概要

　前回はサンプリング法の概要と、基本的なサンプリング法について学んだ。サンプリング法の目的は多くの場合、期待値計算のような積分を評価することである。サンプルを得ることで、積分値の近似をサンプルの有限和で表現することができる（(11.2)式）。サンプルを得る基本的な方法として、「変換法」と「棄却サンプリング」を学んだ。

　今回の範囲である 11.1.3 - 11.1.6 では引き続きいくつかのサンプリング手法を学ぶ。ここで扱うサンプリング手法は互いに関連はしているものの、ストーリー性が乏しい点に注意しながら読みたい。各サンプリング手法がどのように関連しているかをレジュメの最後にまとめるので、理解が一致しているか確認する。

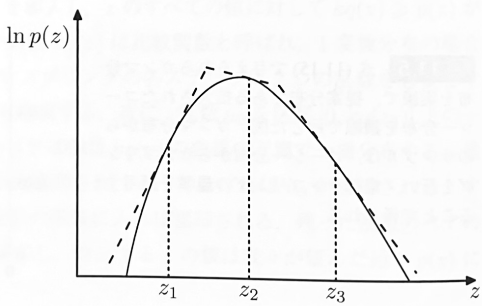
11.1.3 適応的棄却サンプリング

　棄却サンプリングでは受理確率が大きくするために、真の分布に近い提案分布（＝包絡分布？）を設定することが重要であった。しかし、現実には真の分布の形状が分からないまま包絡分布を設定することは難しい。

　適応的棄却サンプリングでは、新しく得られたサンプル値を元に包絡分布を更新し、真の分布に近い包絡分布を得る。ただし、前提として真の分布の対数 $\ln p(z)$ について導関数が計算可能で、さらに $\ln p(z)$ が上に凸な関数である必要がある。

　まず、初期の包絡関数を構築する。そのために、初期のサンプルとそのサンプルにおける $\ln p(z)$ の勾配を用いて、図11.6 破線のような包絡関数 $\ln q(z)$ を構築する。

図 11.6 対数凹である分布の場合，棄却サンプリングで用いる包絡関数はグリッド点の集合において計算された接線を用いて構築できる．サンプル点が棄却されたならば，グリッド点の集合に追加され，包絡分布を改良するために用いられる．



ここで、 z_i における接線部分の包絡関数はを次のように表せる。

$$\ln q(z) = -\lambda_i(z - z_i) + \log(k_i \lambda_i)$$

両辺に指数をとると、包絡分布としての形式が得られる。

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

$$= k_i \lambda_i \exp\{-\lambda_i(z-z_i)\} \qquad (11.17)$$

包絡分布の式が求まると、通常の棄却サンプリングを実施できる。すなわち、z_0 を q(z) に従う分布（正規化したもの）から生成し、次に乱数 u_0 を区間 [0, q(z_0)] 上の一様分布から生成する。$u_0 > p(z_0)$ の場合はサンプルは棄却され、そうでなければサンプルは保持される。

包絡関数の更新は、サンプルが棄却されたときに発動する。サンプルが受理されにくい点で包絡関数が更新されることで、分布 $p(z)$ をよりよく近似できるようになる。

この手続きは、包絡関数が真の分布よりも上部に位置していることを仮定している。これは $\ln p(z)$ が上に凸な関数であることより明らかである。この仮定を必要としない拡張として**適応的棄却 Metropolis サンプリング**と呼ばれるものもあるが、ここでは詳しく解説しない。

この後の教科書の話のつながりが上手く解き明かせなかった。高次元空間における棄却サンプリングの課題について解説しているが、これは「適応的」棄却サンプリングの話ではないように思える。さらに、後ろに続く節もこの問題を解決するものではない。

11.1.4 重点サンプリング

サンプリング法の目的の多くは、期待値のような積分値を計算することであると述べた。**重点サンプリング**では、積分値を直接有限和で近似する手法である。そのため、真の分布 $p(z)$ に従うサンプルは得られない。これは、真の分布に従うサンプルを得てから積分近似を得る棄却サンプリングとは対照的である。

