却サンプリング

- 棄却サンプリングを適用したい多くの場合，適切な提案分布q(z)を解析的に決定することは難しい
- 別のアプローチとして，分布p(z)の観測値に基づいてその場で包絡関数を構築する方法

初期集合を準備

初期集合による包絡分布q(z)でp(z)を覆う

乱数発生

z_4

包絡分布

乱数発生

u_0

区間[0,q(z_4)]で一様分布からu_0発生

$u_0 > \tilde{p}(z_4)$

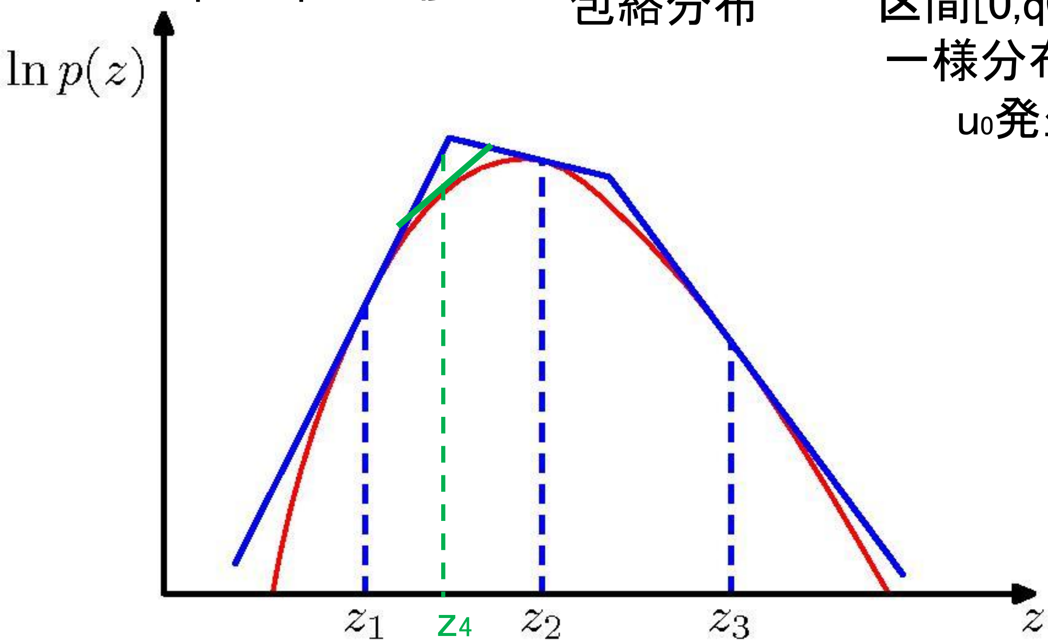
ならz_4は 棄却

だがz_4を包絡分布の接点に生かす

$u_0 \leq \tilde{p}(z_4)$

ならz_4はp(z)に従うサンプリングとして

ガツンと得られる



ここまでのまとめ

- 逆関数法
  - 解析的に正しいし，無駄がないというメリット
  - 複雑な分布だと逆関数を求めるのはまず不可能というデメリット
- 棄却サンプリング
  - 比較的複雑な分布に対してもサンプリングできるメリット
  - 適切な提案分布，定数kを選ぶことが難しい
  - 結果としてせっかくサンプリングしても大量に棄却するので無駄
- 適応的棄却サンプリング
  - 包絡関数を提案分布とすることで棄却が少なくなるメリット
  - 棄却した値も新たな包絡線の接点として取り込む再活用
  - 接線の計算負荷．また多次元で多峰性，するどいピークをもつ分布だとまず対応できないデメリット

棄却サンプリングは次元の多い分布だと棄却率が指数的に増加し非現実的

重点サンプリング

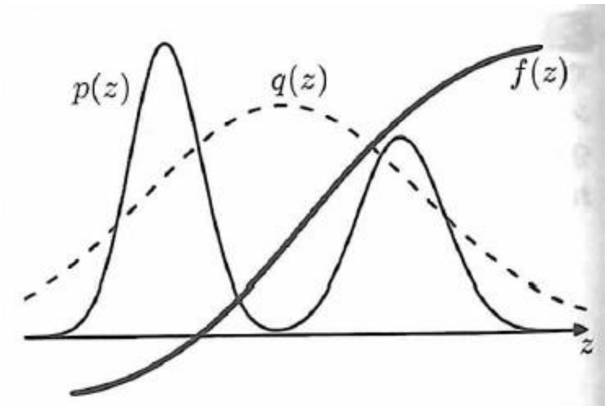
- 分布p(z)の期待値を計算したいことが目的のときもある(重点サンプリングはむしろそれしかできません)
- 一様にサンプリングして期待値もとめるのは非効率的なので提案分布を使います！

