機械学習 Machine Learning

サンプリング法 Sampling Methods

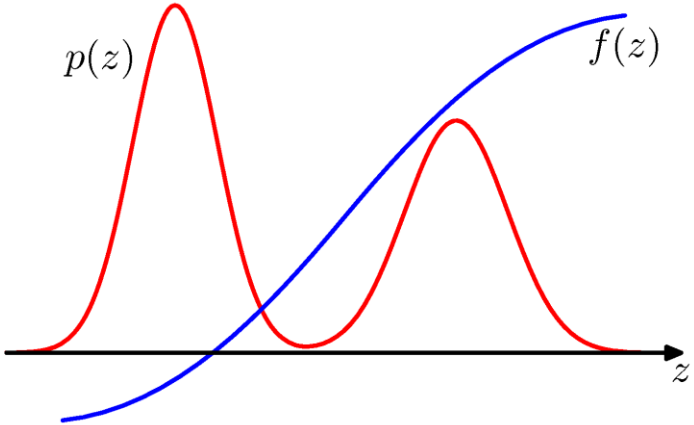
福嶋 誠 Makoto Fukushima

情報科学部
School of Informatics and Data Science

基本的なサンプリングアルゴリズム Basic sampling algorithms
期待値 Expectations

以下の期待値を解析的な方法を用いて厳密に評価するには複雑すぎると仮定する. We suppose that the following expectation is too complex to be evaluated exactly using analytical techniques.

$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z}$$



基本的なサンプリングアルゴリズム Basic sampling algorithms
期待値 Expectations

$p(\mathbf{z})$ から独立に抽出されたサンプルの集合 $\mathbf{z}^{(l)}$（$l = 1, \ldots, L$）より, $\mathbb{E}[f]$を有限和で近似できる.
A set of samples $\mathbf{z}^{(l)}$ (where $l = 1, \ldots, L$) drawn independently from $p(\mathbf{z})$ allows $\mathbb{E}[f]$ to be approximated by a finite sum:

$$\overline{f} = \frac{1}{L} \sum_{l=1}^{L} f(\mathbf{z}^{(l)}) \simeq \mathbb{E}[f(\mathbf{z})]$$

このとき
where

$$\mathbb{E}[\overline{f}] = \mathbb{E}[f(\mathbf{z})]$$

$$\mathrm{var}[\overline{f}] = \frac{1}{L} \mathbb{E}\left[(f - \mathbb{E}[f])^2\right]$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
標準的な分布 Standard distributions

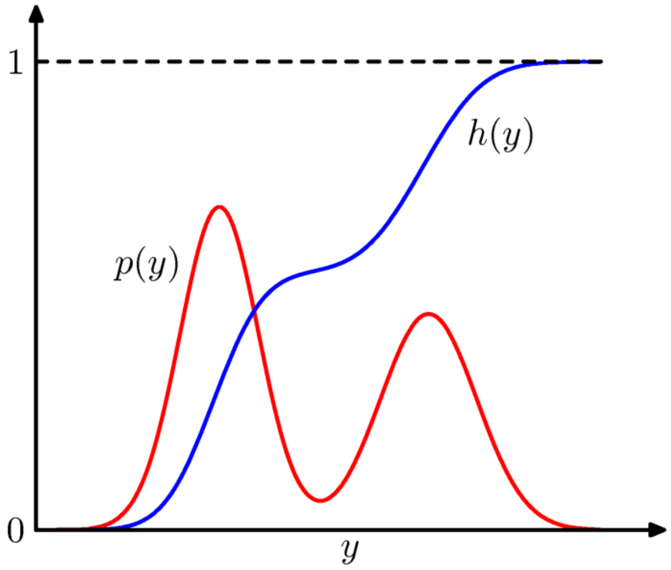
ある関数 $g(\cdot)$を用いて，区間 $(0,1)$ で一様に分布する z の値を $y=g(z)$ と変換する.
We transform the values of z that is uniformly distributed over the interval $(0,1)$ using some function $g(\cdot)$ so that $y=g(z)$.

yの分布は
The distribution of y is

$$p(y) = p(z)\left|\frac{\mathrm{d}z}{\mathrm{d}y}\right|$$

ただし where

$$p(z) = 1$$



<u>目的 Goal</u>

変換後の y の値が求めたい特定の分布 $p(y)$ に従うように $g(z)$ を選ぶこと.
Choose $g(z)$ such that resulting values of y have some specific desired distribution $p(y)$.

基本的なサンプリングアルゴリズム Basic sampling algorithms
標準的な分布 Standard distributions

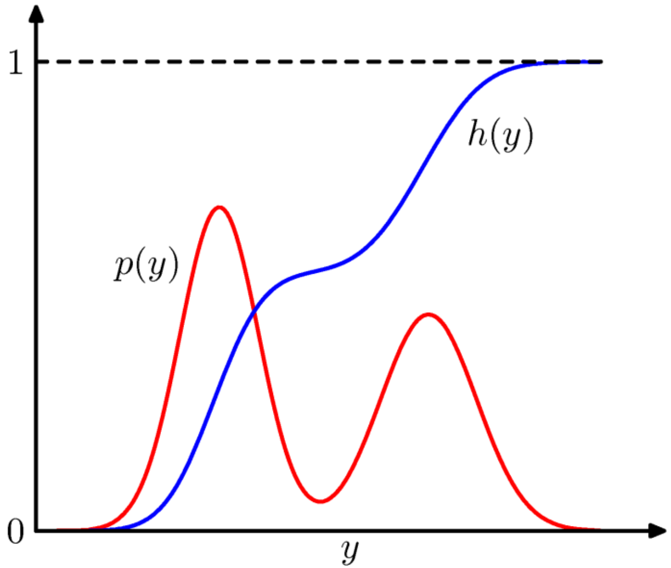
ある関数 $g(\cdot)$を用いて，区間 $(0,1)$ で一様に分布する z の値を $y=g(z)$ と変換する.
We transform the values of z that is uniformly distributed over the interval $(0,1)$ using some function $g(\cdot)$ so that $y=g(z)$.

次式を積分すると
Integrating

$$p(y) = p(z)\left|\frac{\mathrm{d}z}{\mathrm{d}y}\right|$$

$$\Rightarrow \quad z = \int_{-\infty}^{y} p(\widehat{y})\,\mathrm{d}\widehat{y} \equiv h(y)$$

$y = h^{-1}(z)$ を $g(z)$ として選ぶ.
Choose $y = h^{-1}(z)$ as $g(z)$.



