

# 11章 サンプルリング法

小野研 中島優作

以下のスライドを参考、一部引用しています

[http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11\\_chika.pdf](http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11_chika.pdf)

# 前半目次

## Chapter11 サンプリング法 序章

- サンプリング法の基本アイデア

### 11.1 基本的なサンプリングアルゴリズム(モンテカルロ法)

- 11.1.1 標準的な分布
- 11.1.2 棄却サンプリング
- 11.1.3 適応的棄却サンプリング
- 11.1.4 重点サンプリング
- 11.1.5 SIR
- 11.1.6 サンプリングとEMアルゴリズム

# 後半目次

## 11.2 マルコフ連鎖モンテカルロ(MCMC)

- 11.2.1 マルコフ連鎖とメトロポリス法
- 11.2.2 メトロポリス・ヘイスティングス法

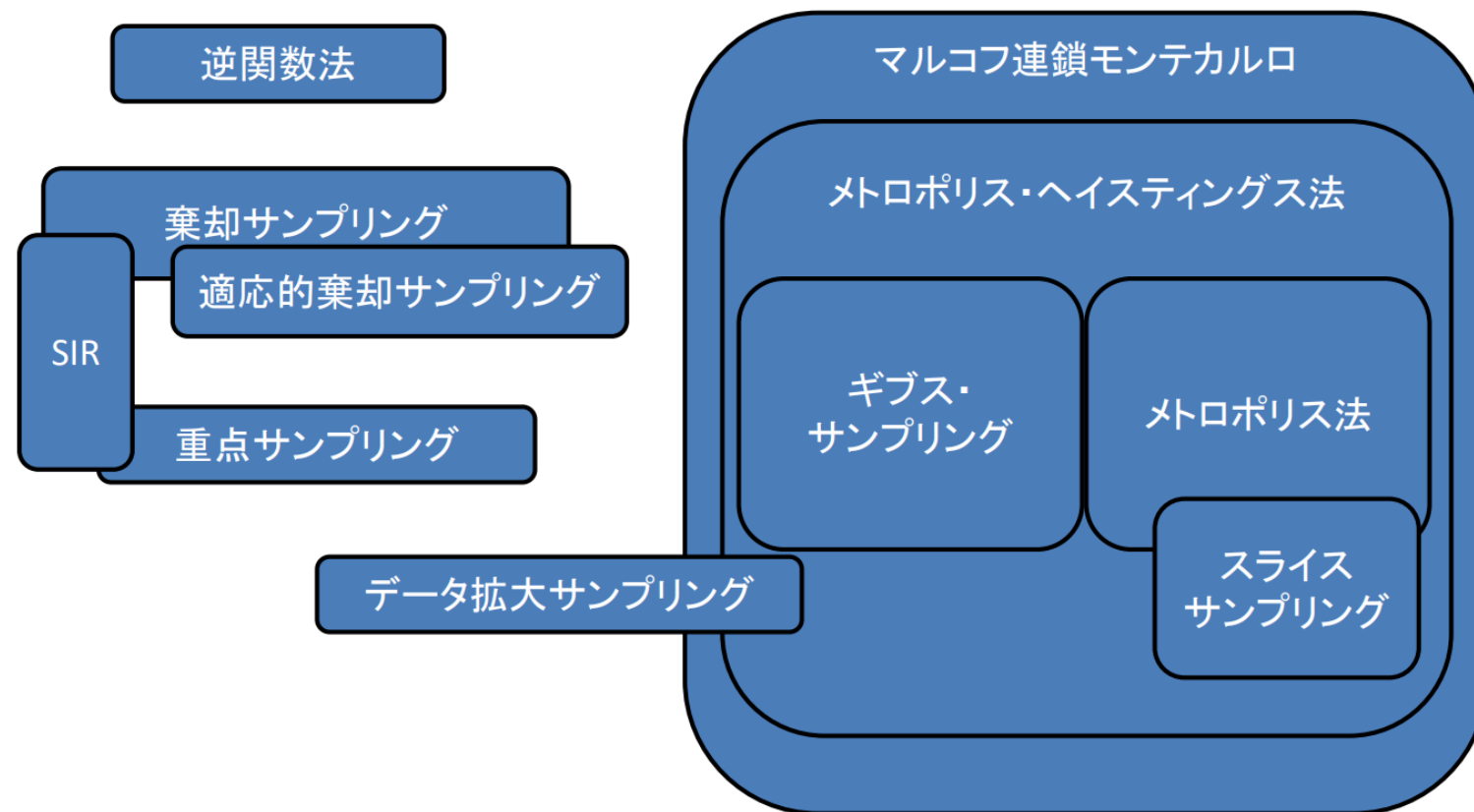
## 11.3 ギブスサンプリング

## 11.4 スライスサンプリング

## 11.5 ハイブリッドモンテカルロ(統計力学との融合)

- 11.5.1 力学系
- 11.5.2 ハイブリッドモンテカルロ

# サンプリング法の俯瞰マップ



以下より引用

[http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11\\_chika.pdf](http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11_chika.pdf)

# Chapter 11 サンプリング法 序章

- **第10章:** 決定論的近似に基づく推論アルゴリズム
  - 変分ベイズ法
  - これらの方法は解析的手法を用いて近似推論を行うが、実用的には正確な推論は困難な場合も多い
- **第11章:** 数値サンプリングに基づく近似推論法
  - 前半：モンテカルロ法(ベイズではなく統計的な手法)
  - 後半：マルコフ連鎖モンテカルロ法

**流れ：前半は単純な分布を解析的、統計的な手法でサンプリング**  
→ **色々工夫してみるけど高次元な分布からのサンプリングは無理**  
→ **後半にマルコフ連鎖を使って高次元でも効率的にサンプリング**

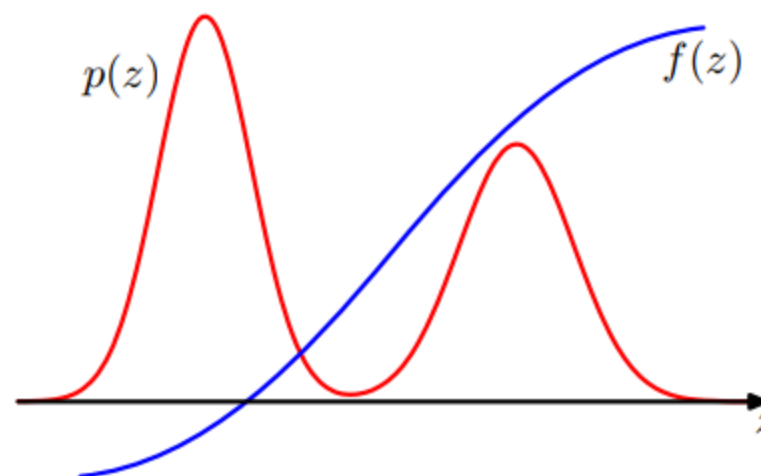
## 近似推論の目的

- 近似推論を用いることで事後分布を知ることができるが、多くの場合は事後分布そのものを使うのではなく**事後分布を元に計算した期待値**を使うことが多い
- 関数  $f(z)$  について、  $p(z)$  のもとでの期待値  $E[f]$  を求める
- 連続変数の場合:  $E[f] = \int f(z)p(z) dz$  (11.1)
- 離散変数の場合:  $E[f] = \sum f(z)p(z)$

# サンプリング法の基本アイデア

- 分布  $p(z)$  から独立にサンプル  $z^{(l)}$  を取得 ( $l = 1, \dots, L$ )
- 式11.1の期待値  $E[f]$  を有限の総和で近似
- 推定量  $f \approx \frac{1}{L} \sum_{l=1}^L f(z^{(l)})$
- 図11.1は関数と分布の関係を示したただの模式図である。後ほど出てくるがこの図は全く良い近似ではない点に注意

**Figure 11.1** Schematic illustration of a function  $f(z)$  whose expectation is to be evaluated with respect to a distribution  $p(z)$ .