平均を取っているようなもの

$$E[f] = \int f(z)p(z)dz = \boxed{\int} f(z)\frac{p(z)}{q(z)}\boxed{q(z)}dz$$

$$\approx \frac{1}{L}\sum_{l=1}^{L} \boxed{\frac{p(z^{(l)})}{q(z^{(l)})}} f(z^{(l)})$$

重要度重みと呼び，求めたいものとは異なった分布からサンプリングするバイアスを補正する．棄却サンプリングと異なり，すべてのサンプルを保持することに注意！



提案分布はp(z)･f(z)が大きくなるところで
分布が大きくなるようなものにすると効率的！！

- 棄却サンプリングのうまいq(z)とkの選び方を発見することは現実的ではない
- そこで，重点サンプリングの重要度重みを有効活用して　なんかうまいサンプリング方法はないか？→SIR
- $\tilde{p}$と$\tilde{q}$は容易に計算できるという過程
- 第１段階で提案分布 $q(z)$からL個のサンプルzを抽出する
- 第２段階で次式によって重みwを計算する

$$w_l = \frac{\boxed{\dfrac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}}{\sum_l \dfrac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}$$

重要度重み

- 最後にzから重いwで与えられる確率に従ってリサンプリングする(このサンプルは$p(z)$に従う！！)
- これはL→∞では分布は正確に従うことが証明されている

SIR(sampling-importance-resampling)

乱数発生

$$\left(z^{(1)},\cdots,z^{(L)}\right)$$

提案分布からL個サンプリング

p(z)から重みwを計算

$$w_l = \frac{\dfrac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}{\sum_l \dfrac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}$$

各lに対応した重みwを計算

リサンプリング

$$\left(z^{(1)},\cdots,z^{(L)}\right)$$

$$\left(w_1,\cdots,w_L\right)$$

zを重みwに対応する確率で再抽出

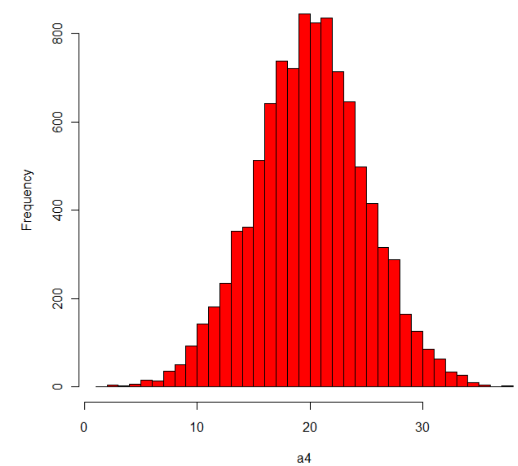
ガツンと得られる

$$\left(z^{(1)}*,\cdots,z^{(l)}*\right)$$

L個のサンプルが得られる

ex) 提案分布 : [0,50]の一様分布でp(z)がN(20,5)とする

z =	p(z) =	w =	z *=
17.08	6.730e-02	1.654e-01	17.08
2.745	2.069e-04	5.087e-04	21.15
38.93	6.136e-05	1.508e-04	17.08
38.40	9.096e-05	2.235e-04	21.34
21.34	7.694e-02	1.891e-01	21.34
41.04	1.133e-05	2.785e-05	21.15
20.52	7.935e-02	1.950e-01	20.52
22.17	7.259e-02	1.784e-01	22.17
21.15	7.770e-02	1.910e-01	21.15
13.30	3.254e-02	7.999e-02	17.08