基本的なサンプリングアルゴリズム Basic sampling algorithms
標準的な分布 Standard distributions

例 Examples

指数分布 Exponential distribution

$$p(y) = \lambda \exp(-\lambda y) \quad 0 \leqslant y < \infty$$

⇨

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

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

コーシー分布 Cauchy distribution

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

⇨

$$z = h(y) = \frac{1}{\pi} \tan^{-1}(y) + \frac{1}{2}$$

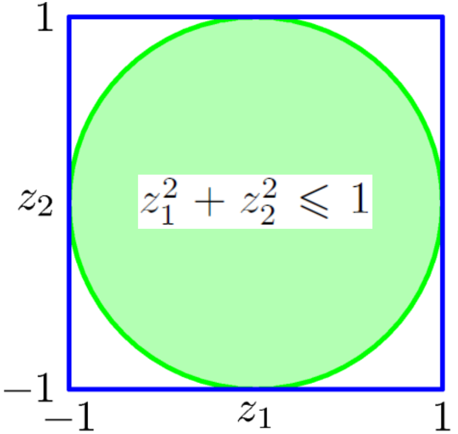
$$y = \tan\left(\pi\left(z - \frac{1}{2}\right)\right)$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
標準的な分布 Standard distributions

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

$$z \rightarrow 2z - 1$$

$$z_1, z_2 \in (-1, 1)$$



$$p(z_1, z_2) = 1/\pi$$

平均0，分散1のガウス分布
Gaussian distribution with zero mean and unit variance

$$\begin{aligned} y_1 &= z_1 \left(\frac{-2 \ln r^2}{r^2} \right)^{1/2} \\ y_2 &= z_2 \left(\frac{-2 \ln r^2}{r^2} \right)^{1/2} \end{aligned}$$

$$r^2 = z_1^2 + z_2^2$$

$$\begin{aligned} p(y_1, y_2) &= p(z_1, z_2) \left| \frac{\partial(z_1, z_2)}{\partial(y_1, y_2)} \right| \\ &= \left[\frac{1}{\sqrt{2\pi}} \exp(-y_1^2/2) \right] \left[\frac{1}{\sqrt{2\pi}} \exp(-y_2^2/2) \right] \end{aligned}$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
標準的な分布 Standard distributions

平均 μ，分散 σ^2 のガウス分布
Gaussian distribution with mean μ and variance σ^2

$$\mu + \sigma y$$

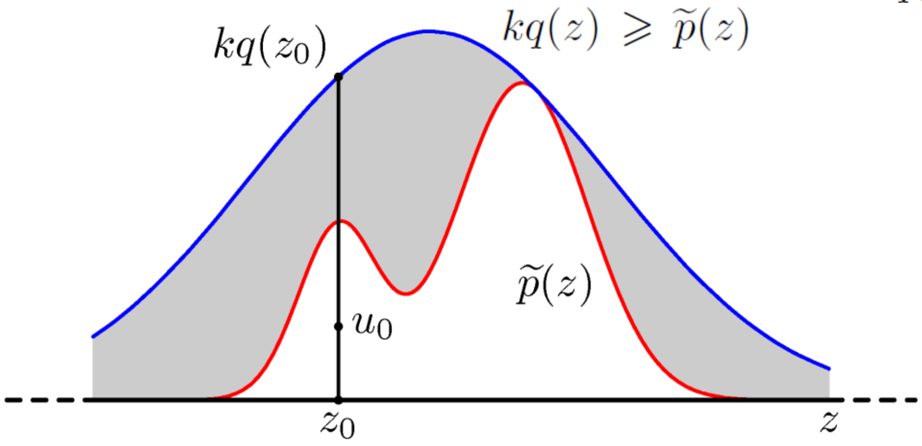
平均 $\boldsymbol{\mu}$，共分散 $\boldsymbol{\Sigma}$ のガウス分布
Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

$$\boldsymbol{\mu} + \mathbf{L}\mathbf{y} \qquad \text{ただし where } \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
棄却サンプリング Rejection sampling

$$p(z) = \frac{1}{Z_p}\widetilde{p}(z)$$

$\widetilde{p}(z)$ はすぐに求められるが, Z_p はわからないとする.
$\widetilde{p}(z)$ readily be evaluated , but Z_p is unknown.



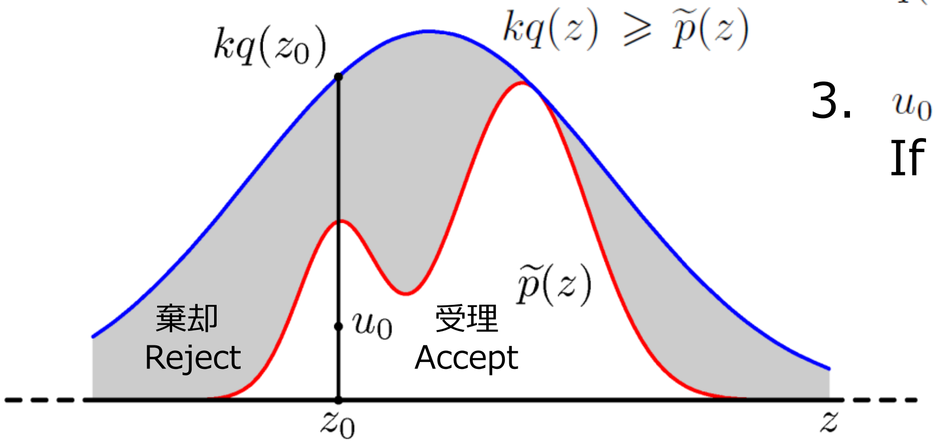
$q(z)$: 提案分布 Proposal distribution

1. z_0 を $q(z)$ から生成する.
   Generate z_0 from $q(z)$.
2. u_0 を $[0, kq(z_0)]$ 上の一様分布から生成する.
   Generate u_0 from the uniform distribution over $[0, kq(z_0)]$.

基本的なサンプリングアルゴリズム Basic sampling algorithms
棄却サンプリング Rejection sampling

$$p(z) = \frac{1}{Z_p}\widetilde{p}(z)$$

$\widetilde{p}(z)$ はすぐに求められるが, Z_p はわからないとする.
$\widetilde{p}(z)$ readily be evaluated , but Z_p is unknown.



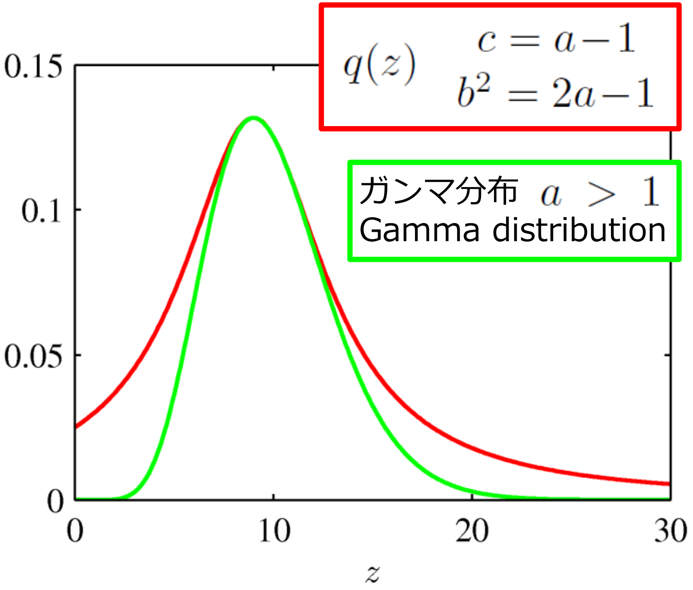
$q(z)$: 提案分布 Proposal distribution

3. $u_0 > \widetilde{p}(z_0)$ ならば，サンプルを棄却する.
If $u_0 > \widetilde{p}(z_0)$, the sample is rejected.