## 理想的な推定量の性質

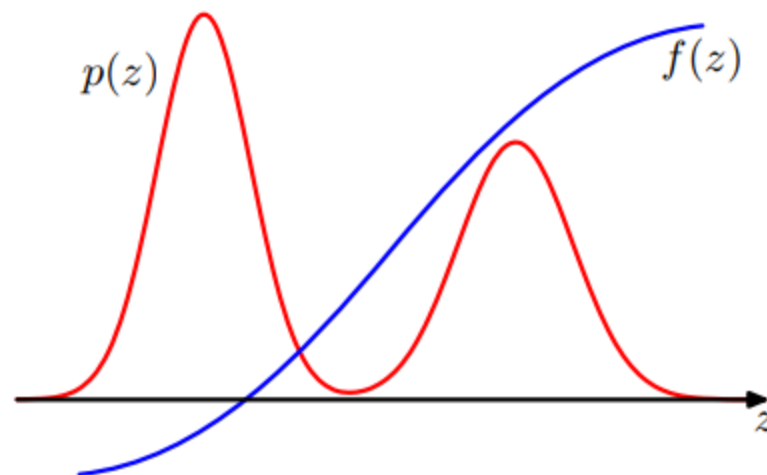
- サンプル  $z^{(l)}$  が分布  $p(z)$  に従って抽出されている限りは、 $E[\hat{f}] = E[f]$  が成り立ち、 $\hat{f}$ は正しい平均を持つ
- これに基づいて分布  $p(z)$  のもとでの  $z^{(l)}$  の分散は  $\text{var}[\hat{f}] = \frac{1}{L} E[(f - E[f])^2]$  となる
- よって  $z^{(l)}$  が分布  $p(z)$  に従って抽出されている限り、推定量の精度(分散)は  $z$  の次元に依存しないため、少数のサンプルでも高精度な期待値の推定が可能となる
  - 実際、十分な精度で期待値を計算するには10~20のサンプリングで十分  
(と書いてあるが根拠は?)



# 実用におけるサンプリングの課題

- 実際はサンプル $z^{(l)}$ が独立でない場合が多く、サンプルが重複するため実効的なサンプルサイズは見かけよりも小さくなる
- また、図11.1の例だと $z^{(l)}$ が分布 $p(z)$ に従って抽出されていない
  - $p(z)$ が小さい(確率の低い)領域で $f(z)$ の期待値が大きいため、小さな確率の領域から期待値に大きく影響している
  - この場合、サンプリング数を多くして精度を上げる必要がある

**Figure 11.1** Schematic illustration of a function  $f(z)$  whose expectation is to be evaluated with respect to a distribution  $p(z)$ .



## 同時分布のサンプリング

- $p(z)$ は同時分布であることが多く、グラフィカルモデルで指定するのが便利
- 観測変数のない有向グラフの場合、伝承サンプリング(8.1.2で紹介)で簡単に行える
- 同時分布  $p(z) = \prod_{i=1}^M p(z_i | pa_i)$ 
  - $z_i$ は、グラフィカルモデルのノードに属する変数集合
  - $pa_i$ は、ノードiの親となる変数集合
- 各ノードの条件付き分布から順番にサンプルを抽出していくが、各ステップで親となるすべての変数の値が決まっているため、サンプリングは常に可能である
  - 最終的に同時分布から1つのサンプルが得られる

## 観測値を持つ有向グラフの場合

- いくつかのノードに対して観測値が与えられる場合、離散変数の場合にはさきほどのアプローチを拡張したロジックサンプリングアプローチが可能
  - これは重点サンプリング(後ほど述べる)の特別な場合
- ロジックサンプリングのアルゴリズム
  - 各ステップで観測値を持つ変数 $z_i$ のサンプル値と観測値が一致する場合にサンプルを保持
  - 一致しない場合は、サンプル全体を破棄して最初のノードから再開
  - つまり、隠れ変数とデータ変数の同時分布からサンプリングし観測データと一致しないサンプルを破棄することと同じであるため、事後分布から正しくサンプリングを行う
    - 矛盾すると即打ち切りのため、厳密にはわずかに時間を節約している
- 観測変数の数が増えると受け入れ確率が急速に減少するため、通常用いない

## 無向グラフとサンプリング

- 観測変数を持たない事前分布でさえ、ワンパスでサンプリングする手段はない
- ギブスサンプリング(11.3節)などの計算コストの高い手法が必要

## 周辺分布からのサンプリング

- 同時分布だけでなく、周辺分布からサンプリングが必要な場合
- 同時分布  $p(u, v)$  からサンプルを取得できる場合は周辺分布の計算は簡単で、各サンプルで  $v$  の値を無視することで周辺分布  $p(u)$  からのサンプリングとなる

## 11.1 基本的なサンプリングアルゴリズム

- 基本的なサンプリング = 解析的もしくは統計的なサンプリング
- このセクションでは、与えられた分布からランダムサンプルを生成するためのいくつかの簡単な戦略を扱う
- コンピュータによるランダムサンプリングは実際には疑似乱数となるが、ここでは一様な乱数生成機として扱う

## 11.1.1 標準的な分布(逆関数法)

- ここでは、一様分布の乱数生成器が利用可能であると仮定して、**単純な非一様分布(既知)から乱数を生成する方法**を考える
  - この手法は、**逆関数法**として知られている
- $z$  が区間  $(0, 1)$  に均一に分布していると仮定し、 $y = f(z)$  という関数を使用して  $z$  の値を変換する
- ここで目的のために、一様分布  $p(z)$  を変換して生成した非一様分布  $p(y)$  から乱数を生成する
- $p(z)$  と  $p(y)$  の変換をヤコビアンを用いて次の式で計算できる

$$p(y) = p(z) \left| \frac{dz}{dy} \right| \quad (11.5)$$

- なお、 $p(z) = 1$

- ここでの目標は、変換後の  $y$  が特定の分布  $p(y)$  に従うように関数  $f(z)$  を選択することである
- 式 (11.5) を積分すると次のようになる