ここまでのまとめ（2）

- 重点サンプリング
  - 提案分布との重み付けで期待値を近似的に計算する
  - 分布全体からサンプリングするものではない
- SIR
  - 棄却サンプリングと重点サンプリングの合わせ技
  - あくまで最初のサンプリングからのリサンプリングなので，反復回数が少ないと偏る…が，∞では近似できることは証明されてる

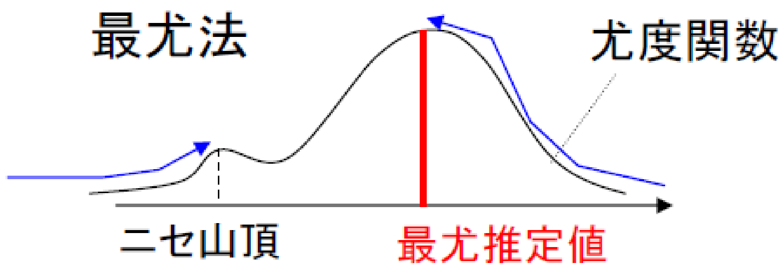
サンプリングしたい確率分布の次元が高いとサンプリングしなければならない数が膨大に…

そんなあなたには！！！！！！！

MCMCならぴったり！！

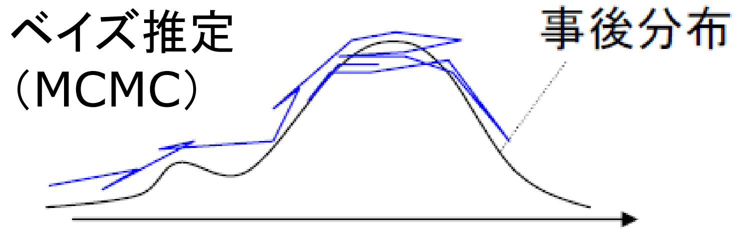
頻度主義とベイズの世界観の違い

- たとえばパラメータ推定の話



ひたすら山頂を目指す
　⇒初期値によっては間違った結果

パラメータはひとつの値をとる



事後分布の形そのものを推定
　⇒大域的最適解にたどり着きやすい