受理

$$\begin{aligned} p(\text{accept}) &= \int \{\widetilde{p}(z)/kq(z)\}\, q(z)\,\mathrm{d}z \\ &= \frac{1}{k}\int \widetilde{p}(z)\,\mathrm{d}z. \end{aligned}$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
棄却サンプリング Rejection sampling



ガンマ分布 Gamma distribution

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

提案分布 Proposal distribution
(スケールされたコーシー分布 Scaled Cauchy distribution)

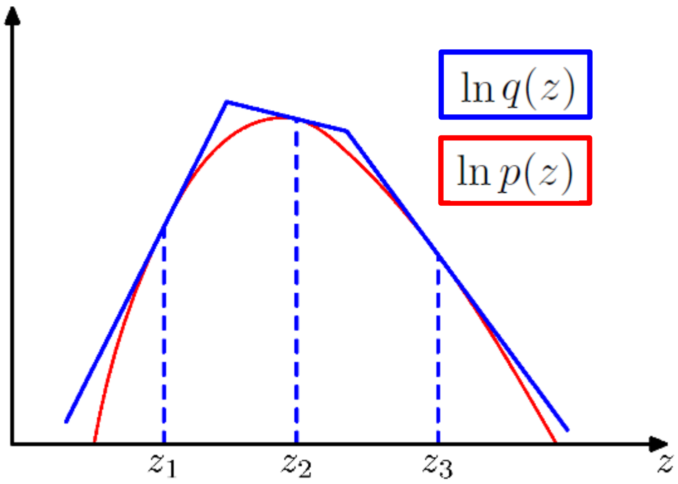
$$q(z) = \frac{k}{1+(z-c)^2/b^2}$$

$$\uparrow \quad z = b\tan y + c$$

区間 $(0,1)$ の一様分布から生成した y
y that is generated from the uniform distribution over the interval $(0,1)$.

基本的なサンプリングアルゴリズム Basic sampling algorithms
適応的棄却サンプリング Adaptive rejection sampling

Gilks and Wild (1992)



対数凹な分布 Log concave distribution

$$p(z)$$

提案分布 Proposal distribution
(区分的な指数分布 Piecewise exponential distribution)

$$q(z) = k_i \lambda_i \exp\{-\lambda_i (z - z_i)\}$$

$$\widehat{z}_{i-1,i} < z \leqslant \widehat{z}_{i,i+1}$$

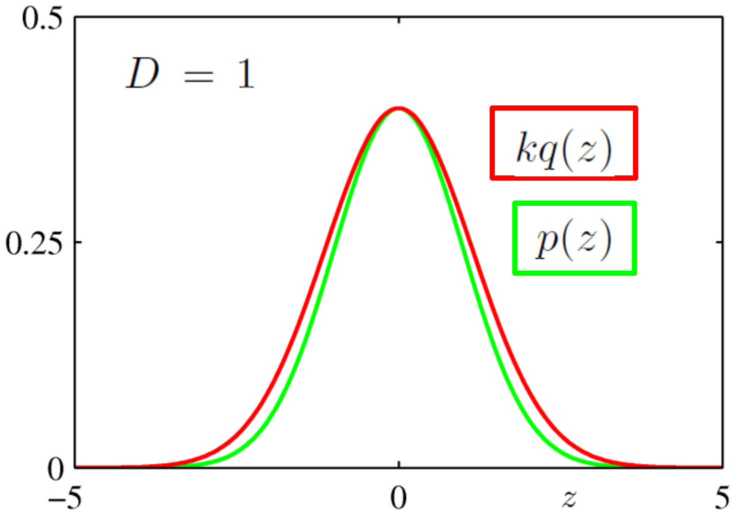
サンプルが棄却された場合には，グリッド点の集合に組み入れられ，新しい接線が計算され，包絡関数がそれによって改良される.
If the sample was rejected, it is incorporated into the set of grid points, a new tangent line is computed, and the envelope function is thereby refined.

基本的なサンプリングアルゴリズム Basic sampling algorithms
適応的棄却サンプリング Adaptive rejection sampling

棄却サンプリングの問題点 Problems of rejection sampling

ゼロ平均ガウス分布 Zero-mean Gaussian distribution

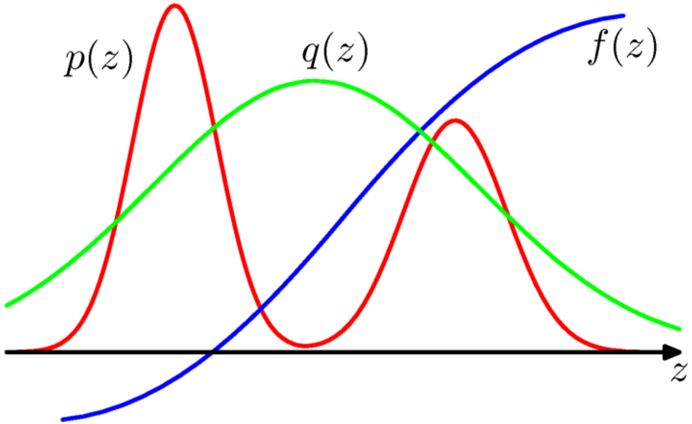
$$q(z) \quad p(z) \quad \sigma_q^2 \geqslant \sigma_p^2$$



最適な k Optimal k $\quad k = (\sigma_q/\sigma_p)^D$

$\sigma_q = 1.01\,\sigma_p$
$D = 1{,}000$
→ 受理率 Acceptance ratio $1/20{,}000$

基本的なサンプリングアルゴリズム Basic sampling algorithms
重点サンプリング Importance sampling



$q(\mathbf{z})$ を用いて期待値 $\mathbb{E}[f]$ を近似する.
Approximate the expectation $\mathbb{E}[f]$ using $q(\mathbf{z})$.

$$\begin{aligned} \mathbb{E}[f] &= \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &= \int f(\mathbf{z})\frac{p(\mathbf{z})}{q(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &\simeq \frac{1}{L}\sum_{l=1}^{L}\frac{p(\mathbf{z}^{(l)})}{q(\mathbf{z}^{(l)})}f(\mathbf{z}^{(l)}) \end{aligned}$$

$\{\mathbf{z}^{(l)}\}$: $q(\mathbf{z})$ から抽出されたサンプル
Samples drawn from $q(\mathbf{z})$

$r_l = p(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})$: 重要度重み
Importance weights