重点サンプリングの説明に入る前に、期待値を近似するためのナイーブな方法を導入する。前提として、真の分布 $p(z)$ から直接サンプリングすることは困難で、与えられた z について $p(z)$ を計算することは簡単にできるとする。この時、z 空間を均一なグリッドで離散化し、積分を次のように近似することができる。

$$
\begin{aligned}
E[f] &= \int f(z)\,p(z)\,dz \\
&\simeq \frac{1}{L}\sum_{l=1}^{L} p(z^{(l)})\,f(z^{(l)}) \qquad (11.18)
\end{aligned}
$$

このアプローチの問題点は次の2点である。

- 和をとる項の数が、z の次元に対して指数的に大きくなる
- 真の分布 $p(z)$ が非常に小さい領域に高い確率であるとき、近似値が少数のサンプルに依存する

以上の問題より、理想的には、$p(z)\,f(z)$ が大きくなるような領域を重点的にサンプリングしたい。

重点サンプリングでは提案分布を利用して、重要度の高いサンプリングを重点的にサンプリングする。これは棄却サンプリングで提案分布を利用していることと同様のアプローチである。この時の期待値は次のような有限和で近似できる。

$$\begin{aligned}
E[f] &= \int f(z)\, p(z)\, dz \\
&= \int f(z) \frac{p(z)}{q(z)} q(z)\, dz \\
&= E_{q(z)}\left[\frac{p(z)}{q(z)} f(z) \right] \\
&\simeq \frac{1}{L} \sum_{l=1}^{L} \frac{p(z^{(l)})}{q(z^{(l)})} f(z^{(l)}) \\
&= \frac{1}{L} \sum_{l=1}^{L} \underbrace{r_l}_{\text{重要度重み}} f(z^{(l)})
\end{aligned}$$

この式は、 $f(z^{(l)})$ の値を、 $z^{(l)}$ の重要度で重みづけた式となっている。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は保持されていることに注意する。代わりに、 r_l が小さいサンプルについては小さい重みが割り当てられ、期待値への寄与が小さくなる。

次に、分布 $p(z)$ が $p(z) = \tilde{p}(z)/Z_p$ のように、正規化定数を除いてしか評価できないときを考える。すなわち、 Z_p はわからないが、 $\tilde{p}(z)$ は容易に計算できるとする。同様に、提案分布についても $q(z) = \tilde{q}(z)/Z_q$ を用いる場合を考える。このときの期待値は次のように近似できる。

$$\begin{aligned}
E[f] &= \int f(z)\, p(z)\, dz \\
&= \int f(z) \frac{Z_q\, \tilde{p}(z)}{Z_p\, \tilde{q}(z)} q(z)\, dz \\
&= \frac{Z_q}{Z_p} \int f(z) \frac{\tilde{p}(z)}{\tilde{q}(z)} q(z)\, dz \\
&\simeq \frac{Z_q}{Z_p} \frac{1}{L} \sum_{l=1}^{L} \frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})} f(z^{(l)}) \\
&= \frac{Z_q}{Z_p} \frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l f(z^{(l)}) \qquad \left(\text{ここで } \tilde{r}_l \equiv \frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})} \right) \\
&= Z_q \left(\int \tilde{p}(z)\, dz \right)^{-1} \frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l f(z^{(l)})
\end{aligned}$$

$$= Z_q \left(\int \frac{Z_p \tilde{p}(z)}{\tilde{q}(z)} q(z) dz \right)^{-1} \frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l f(z^{(l)})$$

$$= \left(\int \frac{\tilde{p}(z)}{\tilde{q}(z)} q(z) dz \right)^{-1} \frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l f(z^{(l)})$$

$$\simeq \left(\frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l \right)^{-1} \frac{1}{L} \sum_{l=1}^{L} \tilde{r}_l f(z^{(l)})$$

$$= \sum_{l=1}^{L} \frac{\tilde{r}_l}{\sum_{m=1}^{L} \tilde{r}_m} f(z^{(l)})$$

$$= \sum_{l=1}^{L} w_l f(z^{(l)}) \qquad \left(ここで \ w_l = \frac{r_l}{\sum_m \tilde{r}_m} \right)$$

このように近似式が得られることから、真の分布と提案分布は必ずしも正規化されている必要がないことがわかる。重点サンプリングが成功するか否かは、提案分布が求めたい分布にどの程度適合しているかに強く影響される。これは棄却サンプリングと同様である。例えば、**z** 空間の比較的小さな領域に高い確率が集中するなら、重要度重みの集合 $\{r_l\}$ は大きな値を持つ少数の重みに支配され、他の多数の重みは相対的に意味のないものになる。そうなると実効的なサンプルサイズは見かけのサンプルサイズよりもずっと小さくなる（章冒頭の「実効的なサンプルサイズ」の意味がようやくわかる）。