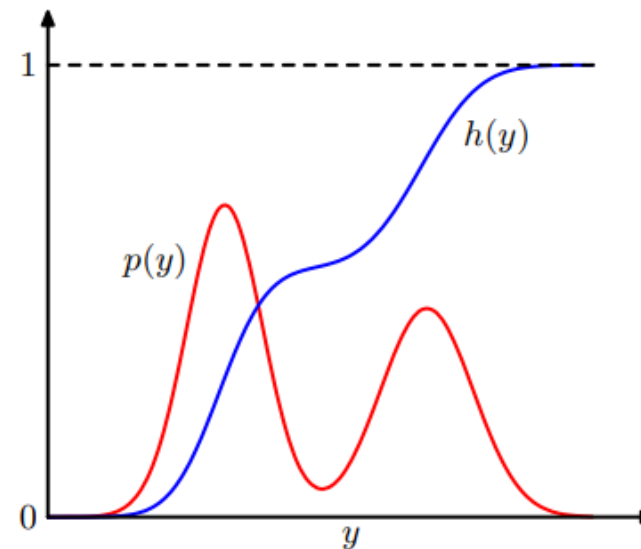
$$z = h(y) \equiv \int_{-\infty}^y p(\hat{y}) d\hat{y} \quad (11.6)$$

- この式は  $p(y)$  の不定積分である



- 式11.6  $z = h(y)$  は  $p(y)$  の不定積分であるため、 $y = h^{-1}(z)$  は  $p(z)$  の不定積分の逆関数である
- $p(z)$  は一様分布を仮定していたため、非一様分布の不定積分の逆関数を使用して、一様分布からのサンプリング(すなわち乱数生成)を行っている
- これを図示したのが図11.2
  - $h(y)$  が  $p(y)$  の面積(不定積分)であり、逆関数で分布を打ち消すことで一様分布となるイメージ

**Figure 11.2** Geometrical interpretation of the transformation method for generating nonuniformly distributed random numbers.  $h(y)$  is the indefinite integral of the desired distribution  $p(y)$ . If a uniformly distributed random variable  $z$  is transformed using  $y = h^{-1}(z)$ , then  $y$  will be distributed according to  $p(y)$ .



## 複数変数への一般化

式11.5では1変数を対象としていたが、複数変数に拡張したヤコビアンを考えることで複数変数にも適用できる

$$p(y_1, \dots, y_M) = p(z_1, \dots, z_M) \left| \frac{\partial(z_1, \dots, z_M)}{\partial(y_1, \dots, y_M)} \right| \quad (11.9)$$