基本的なサンプリングアルゴリズム Basic sampling algorithms
重点サンプリング Importance sampling

正規化定数がわからないとき When normalization constants are unknown,

$$p(\mathbf{z}) = \widetilde{p}(\mathbf{z})/Z_p \qquad q(\mathbf{z}) = \widetilde{q}(\mathbf{z})/Z_q \qquad Z_p = \int \widetilde{p}(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

$$\begin{aligned}\mathbb{E}[f] &= \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &= \frac{Z_q}{Z_p}\int f(\mathbf{z})\frac{\widetilde{p}(\mathbf{z})}{\widetilde{q}(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &\simeq \frac{Z_q}{Z_p}\frac{1}{L}\sum_{l=1}^{L}\widetilde{r}_l f(\mathbf{z}^{(l)})\end{aligned}$$

$$\begin{aligned}\frac{Z_p}{Z_q} &= \frac{1}{Z_q}\int \widetilde{p}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int \frac{\widetilde{p}(\mathbf{z})}{\widetilde{q}(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &\simeq \frac{1}{L}\sum_{l=1}^{L}\widetilde{r}_l \qquad \widetilde{r}_l = \widetilde{p}(\mathbf{z}^{(l)})/\widetilde{q}(\mathbf{z}^{(l)})\end{aligned}$$

⇨ $\mathbb{E}[f] \simeq \sum_{l=1}^{L} w_l f(\mathbf{z}^{(l)})$ ただし where $w_l = \frac{\widetilde{r}_l}{\sum_m \widetilde{r}_m} = \frac{\widetilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \widetilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}$

基本的なサンプリングアルゴリズム Basic sampling algorithms
サンプリング重点リサンプリング Sampling-importance-resampling

定数 k を決定する必要がないサンプリング法
A sampling method that avoids having to determine the constant k

1. L 個のサンプル $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ が $q(\mathbf{z})$ から抽出される.
   L samples $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ are drawn from $q(\mathbf{z})$.

2. 重み $w_1, \ldots, w_L$ が左式を用いて求められる.
   Weights $w_1, \ldots, w_L$ are constructed using

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

3. 2セット目の L 個のサンプルが離散分布 $(\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)})$ から重み $(w_1, \ldots, w_L)$ で与えられる確率に従って抽出される.
   A second set of L samples is drawn from the discrete distribution $(\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)})$ with probabilities given by the weights $(w_1, \ldots, w_L)$.

基本的なサンプリングアルゴリズム Basic sampling algorithms
サンプリング重点リサンプリング Sampling-importance-resampling

定数 k を決定する必要がないサンプリング法
A sampling method that avoids having to determine the constant k

$$p(z \leqslant a) = \sum_{l:z^{(l)} \leqslant a} w_l$$

$$= \frac{\sum_l I(z^{(l)} \leqslant a)\widetilde{p}(z^{(l)})/q(z^{(l)})}{\sum_l \widetilde{p}(z^{(l)})/q(z^{(l)})}$$

$$\xrightarrow{L \to \infty}$$

$$p(z \leqslant a) = \frac{\int I(z \leqslant a)\,\{\widetilde{p}(z)/q(z)\}\,q(z)\,\mathrm{d}z}{\int \{\widetilde{p}(z)/q(z)\}\,q(z)\,\mathrm{d}z}$$

$$= \frac{\int I(z \leqslant a)\widetilde{p}(z)\,\mathrm{d}z}{\int \widetilde{p}(z)\,\mathrm{d}z}$$

$$= \int I(z \leqslant a)p(z)\,\mathrm{d}z$$

$$\boxed{w_l = \frac{\widetilde{r}_l}{\sum_m \widetilde{r}_m} = \frac{\widetilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \widetilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}}$$

基本的なサンプリングアルゴリズム Basic sampling algorithms
サンプリング重点リサンプリング Sampling-importance-resampling

定数 k を決定する必要がないサンプリング法
A sampling method that avoids having to determine the constant k

$$\begin{aligned}\mathbb{E}[f(\mathbf{z})] &= \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &= \frac{\int f(\mathbf{z})[\widetilde{p}(\mathbf{z})/q(\mathbf{z})]q(\mathbf{z})\,\mathrm{d}\mathbf{z}}{\int [\widetilde{p}(\mathbf{z})/q(\mathbf{z})]q(\mathbf{z})\,\mathrm{d}\mathbf{z}} \\ &\simeq \sum_{l=1}^{L} w_l f(\mathbf{z}_l)\end{aligned}$$

$$\boxed{w_l = \frac{\widetilde{r}_l}{\sum_m \widetilde{r}_m} = \frac{\widetilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \widetilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}}$$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo

$p(\mathbf{z}) = \widetilde{p}(\mathbf{z})/Z_p$

$\widetilde{p}(\mathbf{z})$: 計算可能 can be evaluated

Z_p : わからない unknown

提案分布 Proposal distribution: $q(\mathbf{z}|\mathbf{z}^{(\tau)})$

$\mathbf{z}^{(\tau)}$: 現在の状態 Current state

サンプルの系列 $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots$ はマルコフ連鎖を成す.
The sequence of samples $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots$ forms a Markov chain.

アルゴリズムの各サイクルで，サンプル候補 $\mathbf{z}^{\star}$ を提案分布から生成し，
適切な基準に従ってサンプルを受理する.
At each cycle of the algorithm, we generate a candidate sample $\mathbf{z}^{\star}$ from the proposal distribution and then accept the sample according to an appropriate criterion.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
メトロポリス法 The Metropolis algorithm

Metropolis et al. (1953)

仮定 Assumption

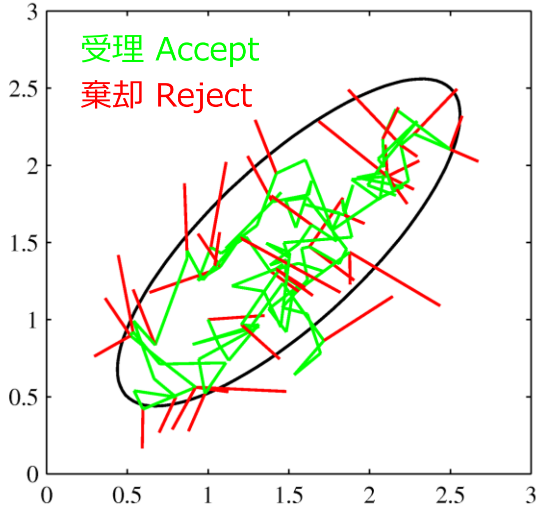
$$q(\mathbf{z}_A|\mathbf{z}_B) = q(\mathbf{z}_B|\mathbf{z}_A)$$

サンプル候補は以下の確率で受理される.
The candidate sample is accepted with probability

$$A(\mathbf{z}^\star, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^\star)}{\widetilde{p}(\mathbf{z}^{(\tau)})}\right)$$

任意の $\mathbf{z}_A$, $\mathbf{z}_B$ の値に対して $q(\mathbf{z}_A|\mathbf{z}_B)$ が正であるならば,
If $q(\mathbf{z}_A|\mathbf{z}_B)$ is positive for any values of $\mathbf{z}_A$ and $\mathbf{z}_B$,

$\mathbf{z}^{(\tau)}$の分布 The distribution of $\mathbf{z}^{(\tau)}$ $\xrightarrow{\tau \to \infty}$ $p(\mathbf{z})$



マルコフ連鎖モンテカルロ Markov chain Monte Carlo
メトロポリス法 The Metropolis algorithm

Metropolis et al. (1953)