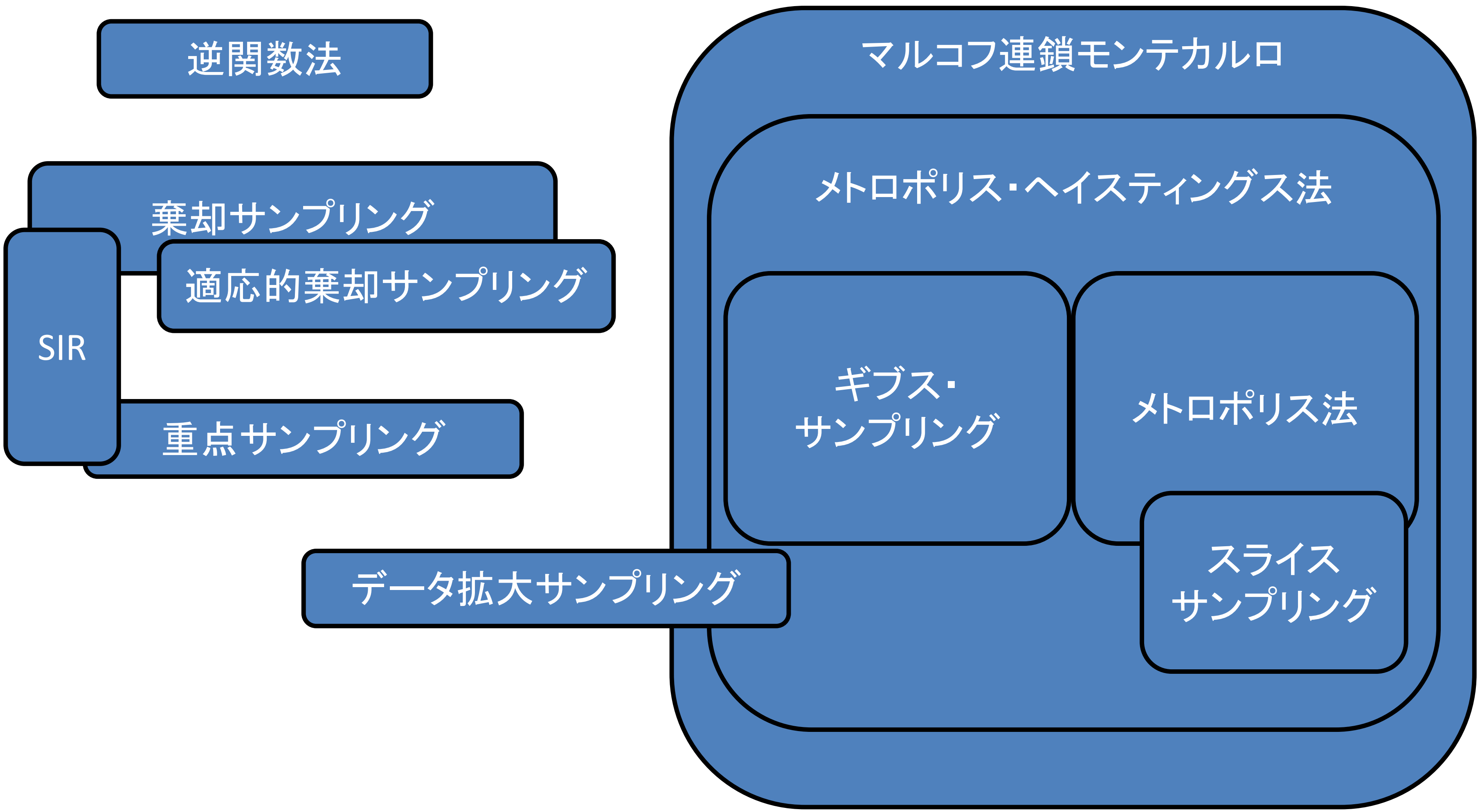
パラメータは確率分布をとる

- 最尤法では山頂にのみ関心があるが
- MCMCは山全体の形に興味がある

なぜMCMCはモテ系なのか？

- 棄却サンプリングやSIR法では効率良いサンプリングのためにサンプラー（提案分布）を上手に選んでやる必要がある
- 非常に複雑な分布（非線形モデル）や大量のパラメータがある（次元が多い）とき，適切にサンプラーを選ぶのは難しい．ときに不可能
- MCMCなら定常分布に収束するという性質を持っており，初期値に依存せず，いい感じのサンプリングが楽に行える
- モテ系というよりマッチョ系

全体像を俯瞰


逆関数法
棄却サンプリング
適応的棄却サンプリング
SIR
重点サンプリング
データ拡大サンプリング
マルコフ連鎖モンテカルロ
メトロポリス・ヘイスティングス法
ギブス・
サンプリング
メトロポリス法
スライス
サンプリング