## 例: 指数分布

具体例として指数分布を考える

$$p(y) = \lambda \exp(-\lambda y) \quad (11.7)$$

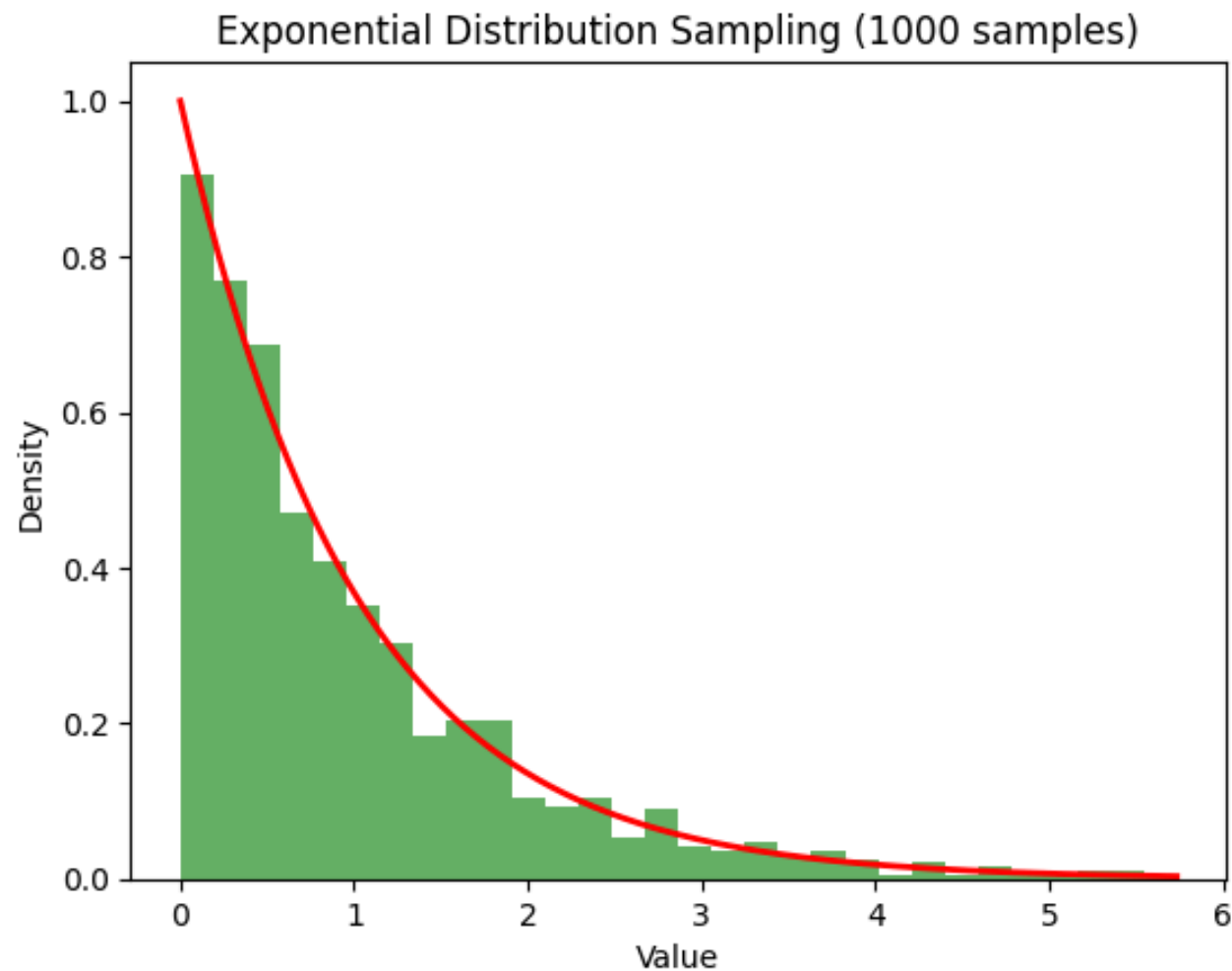
ここで、 $0 \leq y < \infty$  である

この場合、式11.6における積分の下界は0であり、次のようになる

$$h(y) = 1 - \exp(-\lambda y)$$

したがって、変数  $z$  を次のように変換すると、 $y$  は指数分布に従う

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



## 例: コーシー分布

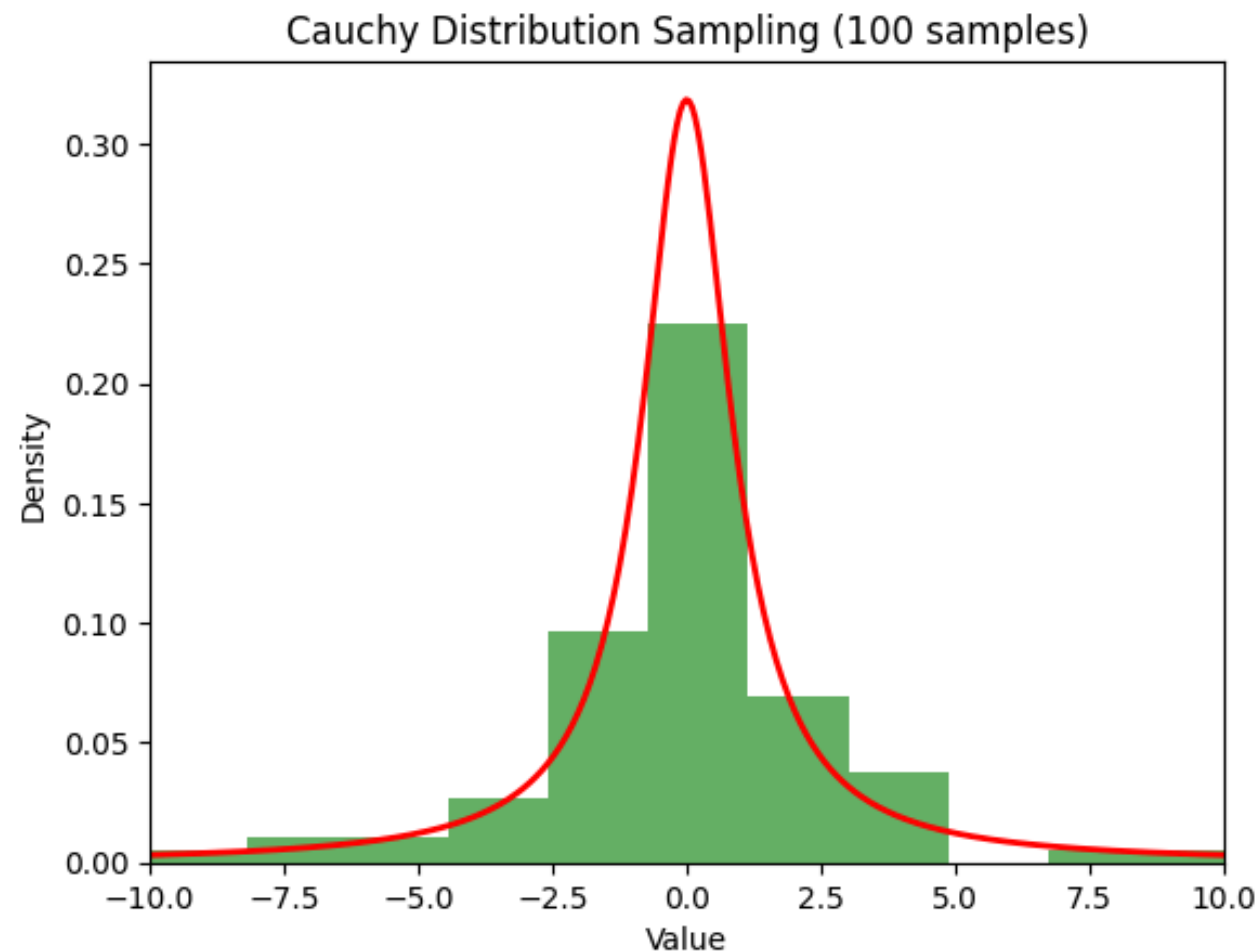
もう一つの例として、コーシー分布を考える

$$p(y) = \frac{1}{\pi} \frac{1}{1 + y^2} \quad (11.8)$$

この場合、不定積分の逆関数は  $\tan$  関数で表現できる

$$y = \tan(\pi(z - 0.5))$$

※コーシー分布は裾が広いので図は見やすいよう可視化範囲を絞った



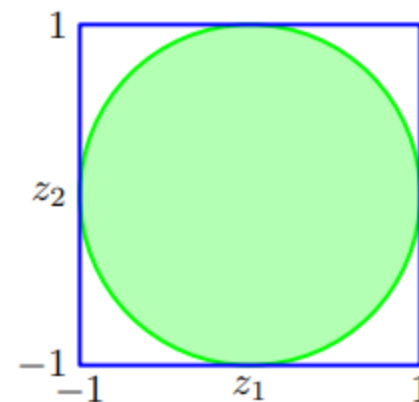
## 例: ガウス分布(ボックス=ミュラー法)

ガウス分布からサンプルを生成するためのボックス=ミュラー法を考える  
まず、区間  $(-1, 1)$  に均一に分布した乱数のペア  $z_1, z_2$  を生成する  
次に、次の条件を満たすペアを破棄する

$$z_1^2 + z_2^2 \leq 1$$

これにより、単位円内の点の均一分布が得られる

**Figure 11.3** The Box-Muller method for generating Gaussian distributed random numbers starts by generating samples from a uniform distribution inside the unit circle.



次に、各ペア  $z_1, z_2$  に対して次の量を評価する

$$y_1 = z_1 \sqrt{-2 \ln r^2 / r^2} \quad (11.10)$$

$$y_2 = z_2 \sqrt{-2 \ln r^2 / r^2} \quad (11.11)$$

ここで、 $r^2 = z_1^2 + z_2^2$  である

このときの  $y_1$  と  $y_2$  の同時分布は以下で与えられる

$$p(y_1, y_2) = \left[ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{y_1^2}{2} \right) \right] \left[ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{y_2^2}{2} \right) \right] \quad (11.12)$$

これにより、 $y_1$  と  $y_2$  は独立であり、それぞれが平均 0、分散 1 のガウス分布に従うため、ガウス分布からのサンプリングが可能となる

## 多変量ガウス分布の変換

もし  $y$  が平均 0、分散 1 のガウス分布に従うならば、 $\sigma y + \mu$  は平均  $\mu$ 、分散  $\sigma^2$  のガウス分布に従う

平均  $\mu$  と共分散  $\Sigma$  を持つ多変量ガウス分布に従うベクトル値変数を生成するために、コレスキー分解を利用する

$$\Sigma = LL^T \quad (\text{Press et al., 1992})$$

$z$  が平均 0、分散 1 の独立したガウス分布に従うベクトル値の確率変数であるとき、その要素が独立で平均 0、分散 1 のガウス分布に従うと仮定すると

$$y = \mu + Lz$$

は平均  $\mu$ 、共分散  $\Sigma$  を持つ

この方法を用いて、多変量ガウス分布に従うサンプルを生成することができる

## 逆関数法の適用限界

ここで紹介した逆関数法は、求めたい分布の不定積分の逆関数が解析的に求められる場合にのみ適用できる

例として、指数分布、コーシー分布、ガウス分布に適用した

実際はそのような単純な分布であることは稀であり、より一般的な戦略が必要である  
次に、棄却サンプリングと重点サンプリングという2つの手法を紹介する



## 11.1.2 棄却サンプリング

- まずは単変量分布を考え、その後に多次元への拡張について説明する
- 分布  $p(z)$  からサンプルを取得したいが、**直接サンプリングするのが難しい場合**を考える
- 任意の  $Z$  が与えられたときに、正規化定数  $Z_p$  を除いて  $p(z)$  を評価できると仮定