Input: Unnormalized distribution $\widetilde{p}(\mathbf{z})$
Proposal distribution $q(\mathbf{z}|\widehat{\mathbf{z}})$
Initial state $\mathbf{z}^{(0)}$
Number of iterations T

Output: $\mathbf{z} \sim p(\mathbf{z})$

```
z_prev ← z^(0)
// Iterative message-passing
for τ ∈ {1, ..., T} do
    z* ~ q(z|z_prev)  // Sample from proposal distribution
    u ~ U(0,1)  // Sample from uniform
    if p̃(z*) / p̃(z_prev) > u then
        z_prev ← z*  // z^(τ) = z*
    else
        z_prev ← z_prev  // z^(τ) = z^(τ−1)
    end if
end for
return z_prev  // z^(T)
```

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
マルコフ連鎖 Markov chains

マルコフ連鎖 Markov chain $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}$

$$p(\mathbf{z}^{(m+1)}|\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}) = p(\mathbf{z}^{(m+1)}|\mathbf{z}^{(m)}) \qquad m \in \{1, \ldots, M-1\}$$

遷移確率 Transition probabilities $T_m(\mathbf{z}^{(m)}, \mathbf{z}^{(m+1)}) \equiv p(\mathbf{z}^{(m+1)}|\mathbf{z}^{(m)})$

遷移確率がすべての m について同一であるとき : **<u>均一</u>**マルコフ連鎖
If the transition probabilities are the same for all m: **<u>Homogeneous</u>** Markov chain

周辺分布 Marginal distribution $p(\mathbf{z}^{(m+1)}) = \int p(\mathbf{z}^{(m+1)}|\mathbf{z}^{(m)})p(\mathbf{z}^{(m)})\,\mathrm{d}\mathbf{z}^{(m)}$

分布がマルコフ連鎖の各ステップで変わらないとき，その分布はそのマルコフ連鎖に関して**<u>不変</u>**（定常）
If each step in a Markov chain leaves the distribution invariant, it is **<u>invariant</u>** (stationary) w.r.t. the Markov chain.

遷移確率が $T(\mathbf{z}', \mathbf{z})$ の均一マルコフ連鎖に対して以下が成り立つとき $p^\star(\mathbf{z})$ は不変.
For a homogeneous Markov chain with $T(\mathbf{z}', \mathbf{z})$, $p^\star(\mathbf{z})$ is invariant if

$$p^\star(\mathbf{z}) = \int T(\mathbf{z}', \mathbf{z})p^\star(\mathbf{z}')\,\mathrm{d}\mathbf{z}'$$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
マルコフ連鎖 Markov chains

求めたい分布 $p(\mathbf{z})$ が不変であることを保証する十分条件は，分布 $p^\star(\mathbf{z})$ に対して
A sufficient condition for ensuring that the required distribution $p(\mathbf{z})$ is invariant is to choose the transition probabilities to satisfy the property of **detailed balance**, defined by

$$p^\star(\mathbf{z})T(\mathbf{z},\mathbf{z}') = p^\star(\mathbf{z}')T(\mathbf{z}',\mathbf{z})$$

で定義される**詳細釣り合い条件**が満たされるように遷移確率を選ぶことである.
for the particular distribution $p^\star(\mathbf{z})$.

なぜなら
Because

$$\begin{aligned}\int p^\star(\mathbf{z}')T(\mathbf{z}',\mathbf{z})\,\mathrm{d}\mathbf{z}' &= \int p^\star(\mathbf{z})T(\mathbf{z},\mathbf{z}')\,\mathrm{d}\mathbf{z}' \\ &= p^\star(\mathbf{z})\int p(\mathbf{z}'|\mathbf{z})\,\mathrm{d}\mathbf{z}' \\ &= p^\star(\mathbf{z}).\end{aligned}$$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
マルコフ連鎖 Markov chains

目的：与えられた分布からのサンプリングをマルコフ連鎖を用いて行うこと.
Goal: Use Markov chains to sample from a given distribution.

求めたい分布が不変となるようなマルコフ連鎖を用意し・・・
We can achieve this if we set up a Markov chain such that the desired distribution is invariant and ...

加えて，$m \to \infty$ のとき，初期分布 $p(\mathbf{z}^{(0)})$ の選択にかかわらず，分布 $p(\mathbf{z}^{(m)})$ が求めたい不変分布 $p^{\star}(\mathbf{z})$ に収束すれば達成できる（**エルゴード性**）.
and if the distribution $p(\mathbf{z}^{(m)})$ converges to the required invariant distribution $p^{\star}(\mathbf{z})$ for $m \to \infty$, irrespective of the choice of initial distribution $p(\mathbf{z}^{(0)})$ (**ergodicity**).

均一マルコフ連鎖は，弱い制限を設けるとエルゴード的になる. Neal (1993)
A homogeneous Markov chain will be ergodic, subject to weak restrictions.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
マルコフ連鎖 Markov chains

実際には多くの場合，遷移確率は「基本」遷移 $B_1, \ldots, B_K$ の組から構築される.
In practice, we often construct the transition probabilities from a set of 'base' transitions $B_1, \ldots, B_K$.

<u>例 Examples:</u>

$$T(\mathbf{z}', \mathbf{z}) = \sum_{k=1}^{K} \alpha_k B_k(\mathbf{z}', \mathbf{z}) \qquad \alpha_k \geqslant 0 \quad \sum_k \alpha_k = 1$$

$$T(\mathbf{z}', \mathbf{z}) = \sum_{\mathbf{z}_1} \ldots \sum_{\mathbf{z}_{K-1}} B_1(\mathbf{z}', \mathbf{z}_1) \ldots B_{K-1}(\mathbf{z}_{K-2}, \mathbf{z}_{K-1}) B_K(\mathbf{z}_{K-1}, \mathbf{z})$$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
メトロポリス–ヘイスティングス法 The Metropolis–Hastings algorithm

Hastings (1970)

分布 $q_k(\mathbf{z}|\mathbf{z}^{(\tau)})$ からサンプル $\mathbf{z}^\star$ を抽出し，以下の確率に従って受理する.
We draw a sample $\mathbf{z}^\star$ from the distribution $q_k(\mathbf{z}|\mathbf{z}^{(\tau)})$ and then accept it with probability

$$A_k(\mathbf{z}^\star, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^\star)q_k(\mathbf{z}^{(\tau)}|\mathbf{z}^\star)}{\widetilde{p}(\mathbf{z}^{(\tau)})q_k(\mathbf{z}^\star|\mathbf{z}^{(\tau)})}\right)$$

$$\boxed{p(\mathbf{z}) = \widetilde{p}(\mathbf{z})/Z_p}$$

以下の詳細釣り合い条件により，上記のマルコフ連鎖で $p(\mathbf{z})$ が不変分布となる.
$p(\mathbf{z})$ is an invariant distribution of the Markov chain above because of the detailed balance below.

$$\begin{aligned}
p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z})A_k(\mathbf{z}', \mathbf{z}) &= \min\left(p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z}), p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}')\right) \\
&= \min\left(p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}'), p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z})\right) \\
&= p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}')A_k(\mathbf{z}, \mathbf{z}')
\end{aligned}$$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
メトロポリス–ヘイスティングス法 The Metropolis–Hastings algorithm