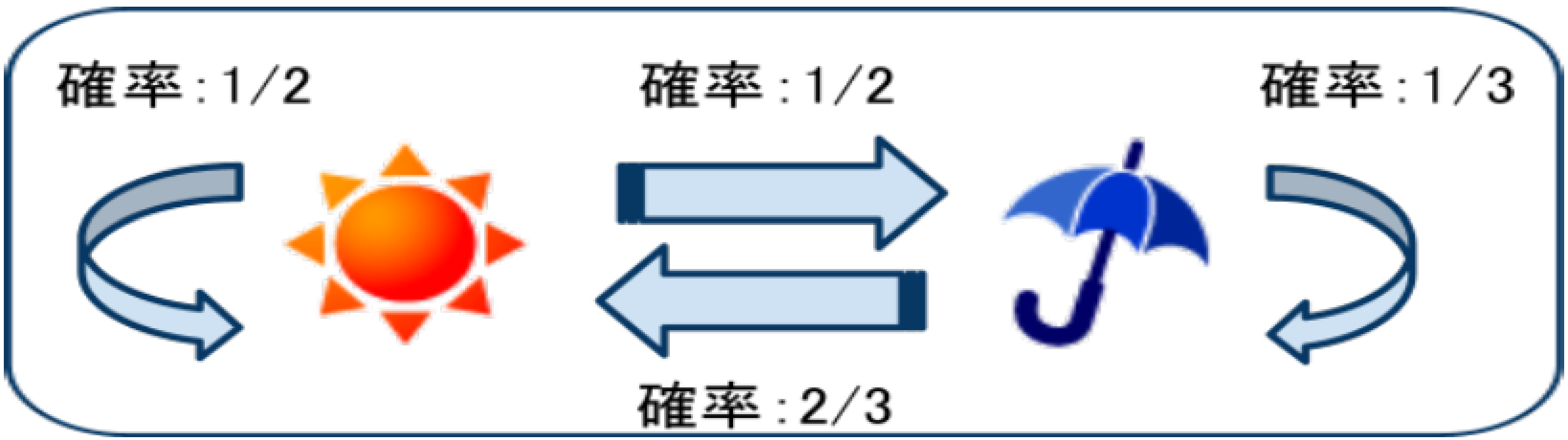
マルコフ連鎖モンテカルロ法とは

- 任意の分布p(z)を不変分布にもつマルコフ連鎖を生成することで，望む分布に従った乱数を生成するモンテカルロ法のこと！！

- 不変分布とは？

不変分布とは

- 次の例を考えてみる



$$\begin{pmatrix} F_{t+1} \\ R_{t+1} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} F_t \\ R_t \end{pmatrix} \qquad \begin{pmatrix} F_0 \\ R_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

この漸化式を解いてやると…　　　　つまり…

$$\begin{pmatrix} F_t \\ R_t \end{pmatrix} = \begin{pmatrix} \frac{3}{7}\left(-\frac{1}{6}\right)^t + \frac{4}{7} \\ -\frac{3}{7}\left(-\frac{1}{6}\right)^t + \frac{3}{7} \end{pmatrix} \qquad \begin{pmatrix} F_\infty \\ R_\infty \end{pmatrix} = \boxed{\begin{pmatrix} \frac{4}{7} \\ \frac{3}{7} \end{pmatrix}}$$

これが
不変分布

不変分布とは

- マルコフ連鎖では最終的にある一定の確率分布へと収束し，これが不変分布である
- 初期値には依存しない
- この性質はかなり大事

マルコフ連鎖モンテカルロ法とは

- 任意の分布p(z)を不変分布にもつマルコフ連鎖を生成することで，望む分布に従った乱数を生成するモンテカルロ法のこと！！

- マルコフ連鎖の生成による乱数生成は遷移確率が与えられているならばただのすごろく！

さっきの例を使ってサンプリング


```
#晴⇒晴：1/2,雨⇒雨：1/3な推移確率
transition <- c(Fine=1/2,Rain=1/3)
size <- 10000 #サンプリング回数
result <- c() #結果格納
weather.now <- "晴” #現在のお天気
#マルコフ連鎖の生成＆不変分布からのサンプリング.
for(i in 1:size){
  result <- c(result,weather.now)
  #現在の状態に応じて次の状態を決める⇒マルコフ連鎖の生成
  if(weather.now=="晴”){
    weather.now <- ifelse(runif(1) > transition["Fine"],"晴”,”雨”)
  }else if(weather.now=="雨”){
    weather.now <- ifelse(runif(1) > transition["Rain"],"晴”,”雨”)
  }
}
#結果表示、標本から確率を計算
#初めの10個は不変分布に達してないと思って捨てる
print(sum(result[-(10:1)]=="晴”)/(size-10))
print(sum(result[-(10:1)]=="雨”)/(size-10)
```

```
> print(sum(result[-(10:1)]=="晴”)/(size-10))
[1] 0.5705706
> print(sum(result[-(10:1)]=="雨”)/(size-10))
[1] 0.4294294
```

マルコフ連鎖モンテカルロ法の発想 26

- 推移確率が与えられていたら簡単
- 例とは違って不変分布が与えられていて推移確率が与えられていない場合は？？
- 不変分布に到達するような推移確率を与え，マルコフ連鎖を生成することで，不変分布を再現(近似)することが可能
- 推移確率を与える手段は？？

  → メトロポリス法，ギブスサンプリング！

推移確率決定の指針

- 詳細釣り合いの条件を満たすように推移確率を決めれば，望む不変分布を得る！
- 詳細釣り合いの条件
  - お天気の例で行くと

$$\underbrace{\frac{4}{7}}_{\text{不変分布晴の確率}} \times \underbrace{\Pi_{Fine \to Rain}}_{\text{晴⇒雨の推移確率}} = \underbrace{\frac{3}{7}}_{\text{不変分布雨の確率}} \times \underbrace{\Pi_{Rain \to Fine}}_{\text{雨⇒晴の推移確率}}$$