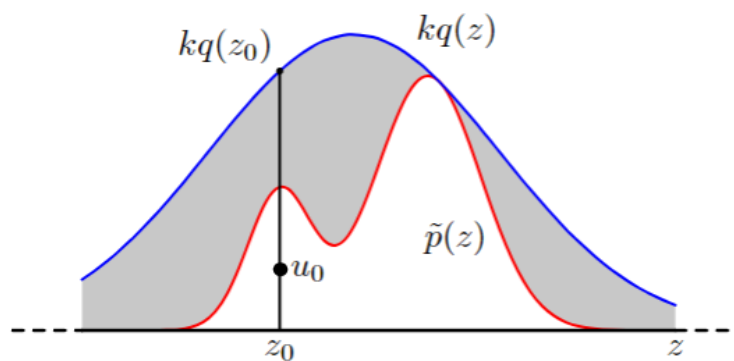
$$p(z) = \frac{1}{Z_p} \tilde{p}(z) \quad (11.13)$$

- $\tilde{p}(z)$  は容易に評価できるが、 $Z_p$  は未知
- ここで**簡単にサンプリングできる提案分布  $q(z)$  を使って  $p(z)$  を求めていく**
  - ベイズではないが提案分布を事前分布のように扱ってサンプリングを行う

# 棄却サンプリングの手順

1. 提案分布  $q(z)$  からサンプル  $z_0$  を生成
2.  $k$  を導入して分布  $kq(z)$  を作成(青線,  $k$ はすべての  $z$  に対して  $kq(z) \geq \tilde{p}(z)$  となるように調整したい)
3. 一様分布  $[0, kq(z_0)]$  からサンプル  $u_0$  を生成。
4.  $u_0 > \tilde{p}(z_0)$  ならサンプルを棄却(灰色の領域)し、そうでなければ受け入れる(赤線の内側)

**Figure 11.4** In the rejection sampling method, samples are drawn from a simple distribution  $q(z)$  and rejected if they fall in the grey area between the unnormalized distribution  $\tilde{p}(z)$  and the scaled distribution  $kq(z)$ . The resulting samples are distributed according to  $p(z)$ , which is the normalized version of  $\tilde{p}(z)$ .



## 受け入れ確率

- サンプルが受け入れられる確率は次のように与えられる

$$p(\text{accept}) = \int \frac{\tilde{p}(z)}{kq(z)} q(z) dz = \frac{1}{k} \int \tilde{p}(z) dz \quad (11.14)$$

- 棄却される点の割合は、非正規化分布  $\tilde{p}(z)$  の面積と比較関数  $kq(z)$  の面積の割合に依存する
- 当然だが、 $k$ が大きいほどreject領域が大きくなり、accept確率が小さくなるため、 $k$ の値はなるべく小さくしたほうが良い

## 例：ガンマ分布

$$\text{Gam}(z|a, b) = \frac{b^a z^{a-1} \exp(-bz)}{\Gamma(a)} \quad (11.15)$$

ここで、 $a$  は形状パラメータ,  $b$  は尺度パラメータ

$a > 1$  の場合、ガンマ分布は  $z$  に対して対数凹関数であるため、提案分布としては似た形かつ逆関数法が適用できるコーシー分布が適している

ただし、コーシー分布がガンマ分布よりも小さな値となるように一般化すると、 $z = b \tan y + c$  という変換を使用する

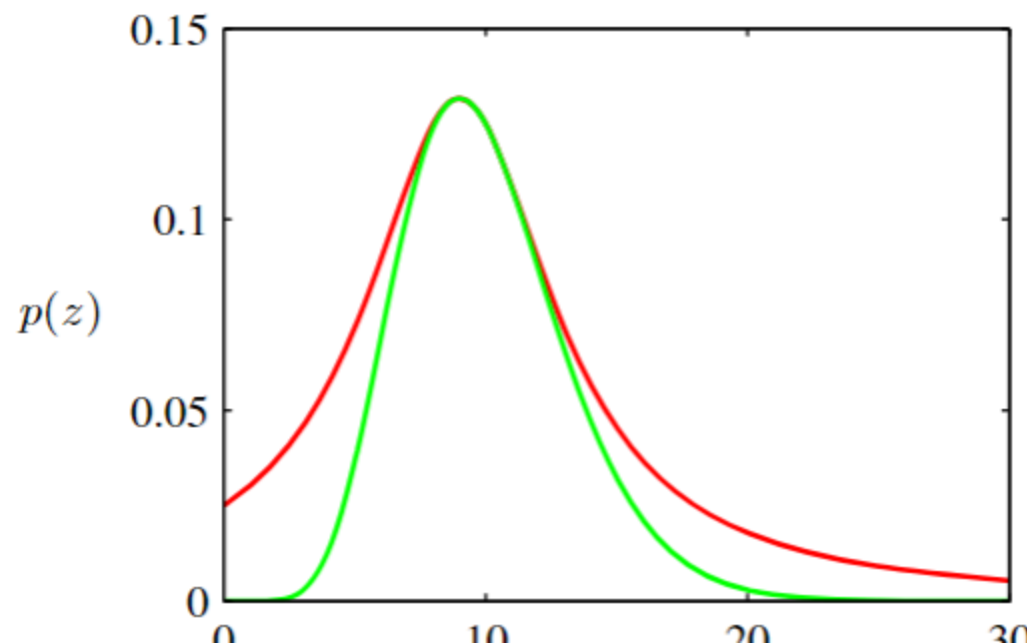
この変換により、乱数は

$$q(z) = \frac{k}{1 + \frac{(z-c)^2}{b^2}} \quad (11.16)$$

に従って分布する

ここで、棄却率を最小とするため、パラメータ  $c = a - 1$ ,  $b^2 = 2a - 1$  を選択する  
 $k$  は  $kq(z) \geq \tilde{p}(z)$  を満たした上でなるべく小さな値となるように選択する

**Figure 11.5** Plot showing the gamma distribution given by (11.15) as the green curve, with a scaled Cauchy proposal distribution shown by the red curve. Samples from the gamma distribution can be obtained by sampling from the Cauchy and then applying the rejection sampling criterion.



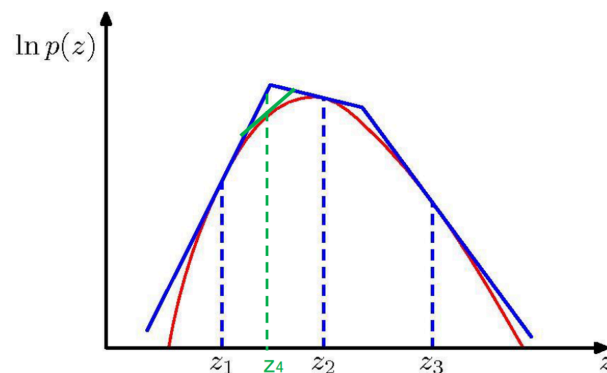
## 11.1.3 適応的棄却サンプリング

- 棄却サンプリングでは提案分布  $q(z)$  が既知だったが実際には未知の場合も多い
- 代替アプローチとして、分布  $p(z)$  の測定値に基づいて包絡関数を動的に構築する (Gilks and Wild, 1992)
  - この手法は、 $p(z)$  が対数凹関数( $\ln p(z)$  の導関数が  $z$  の非増加関数)である場合に特に簡単である

# 適応的棄却サンプリングの手順

1.  $q(z)$  で  $p(z)$  を覆うことができる、包絡線の初期点の集合を用意
2. 次の点  $z_4$  をサンプリングし、 $q(z)$  を計算
3. 一様分布  $[0, kq(z_4)]$  からサンプル  $u_0$  を生成。
4.  $u_0 > \tilde{p}(z_0)$  ならサンプルを棄却し、そうでなければ受け入れる
  - 棄却した場合も包絡線の更新に使われるため棄却サンプリングより無駄が少ない

**Figure 11.6** In the case of distributions that are log concave, an envelope function for use in rejection sampling can be constructed using the tangent lines computed at a set of grid points. If a sample point is rejected, it is added to the set of grid points and used to refine the envelope distribution.