Hastings (1970)

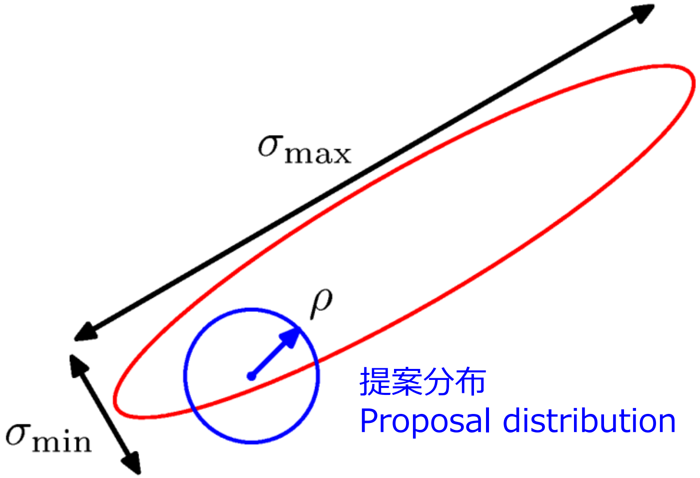
Input: Unnormalized distribution $\widetilde{p}(\mathbf{z})$
Proposal distributions $\{q_k(\mathbf{z}|\widehat{\mathbf{z}}) : k \in 1, \ldots, K\}$
Mapping from iteration index to distribution index $M(\cdot)$
Initial state $\mathbf{z}^{(0)}$
Number of iterations T

Output: $\mathbf{z} \sim p(\mathbf{z})$

$\mathbf{z}_{\mathrm{prev}} \leftarrow \mathbf{z}^{(0)}$
// Iterative message-passing
for $\tau \in \{1, \ldots, T\}$ **do**
  $k \leftarrow M(\tau)$ // get distribution index for this iteration
  $\mathbf{z}^{\star} \sim q_k(\mathbf{z}|\mathbf{z}_{\mathrm{prev}})$ // sample from proposal distribution
  $u \sim \mathcal{U}(0,1)$ // sample from uniform
  if $\widetilde{p}(\mathbf{z}^{\star})q(\mathbf{z}_{\mathrm{prev}}|\mathbf{z}^{\star}) \,/\, \widetilde{p}(\mathbf{z}_{\mathrm{prev}})q(\mathbf{z}^{\star}|\mathbf{z}_{\mathrm{prev}}) > u$ **then**
    $\mathbf{z}_{\mathrm{prev}} \leftarrow \mathbf{z}^{\star}$ // $\mathbf{z}^{(\tau)} = \mathbf{z}^{\star}$
  else
    $\mathbf{z}_{\mathrm{prev}} \leftarrow \mathbf{z}_{\mathrm{prev}}$ // $\mathbf{z}^{(\tau)} = \mathbf{z}^{(\tau-1)}$
  end if
end for
return $\mathbf{z}_{\mathrm{prev}}$ // $\mathbf{z}^{(T)}$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
メトロポリス–ヘイスティングス法 The Metropolis–Hastings algorithm

Hastings (1970)



もともとの状態から多少なりとも独立な状態に達するのに必要なステップ数は $(\sigma_{\max}/\sigma_{\min})^2$ のオーダーとなる.
The number of steps to arrive at a state that is more or less independent of the original state is of order $(\sigma_{\max}/\sigma_{\min})^2$.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
ギブスサンプリング Gibbs sampling

Geman and Geman (1984)

$$p(\mathbf{z}) = p(z_1, \ldots, z_M)$$ サンプリングしたい分布
The distribution from which we wish to sample

z_i を分布 $p(z_i|\mathbf{z}_{\setminus i})$ から抽出する．ただし $\mathbf{z}_{\setminus i}$ は $\{z_1, \ldots, z_M\}$ から z_i を除いたもの.
Draw z_i from the distribution $p(z_i|\mathbf{z}_{\setminus i})$, where $\mathbf{z}_{\setminus i}$ is $\{z_1, \ldots, z_M\}$ with z_i omitted.

例 Example $p(z_1, z_2, z_3)$

1. $z_1^{(\tau)}$を $p(z_1|z_2^{(\tau)}, z_3^{(\tau)})$ から抽出して得た $z_1^{(\tau+1)}$ で置き換える.
   Replace $z_1^{(\tau)}$ by $z_1^{(\tau+1)}$ obtained by sampling from $p(z_1|z_2^{(\tau)}, z_3^{(\tau)})$.
2. $z_2^{(\tau)}$を $p(z_2|z_1^{(\tau+1)}, z_3^{(\tau)})$ から抽出して得た $z_2^{(\tau+1)}$ で置き換える.
   Replace $z_2^{(\tau)}$ by $z_2^{(\tau+1)}$ obtained by sampling from $p(z_2|z_1^{(\tau+1)}, z_3^{(\tau)})$.
3. $z_3^{(\tau)}$を $p(z_3|z_1^{(\tau+1)}, z_2^{(\tau+1)})$から抽出して得た $z_3^{(\tau+1)}$ で置き換える.
   Replace $z_3^{(\tau)}$ by $z_3^{(\tau+1)}$ obtained by sampling from $p(z_3|z_1^{(\tau+1)}, z_2^{(\tau+1)})$.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
ギブスサンプリング Gibbs sampling

Geman and Geman (1984)

Input: Initial values $\{z_i : i \in 1, \ldots, M\}$
Conditional distributions $\{p(z_i|\{z_{j\neq i}\}) : i \in 1, \ldots, M\}$
Number of iterations T
Output: Final values $\{z_i : i \in 1, \ldots, M\}$

---

for $\tau \in \{1, \ldots, T\}$ **do**
  for $i \in \{1, \ldots, M\}$ **do**
    $z_i \sim p(z_i|\{z_{j\neq i}\})$
  end for
end for
return $\{z_i : i \in 1, \ldots, M\}$

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
ギブスサンプリング Gibbs sampling

Geman and Geman (1984)

ギブスサンプリングの性質 Properties of Gibbs sampling

分布 $p(\mathbf{z})$ はサンプリングの各ステップで不変.
The distribution $p(\mathbf{z})$ is invariant of each of sampling steps.

条件付き分布の確率が0となる場合がなければ，エルゴード性を満たす.
It is ergodic if none of the conditional distributions are anywhere zero.

メトロポリス–ヘイスティングス法の特別な場合とみなすことができる.
It can be regarded as a special case of the Metropolis–Hastings algorithm.