推移確率決定の指針

- 先ほどの式を満たす推移確率ならば不変分布を得ることが本当にできるのか
- 推移確率＊不変分布＝不変分布になるはず

$$\boxed{\begin{pmatrix} 1-\Pi_{Fine\to Rain} & \Pi_{Rain\to Fine} \\ \Pi_{Fine\to Rain} & 1-\Pi_{Rain\to Fine} \end{pmatrix}} \boxed{\begin{pmatrix} \frac{4}{7} \\ \frac{3}{7} \end{pmatrix}} =$$

$$\begin{pmatrix} \frac{4}{7}-\frac{4}{7}\Pi_{Fine\to Rain}+\frac{3}{7}\Pi_{Rain\to Fine} \\ \frac{4}{7}\Pi_{Fine\to Rain}+\frac{3}{7}-\frac{3}{7}\Pi_{Rain\to Fine} \end{pmatrix} = \begin{pmatrix} \frac{4}{7} \\ \frac{3}{7} \end{pmatrix}$$

↑詳細釣り合いの条件で消える

より一般的な証明

詳細釣り合いの条件は

$$Pr\{x\} \times \Pi_{x \to x'} = Pr\{x'\} \times \Pi_{x' \to x}$$

証明：不変確率＝推移確率＊不変分布

$$
\begin{aligned}
Pr\{x\} &= \sum_{x'} \Pi_{x' \to x} \times Pr\{x'\} \\
&= \sum_{x'} \Pi_{x \to x'} \times Pr\{x\} \\
&= Pr\{x\} \sum_{x'} \Pi_{x \to x'} = Pr\{x\}
\end{aligned}
$$

↑お天気推移モデルの一般化式

↑詳細釣り合いの条件↓

↓推移確率の和は1

Q.E.D

メトロポリス法

- 推移確率を以下の式で与える

$$\Pi_{x \rightarrow x'} = \min(1, \frac{Pr\{x'\}}{Pr\{x\}})$$

- このとき詳細釣り合いの条件式が満たされる
- 不変分布の比をとるので不変分布は規格化の必要がない！！
- 比は状態xと状態x’間で取るので，推移先x’を決めてあげなければならない
- 比が1より大きければ必ず推移し，比が1より小さければその確率で推移する
- つまり推移しないときも当然ある

なんとなく分布になりそうなイメージ！

メトロポリス法の簡単アルゴリズム

- step0:初期位置を設定
- step1:推移候補先を選定
- step2:メトロポリス法で決まる推移確率で推移
- step3:気が済んだ→end　済まない→step1

※推移候補先の選定に使う確率分布は詳細釣り合いの条件を満たすように！！(一様分布が一般的)

みてみましょう

二次元正規分布のサンプリング

http://visualize-mcmc.appspot.com/2_metropolis.html

ギブス・サンプリング

- メトロポリス法は推移位置を自分で設定して与える必要がある
- ギブスサンプリングはそんなことないです
- 推移確率は以下の通り

$$
\begin{aligned}
x &= (x_0, x_1) \\
x' &= (x_0, x_1') \\
\Pi_{x \to x'} &= \frac{Pr\{x'\}}{\sum_{x_1'} Pr\{x'\}} = \frac{Pr\{x'\}}{Pr\{x_0\}} \\
&= Pr\{x_1' | x_0\}
\end{aligned}
$$

x_0,x_1がそれぞれの次元での状態(位置)を表す

分母は規格化定数！

Pr{x}は確率(密度)

周辺化してる

ギブスサンプリング

- 詳細釣り合いの条件を満たすことの証明

$$
\begin{aligned}
Pr\{x\}\Pi_{x\to x'} &= Pr\{x\}\frac{Pr\{x'\}}{Pr\{x_0\}} \\
= Pr\{x'\}\frac{Pr\{x\}}{Pr\{x_0\}} &= Pr\{x'\}\Pi_{x'\to x}
\end{aligned}
$$