↑説明のために、Fig. 11.6に $z_4$ を足した

## 包絡分布

包絡分布の対数は線形関数の連なりで、崩落分布自体も区分的指数分布で構成される

$$q(z) = k_i \lambda_i \exp\{-\lambda_i(z - z_{i-1})\} \quad z_{i-1} < z \leq z_i \quad (11.17)$$



## 対数凹関数以外への拡張

導関数の評価を避ける手法も存在する (Gilks, 1992)

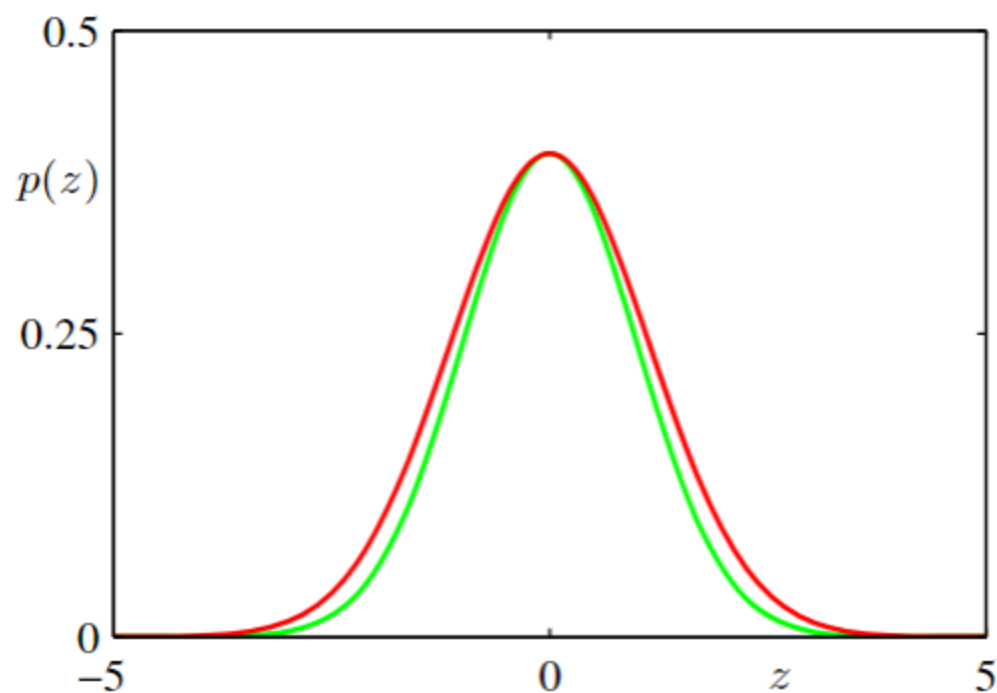
適応的棄却サンプリングの枠組みで、棄却サンプリングの各ステップの後にメトロポリス・ヘイスティングスステップ(11.2.2節で登場)を追加することで、対数凹関数でない分布にも拡張できる (Gilks et al., 1995)

# 高次元での棄却サンプリング(棄却サンプリング全般に当てはまる)

- 棄却サンプリングが実用的であるためには、比較関数が求めたい分布に近いことが重要である
- しかし、高次元空間で棄却サンプリングを使用すると、受け入れ率が次第に低下する
- 例えば、ゼロ平均で共分散が  $\sigma_p^2 I$  の多変量ガウス分布からサンプルを取得する問題を考える
  - 提案分布もゼロ平均のガウス分布であり、共分散が  $\sigma_q^2 I$  であるとする
- $kq(z) \geq p(z)$  となるように  $k$  が存在するには、 $\sigma_q^2 I > \sigma_p^2 I$  である必要がある
- この場合、最適な  $k$  の値は  $k = (\sigma_q / \sigma_p)^D$  であり、受け入れ率は  $kq(z)$  より下の体積と  $p(z)$  より下の体積の比であり、両方正規化されていて体積は同じであるため  $\frac{1}{k}$  となる
- つまり、受け入れ率は次元数に対して指数関数的に減少する

## 1次元での棄却サンプリングの例

**Figure 11.7** Illustrative example of rejection sampling involving sampling from a Gaussian distribution  $p(z)$  shown by the green curve, by using rejection sampling from a proposal distribution  $q(z)$  that is also Gaussian and whose scaled version  $kq(z)$  is shown by the red curve.



## まとめ

- 棄却サンプリングは提案分布や定数 $k$ の選択が難しく、より柔軟な包絡関数を提案分布とした適応的棄却サンプリングを提案
- 依然として、多峰性、鋭いピークを持つ分布の対応が難しい
- また、高次元空間での棄却率は指数関数的に増加するため、高次元への対応は難しい(せいぜい1,2次元)

## 11.1.4 重点サンプリング

サンプリング法の目的は(11.1) の形式の期待値を得ることであった  
重点サンプリングは、**分布からサンプルを抽出するのではなく、期待値を直接近似する**

## 基本的なアイデア

- $p(z)$  から直接サンプルを取得するのが実際には難しいが、 $p(z)$  を任意の  $z$  の値に対して計算できると仮定する
- 単純な戦略として、 $z$  空間を一様なグリッドに離散化し、次の形式の和として積分を評価する方法がある

$$E[f] \approx \frac{1}{L} \sum_{l=1}^L p(z^{(l)}) f(z^{(l)}) \quad (11.18)$$

- しかし、既に述べたように  $z$  の次元が大きい場合、和の項数が指数関数的に増加するため、この方法は実用的ではない
  - また、この場合はサンプルの一部が和に対して大きく寄与するため、一様サンプリングは非効率である
  - そこで、重点サンプリングを考える

## 重点サンプリングの手順

棄却サンプリングと同様に、重点サンプリングは提案分布  $q(z)$  を使用する  
**提案分布  $q(z)$  からサンプルを取得し、次の形式の有限和として期待値を表現する**

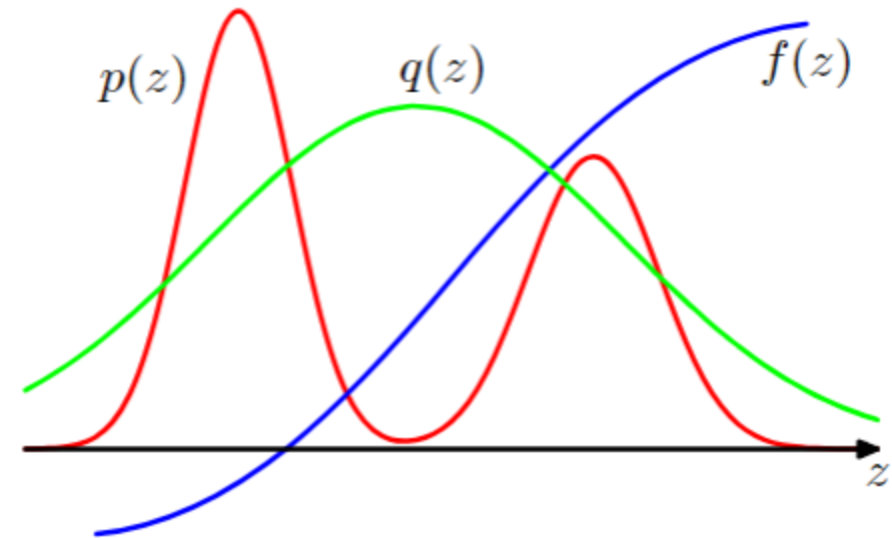
$$E[f] = \int f(z)p(z)dz = \int f(z)\frac{p(z)}{q(z)}q(z)dz \approx \frac{1}{L} \sum_{l=1}^L \frac{p(z^{(l)})}{q(z^{(l)})} f(z^{(l)}) \quad (11.19)$$