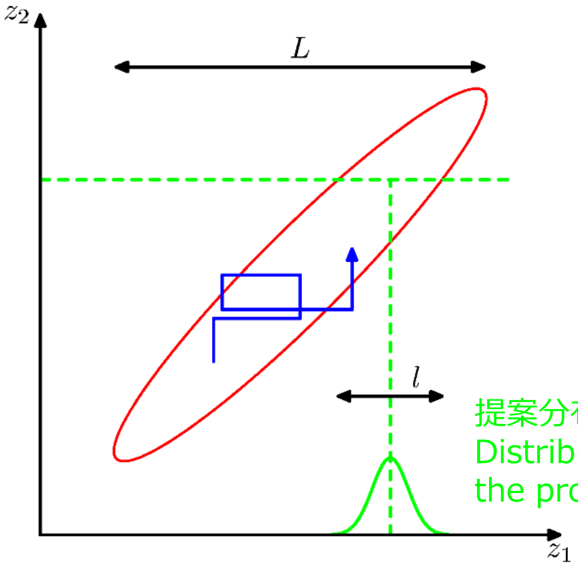
$$q_k(\mathbf{z}^\star|\mathbf{z}) = p(z_k^\star|\mathbf{z}_{\setminus k}) \qquad \mathbf{z}^\star_{\setminus k} = \mathbf{z}_{\setminus k}$$

$$\Rightarrow \quad A(\mathbf{z}^\star, \mathbf{z}) = \frac{p(\mathbf{z}^\star)q_k(\mathbf{z}|\mathbf{z}^\star)}{p(\mathbf{z})q_k(\mathbf{z}^\star|\mathbf{z})} = \frac{p(z_k^\star|\mathbf{z}^\star_{\setminus k})p(\mathbf{z}^\star_{\setminus k})p(z_k|\mathbf{z}^\star_{\setminus k})}{p(z_k|\mathbf{z}_{\setminus k})p(\mathbf{z}_{\setminus k})p(z_k^\star|\mathbf{z}_{\setminus k})} = 1$$

常に受理
Always accepted

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
ギブスサンプリング Gibbs sampling

Geman and Geman (1984)



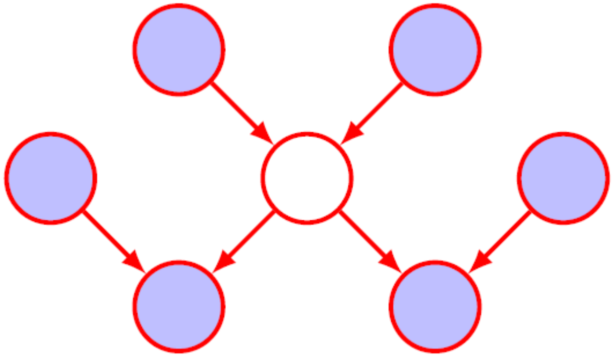
提案分布に相当する分布
Distribution corresponding to the proposal distribution

独立なサンプルを分布から得るために必要なステップ数は $(L/l)^2$ のオーダーとなる.
The number of steps needed to obtain independent samples from the distribution will be of order $(L/l)^2$.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
ギブスサンプリング Gibbs sampling

Geman and Geman (1984)

有向グラフィカルモデルによって定まる確率分布の場合，各ノードの条件付き分布 $p(z_k|\mathbf{z}_{\setminus k})$ は対応するマルコフブランケットの中の変数にのみ依存する．
For probability distributions specified using directed graphical models, the conditional distributions for individual nodes $p(z_k|\mathbf{z}_{\setminus k})$ depend only on the variables in the corresponding Markov blanket.



マルコフ連鎖モンテカルロ Markov chain Monte Carlo
伝承サンプリング Ancestral sampling

伝承サンプリング Ancestral sampling

変数の集合を $\mathbf{z}_1, \ldots, \mathbf{z}_M$ の順番で $p(\mathbf{z}_i|\mathrm{pa}(i))$ からサンプリングする.
We make one pass through the set of variables in the order $\mathbf{z}_1, \ldots, \mathbf{z}_M$ sampling from the conditional distributions $p(\mathbf{z}_i|\mathrm{pa}(i))$.

$$p(\mathbf{z}) = \prod_{i=1}^{M} p(\mathbf{z}_i|\mathrm{pa}(i))$$

ロジックサンプリング Logic sampling Henrion (1988)

各ステップで，観測値をもつ変数 $\mathbf{z}_i$ のサンプル値が得られたとき，
サンプル値は観測値と比較される.
At each step, when a sampled value is obtained for $\mathbf{z}_i$ whose value is observed, the sampled value is compared to the observed value.

マルコフ連鎖モンテカルロ Markov chain Monte Carlo
伝承サンプリング Ancestral sampling

尤度重み付きサンプリング Likelihood weighted sampling

Fung and Chang (1990)
Shachter and Peot (1990)

各変数を順番に，もしその変数が証拠集合に含まれるならば，単にその値を実現値に設定する．もし証拠集合に含まれないなら，$p(\mathbf{z}_i|\mathrm{pa}(i))$からサンプリングする．
For each variable in turn, if that variable is in the evidence set, then it is just the set to its instantiated value. If it is not in the evidence set, then it is sampled from $p(\mathbf{z}_i|\mathrm{pa}(i))$.

$\mathbf{e}$: 証拠集合 Evidence set
（観測変数の集合 The set of observed variables）

サンプリング結果 $\mathbf{z}$ に対する重み付けは
The weighting associated with the resulting sample $\mathbf{z}$ is given by

$$r(\mathbf{z}) = \prod_{\mathbf{z}_i \notin \mathbf{e}} \frac{p(\mathbf{z}_i|\mathrm{pa}(i))}{p(\mathbf{z}_i|\mathrm{pa}(i))} \prod_{\mathbf{z}_i \in \mathbf{e}} \frac{p(\mathbf{z}_i|\mathrm{pa}(i))}{1} = \prod_{\mathbf{z}_i \in \mathbf{e}} p(\mathbf{z}_i|\mathrm{pa}(i))$$

ランジュバンサンプリング Langevin sampling
エネルギーベースモデル Energy-based models

LeCun et al. (2006)

$$p(\mathbf{x}|\mathbf{w}) = \frac{1}{Z(\mathbf{w})} \exp\{-E(\mathbf{x},\mathbf{w})\}$$

$E(\mathbf{x},\mathbf{w})$: エネルギー関数 Energy function

$Z(\mathbf{w})$: 分配関数 Partition function

わからない unknown

$$\int p(\mathbf{x}|\mathbf{w})\,\mathrm{d}\mathbf{x} = 1 \quad \Rightarrow \quad Z(\mathbf{w}) = \int \exp\{-E(\mathbf{x},\mathbf{w})\}\,\mathrm{d}\mathbf{x}$$

データセット $\mathcal{D} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ に対する対数尤度関数
The log likelihood function for a data set $\mathcal{D} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$

$$\ln p(\mathcal{D}|\mathbf{w}) = -\sum_{n=1}^{N} E(\mathbf{x}_n,\mathbf{w}) - N \ln Z(\mathbf{w})$$

ランジュバンサンプリング Langevin sampling
尤度最大化 Maximizing the likelihood

単一のデータ点 $\mathbf{x}$ に対して
For a single data point $\mathbf{x}$

$$\nabla_{\mathbf{w}} \ln p(\mathbf{x}|\mathbf{w}) = -\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w}) - \nabla_{\mathbf{w}} \ln Z(\mathbf{w})$$