- 条件付分布が求まらなければ使えない

ギブス・サンプリング

- たとえば二変量正規分布の場合

（本当はもっと高次元がMCMCの腕の見せ所であるが…）

$$p(x_1, x_2) = \frac{1}{Z}\exp\left(-\frac{x_1^2 - 2bx_1x_2 + x_2^2}{2}\right)$$

ここで，一方を固定したときの条件付き密度は

$$p(x_1 \mid x_2) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_1 - bx_2)^2}{2}\right) \qquad p(x_2 \mid x_1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_2 - bx_1)^2}{2}\right)$$

単なる正規分布からのサンプリングになったぞー！→余裕

- 条件付き分布からのサンプリングが簡単な場合，**魅力的**
- 条件付き分布が適切に書き下せる場合がギブスサンプラーの使いどころ
- 条件付き分布からのサンプリングが簡単でない場合→MH

ギブス・サンプリングアルゴリズム

1. $\{z_i : i = 1, \ldots, M\}$ の初期値 $\{z_1^{(1)}, z_2^{(1)}, \ldots, z_M^{(1)}\}$ を与える
2. $\tau = 1, \ldots, T$ に対して以下を行う
   - $z_1^{(\tau+1)} \sim p(z_1 \mid z_2^{(\tau)}, \ldots\ldots, z_M^{(\tau)})$ をサンプリングする
   - $z_2^{(\tau+1)} \sim p(z_2 \mid z_1^{(\tau+1)}, \ldots\ldots, z_M^{(\tau)})$ をサンプリングする
   - $z_j^{(\tau+1)} \sim p(z_j \mid z_1^{(\tau+1)}, \ldots\ldots, z_M^{(\tau)})$ をサンプリングする

     …
   - $z_M^{(\tau+1)} \sim p(z_M \mid z_1^{(\tau+1)}, \ldots\ldots, z_{M-1}^{(\tau+1)})$ をサンプリングする
3. 収束したと判定されるまでステップ2を繰り返す