$r_l = \frac{p(z^{(l)})}{q(z^{(l)})}$  は重要度重みであり、誤った分布からサンプリングすることによって生じるバイアスを補正する

なお、棄却サンプリングとは異なり、生成されたすべてのサンプルが保持される

- 提案分布  $q(z)$  から少しずつ  $p(z)$  に近づけていく
- 目的の分布  $p(z)$  から外れていても棄却せず、小さい重みをつけた上で取り込む
- 形状が目的の分布から外れることもあるのでサンプリングに使うのは難しいが、期待値の近似としては使える

**Figure 11.8** Importance sampling addresses the problem of evaluating the expectation of a function  $f(z)$  with respect to a distribution  $p(z)$  from which it is difficult to draw samples directly. Instead, samples  $\{z^{(l)}\}$  are drawn from a simpler distribution  $q(z)$ , and the corresponding terms in the summation are weighted by the ratios  $p(z^{(l)})/q(z^{(l)})$ .





## 正規化定数の評価

$p(z) = \frac{1}{Z_p} \tilde{p}(z)$  の場合、正規化定数を除いてしか評価できない

同様に、重点サンプリング分布  $q(z)$  も正規化定数  $Z_q$  を持つと仮定する

$$E[f] = \frac{Z_q}{Z_p} \frac{1}{L} \sum_{l=1}^L \tilde{r}_l f(z^{(l)}) \quad (11.20)$$

ここで、 $\tilde{r}_l = \frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})}$  である

## 正規化定数の比の評価

同じサンプルセットを使用して、正規化定数の比  $Z_p/Z_q$  を評価できる

$$\frac{Z_p}{Z_q} = \frac{1}{L} \sum_{l=1}^L \tilde{r}_l \quad (11.21)$$

したがって、

$$E[f] \approx \sum_{l=1}^L w_l f(z^{(l)}) \quad (11.22)$$

ここで、

$$w_l = \frac{\tilde{r}_l}{\sum_m \tilde{r}_m} = \frac{\tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_m \tilde{p}(z^{(m)})/q(z^{(m)})} \quad (11.23)$$

## 重点サンプリングの成功

- 重点サンプリングの成功は、提案分布  $q(z)$  が目的の分布  $p(z)$  にどれだけ一致するか大きく依存する
- $p(z)f(z)$  が強く変動し、その大部分が  $z$  空間の比較的小さな領域に集中している場合、重要度重み  $\{r_l\}$  は大きな値を持ついくつかの重みによって支配される可能性がある
  - つまり重みが極端であり、重みが大きい領域をサンプリングしたときだけ大幅に動くが、小さい領域はほとんど動かないので機よが小さい
- この場合、**実効サンプルサイズ**は見かけのサンプルサイズ  $L$  よりもはるかに小さくなる可能性がある

## 11.1.5 SIR(sampling-importance-resampling)

棄却サンプリング法において定数  $k$  の適切な選択は実際には困難

SIRアプローチも提案分布  $q(z)$  を使用するが、定数  $k$  を決定する必要がない  
この手法は3つのステップ(SIR)からなる

1. 提案分布  $q(z)$  から  $L$  個のサンプル  $z^{(1)}, \dots, z^{(L)}$  を生成する
2. (11.23) 式を用いて重み  $w_1, \dots, w_L$  を計算する
3. 離散分布  $(z^{(1)}, \dots, z^{(L)})$  から  $L$  個のサンプルを再サンプリングする

重み  $w_l$  は次のように計算される

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

## 再サンプリングの結果

再サンプリングされた  $L$  個のサンプルは、 $p(z)$  に従って分布するが、これは  $L \rightarrow \infty$  の極限で正確になる

一変量の場合、再サンプリングされた値の累積分布は次のように与えられる

$$p(z \leq a) = \sum_{l: z^{(l)} \leq a} w_l = \frac{\sum_l I(z^{(l)} \leq a) \tilde{p}(z^{(l)}) / \tilde{q}(z^{(l)})}{\sum_l \tilde{p}(z^{(l)}) / \tilde{q}(z^{(l)})} \quad (11.25)$$

## 極限での挙動

$L \rightarrow \infty$  の極限を取ると、和は元のサンプリング分布  $q(z)$  に従って重み付けされた積分に置き換えられる

$$p(z \leq a) = \frac{\int I(z \leq a) \tilde{p}(z) / \tilde{q}(z) q(z) dz}{\int \tilde{p}(z) / \tilde{q}(z) q(z) dz} = \int I(z \leq a) p(z) dz \quad (11.26)$$

これは  $p(z)$  の累積分布関数である

## 有限サンプルの場合

有限の  $L$  の場合、再サンプリングされた値は目的の分布からのみ近似的に取得される  
棄却サンプリングと同様に、提案分布  $q(z)$  が目的の分布  $p(z)$  に近づくほど近似は改善される

## 期待値

分布  $p(z)$  に関するモーメント(期待値)が必要な場合、元のサンプルと重みを使用して直接評価できる

$$E[f(z)] = \int f(z)p(z)dz = \int f(z)\frac{\tilde{p}(z)}{q(z)}q(z)dz \approx \sum_{l=1}^L w_l f(z^{(l)}) \quad (11.27)$$



## まとめ

SIRは、定数  $k$  を決定する必要がないため、棄却サンプリングに対する有用な代替手法である

最初の提案分布  $q(z)$  からのリサンプリングばびでサンプリング回数が少ないと偏りが生じるが、無限回行くと近似できることが証明されている

高次元のサンプリングは依然として課題→MCMCへ

## 11.1.6 サンプリングとEMアルゴリズム

モンテカルロ法はベイズフレームワークを提供するだけでなく、頻度主義においても役割を果たす

例えば、最大尤度解を見つけるために使用される

特に、サンプリング法は、Eステップを解析的に実行できないモデルのEMアルゴリズムのEステップを近似するために使用できる

隠れ変数  $Z$ 、観測変数  $X$ 、パラメータ  $\theta$  を持つモデルを考える

Mステップで  $\theta$  に関して最適化される関数は、次のように与えられる完全データ対数尤度の期待値である

$$Q(\theta, \theta_{\text{old}}) = \int p(Z|X, \theta_{\text{old}}) \ln p(Z, X|\theta) dZ \quad (11.28)$$