$p_{\mathcal{D}}(\mathbf{x})$についての期待値
The expectation w.r.t. $p_{\mathcal{D}}(\mathbf{x})$

$$\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\nabla_{\mathbf{w}} \ln p(\mathbf{x}|\mathbf{w})\right] = -\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\right] - \nabla_{\mathbf{w}} \ln Z(\mathbf{w})$$

$$-\nabla_{\mathbf{w}} \ln Z(\mathbf{w}) = \int \{\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\}\, p(\mathbf{x}|\mathbf{w})\, \mathrm{d}\mathbf{x}$$

$$p(\mathbf{x}|\mathbf{w}) = \frac{1}{Z(\mathbf{w})} \exp\{-E(\mathbf{x},\mathbf{w})\}$$

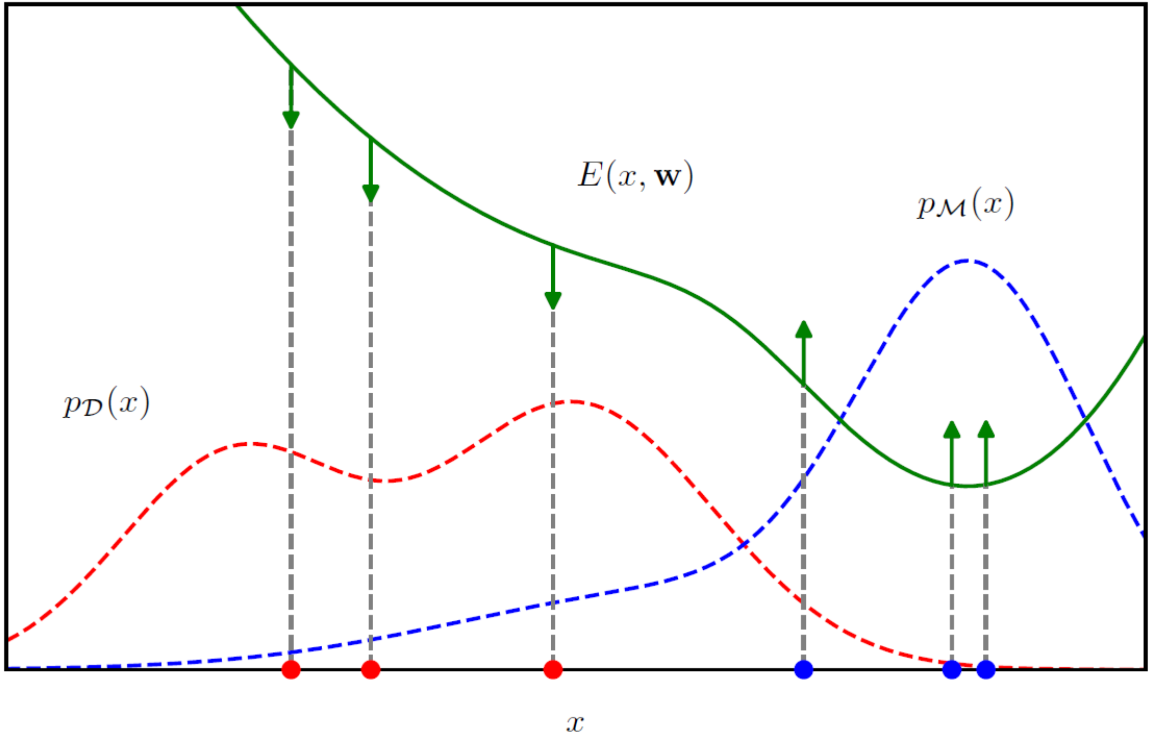
$$Z(\mathbf{w}) = \int \exp\{-E(\mathbf{x},\mathbf{w})\}\, \mathrm{d}\mathbf{x}$$

$$\Rightarrow \quad \nabla_{\mathbf{w}} \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\ln p(\mathbf{x}|\mathbf{w})\right] = -\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\right] + \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{M}}(\mathbf{x})}\left[\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\right]$$

ただし where $p_{\mathcal{M}}(\mathbf{x}) = p(\mathbf{x}|\mathbf{w})$

ランジュバンサンプリング Langevin sampling
尤度最大化 Maximizing the likelihood

$$\nabla_{\mathbf{w}} \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\ln p(\mathbf{x}|\mathbf{w})\right] = -\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\right] + \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{M}}(\mathbf{x})}\left[\nabla_{\mathbf{w}} E(\mathbf{x},\mathbf{w})\right]$$



ランジュバンサンプリング Langevin sampling
ランジュバン動力学 Langevin dynamics

Parisi (1981), Welling and Teh (2011)

$$\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}}\left[\nabla_{\mathbf{w}}E(\mathbf{x},\mathbf{w})\right] \simeq \frac{1}{N}\sum_{n=1}^{N}\nabla_{\mathbf{w}}E(\mathbf{x}_n,\mathbf{w})$$

$$\mathbb{E}_{\mathbf{x}\sim p_{\mathcal{M}}(\mathbf{x})}\left[\nabla_{\mathbf{w}}E(\mathbf{x},\mathbf{w})\right] \simeq \frac{1}{M}\sum_{m=1}^{M}\nabla_{\mathbf{w}}E(\mathbf{x}_m,\mathbf{w})$$

$$\{\mathbf{x}_1,\ldots,\mathbf{x}_M\}$$

ランジュバンサンプリング Langevin sampling

$$\tau \in 1,\ldots,\mathcal{T} \quad \mathbf{x}^{(\tau+1)} = \mathbf{x}^{(\tau)} + \eta\nabla_{\mathbf{x}}\ln p(\mathbf{x}^{(\tau)},\mathbf{w}) + \sqrt{2\eta}\boldsymbol{\epsilon}^{(\tau)} \quad \boldsymbol{\epsilon}^{(\tau)} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$$

スコア関数 Score function $\mathbf{s}(\mathbf{x},\mathbf{w}) = \nabla_{\mathbf{x}}\ln p(\mathbf{x}|\mathbf{w}) = -\nabla_{\mathbf{x}}E(\mathbf{x},\mathbf{w})$

$\eta \to 0$, $\mathcal{T} \to \infty$ ⇨ $\mathbf{z}^{(\mathcal{T})}$ の値は分布 $p(\mathbf{x})$ からの独立サンプルとなる.
The value of $\mathbf{z}^{(\mathcal{T})}$ is an independent sample from the distribution $p(\mathbf{x})$.

ランジュバンサンプリング Langevin sampling
ランジュバン動力学 Langevin dynamics

Input: Initial value $\mathbf{x}^{(0)}$

Probability density $p(\mathbf{x}, \mathbf{w})$

Learning rate parameter η

Number of iterations T

Output: Final value $\mathbf{x}^{(T)}$

$\mathbf{x} \leftarrow \mathbf{x}_0$

for $\tau \in \{1, \ldots, T\}$ **do**

  $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \mathbf{I})$

  $\mathbf{x} \leftarrow \mathbf{x} + \eta \nabla_{\mathbf{x}} \ln p(\mathbf{x}, \mathbf{w}) + \sqrt{2\eta}\boldsymbol{\epsilon}$

end for

return $\mathbf{x}$ // Final value $\mathbf{x}^{(T)}$