一つ前の期の自分以外に依存しているのでマルコフ連鎖の一種である

ギブス・サンプリング

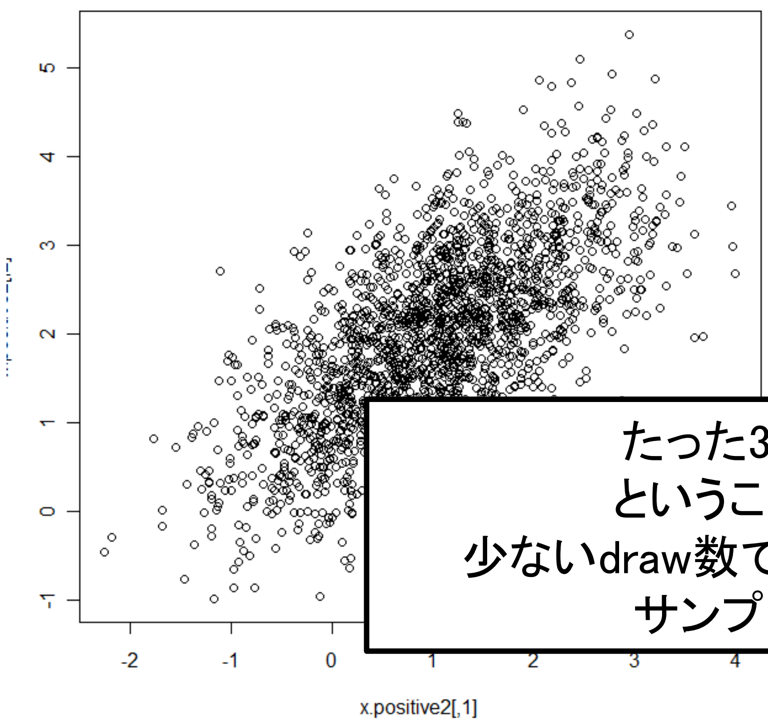
- たとえば二変量正規分布の場合

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right)$$

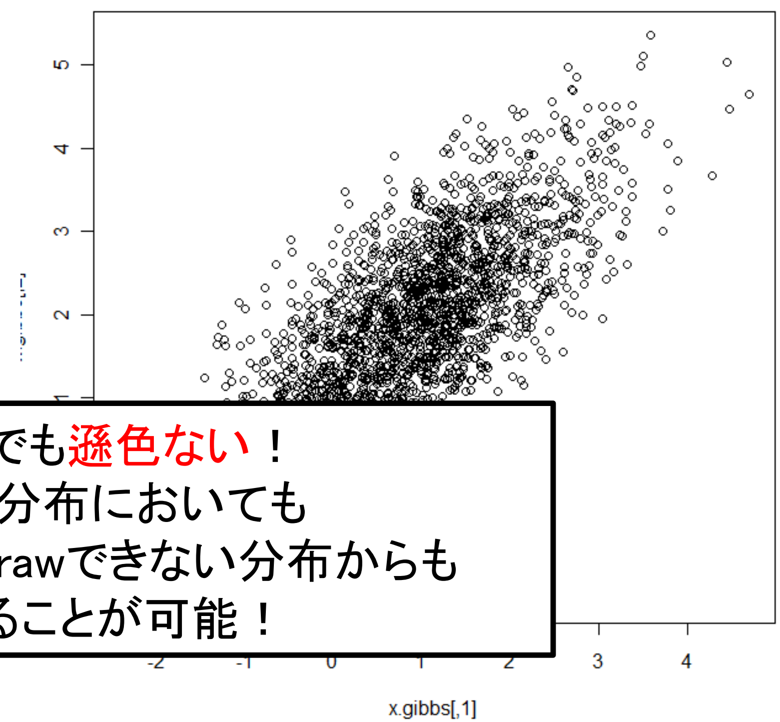
$$x_1 \sim N(1+0.7(x_2-2), 1-0.7^2)$$

$$x_2 \sim N(2+0.7(x_1-1), 1-0.7^2)$$

Rの多変量正規分布の乱数発生関数から2000個



ギブスサンプラー(3000draw 1000はburn-in)



たった3000drawでも遜色ない！
ということは他の分布においても
少ないdraw数で簡単にdrawできない分布からも
サンプリングすることが可能！

やっぱりみてみよう

二次元正規分布のサンプリング

http://visualize-mcmc.appspot.com/3_gibbs.html

メトロポリス・ヘイスティングス法

- 今までのなかで一番一般的な形
- メトロポリス法，ギブスサンプリング∈MH法
- MH法では推移確率を以下のように与える

$$\Pi_{x \to x'} = Q(x, x') \min(1, \frac{Pr\{x'\}Q(x', x)}{Pr\{x\}Q(x, x')})$$

と決定する。ここでQ(x,x')はxという状態にいるときにx'という状態が選択される確率です。

Q(x,x')を適切に選ぶことでメトロポリス法・熱浴法が再現される

MH法

■Q(x,x')=Q(x',x)とすると、約分されて・・・

$$\Pi_{x \to x'} = Q(x, x') \min(1, \frac{Pr\{x'\}}{Pr\{x\}})$$

となってメトロポリス法が再現される。

■Q(x,x')=Pr{x' | x}とすると

$$\begin{aligned} \Pi_{x \to x'} &= Q(x, x') \min(1, \frac{Pr\{x'\}Q(x', x)}{Pr\{x\}Q(x, x')}) = Pr\{x'|x\} \min(1, \frac{Pr\{x'\}Pr\{x|x'\}}{Pr\{x\}Pr\{x'|x\}}) \\ &= Pr\{x'|x\} \min(1, \frac{Pr\{x, x'\}}{Pr\{x', x\}}) = Pr\{x'|x\} \min(1, 1) \\ &= Pr\{x'|x\} \end{aligned}$$

となってギブサンプリングが再現される。

MCMCのまとめ(3)

- メトロポリス法
  - 条件付確率が書き下せないときはこれ
  - ステップ幅の設定が難しい(大きすぎたら棄却されやすく，小さすぎたらランダムウォークしてしまう)
- ギブスサンプリング
  - 条件付き確率が書き下せればかなり効率良い
  - 1つ1つ固定していったうえでのサンプリング
- メトロポリスヘイスティングス法
  - 一般化
- スライスサンプリング
  - ステップ幅を適切に設定することができる
  - 今回は割愛