サンプリング法を使用して、この積分を現在の事後分布  $p(Z|X, \theta_{\text{old}})$  から抽出されたサンプル  $\{Z^{(l)}\}$  による有限和で近似する

$$Q(\theta, \theta_{\text{old}}) \approx \frac{1}{L} \sum_{l=1}^L \ln p(Z^{(l)}, X|\theta) \quad (11.29)$$

Q関数は通常の方法でMステップで最適化される

この手順はモンテカルロEMアルゴリズムと呼ばれる

事前分布  $p(\theta)$  が定義されている場合、Q関数に  $\ln p(\theta)$  を追加してからMステップを実行することで、 $\theta$  の事後分布のモード（MAP推定）を見つける問題に拡張できる

## 統計的EM

モンテカルロEMアルゴリズムの特定のインスタンスである統計的EMは、有限混合モデルを考え、各Eステップで1つのサンプルを抽出する場合に発生する

ここで、潜在変数  $Z$  は、各データポイントを生成する混合成分のどれかを特徴付ける  
Eステップでは、データセット  $X$  に対して事後分布  $p(Z|X, \theta_{\text{old}})$  から  $Z$  のサンプルを抽出する

これにより、各データポイントを混合成分の1つにハードに割り当てる

Mステップでは、このサンプル化された事後分布の近似を使用して、通常の方法でモデルパラメータを更新する

# データ補完アルゴリズム

最大尤度アプローチから完全なベイズ処理に移行し、パラメータベクトル  $\theta$  の事後分布からサンプルを取得する場合を考える

原則として、結合事後分布  $p(\theta, Z|X)$  からサンプルを取得したいが、これは計算的に困難であると仮定する

完全データパラメータ事後分布  $p(\theta|Z, X)$  からサンプルを取得するのが比較的簡単であると仮定する

これにより、データ補完アルゴリズムがインスパイアされ、Iステップ（Eステップに類似）とPステップ（Mステップに類似）の2つのステップを交互に実行する

# IPアルゴリズム

## Iステップ

$p(Z|X)$  からサンプルを取得したいが、直接行うことはできない  
したがって、次の関係を利用する

$$p(Z|X) = \int p(Z|\theta, X)p(\theta|X)d\theta \quad (11.30)$$

まず、現在の事後分布  $p(\theta|X)$  からサンプル  $\theta^{(l)}$  を抽出し、次にこれを使用して  $p(Z|\theta^{(l)}, X)$  からサンプル  $Z^{(l)}$  を抽出する

## Pステップ

次の関係を利用する

$$p(\theta|X) = \int p(\theta|Z, X)p(Z|X)dZ \quad (11.31)$$

Iステップから得られたサンプル  $\{Z^{(l)}\}$  を使用して、次のように  $\theta$  の事後分布の改訂推定を計算する

$$p(\theta|X) \approx \frac{1}{L} \sum_{l=1}^L p(\theta|Z^{(l)}, X) \quad (11.32)$$

## まとめ

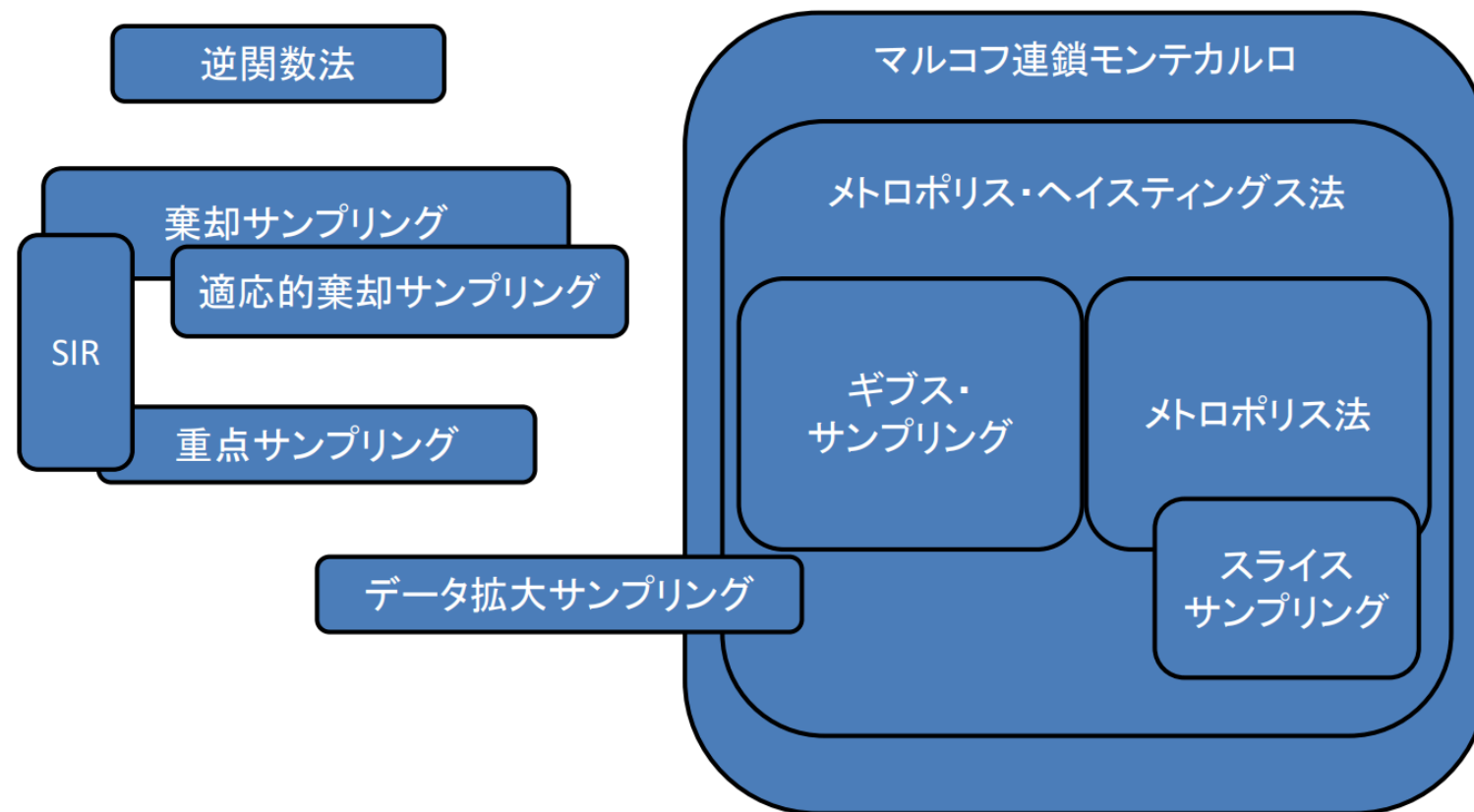
サンプリングとEMアルゴリズムは、ベイズフレームワークと頻度主義の両方で重要な役割を果たす

モンテカルロ法は、解析的に実行できないEステップを近似するために使用される

データ補完アルゴリズムは、完全なベイズ処理において事後分布からサンプルを取得するための強力な手法である



# サンプリング法の俯瞰マップ



以下より引用

[http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11\\_chika.pdf](http://bin.t.u-tokyo.ac.jp/summercamp2015/document/prml11_chika.pdf)