もし、 $p(z)$ の大きな領域にどのサンプルも入らなければ、問題はもっと深刻になる。この場合、たとえ期待値の推定値が大きく誤っていたとしても、r_l と $r_l f(z^{(l)})$ の見かけの分散は小さい。すなわち、どのサンプルの重要度重みもそれなりに高い値をとり、一見良いサンプルを得られているように見えてしまう。このように、重点サンプリング法はたとえ近似値の誤差が大きい場合でも、それを診断するための指標が得られないという大きな欠点がある。

この後のグラフィカルモデルの話が理解できなかったので読み合わせたい。特に未知の用語についての説明がないので、諸々確信がもてなかった。「証拠集合」「実現値」「数式中の変数e」など。

11.1.5 SIR

棄却サンプリングが成功するかどうかは、適切な定数 k を設定できるかに依存している。k が小さすぎる場合は $k\,q(z)$ が $p(z)$ を上から抑えることができないが、k が大きすぎる場合は受理率が小さくなってしまう。

SIR（**sampling-importance-resampling**）は棄却サンプリングと同様にサンプリング分布（=提案分布と同一？）を利用するが、定数 k を決定する必要はない。この方式は次の2段階からなる。

■ ステップ 1

L 個のサンプル $z^{(1)}, \cdots, z^{(L)}$ を $q(z)$ から抽出する。

■ ステップ2

次式を用いて重み $(w_1, \cdots, w_L)$ を求める。

$$w_l = \frac{r_l}{\sum \tilde{r}_m} = \frac{\hat{p}(z^{(l)})/q(z^{(l)})}{\sum_m \tilde{p}(z^{(m)})/q(z^{(m)})} \qquad (11.23)$$

2セット目の L 個のサンプルを、離散分布 $(z^{(1)}, \cdots, z^{(L)})$ から重み $(w_1, \cdots, w_L)$ で与えられる確率に従って抽出し、これを出力のサンプルとする。

結果として得られる L 個のサンプルは、分布 $p(z)$ に近似的に従って（「近似的に従う」のイメージがむずい）分布しているに過ぎないが、$L \to \infty$ の極限では分布に正確に従う。これを1変数の場合で確かめる。再サンプリングされた値の累積分布は次のようになる。

$$\hat{p}(z \le a) = \sum_{l: z^{(l)} \le a} w_l \qquad (*)$$

SIR により得られた
サンプルの累積分布なので、$p(z \le a)$ とは別物として置く

(*) の理由

$$\begin{cases} \hat{p}(z = z_1) = w_1 \\ \hat{p}(z = z_2) = w_2 \\ \quad \vdots \\ \hat{p}(z = z_L) = w_L \end{cases}$$

より、$\hat{p}(z \le a)$ は $z_l \le a$ となる l について w_l を足し合わせれば良い。

$$= \sum_l I(z^{(l)} \le a)\, w_l$$

$$= \frac{\sum_l I(z^{(l)} \le a)\ \tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_m \tilde{p}(z^{(m)})/q(z^{(m)})} \qquad (11.25)$$

一方で、求めたい分布の累積分布は次のように変形できる。これは上記の (11.25) 式の極限と一致する。

$$
\begin{aligned}
P(z \leq a) &= \int \mathbb{I}(z \leq a)\, p(z)\, dz \\
&= \frac{\int \mathbb{I}(z \leq a)\, \tilde{p}(z)\, dz}{\int \tilde{p}(z)\, dz} \\
&= \frac{\int \mathbb{I}(z \leq a) \{\tilde{p}(z)/q(z)\}\, q(z)\, dz}{\int \{\tilde{p}(z)/q(z)\}\, q(z)\, dz} \\
&= \frac{\lim_{L\to\infty} \frac{1}{L} \sum_{l}^{L} \mathbb{I}(z^{(l)} \leq a)\, \tilde{p}(z^{(l)})/q(z^{(l)})}{\lim_{L\to\infty} \frac{1}{L} \sum_{l}^{L} \tilde{p}(z^{(l)})/q(z^{(l)})} \\
&= \lim_{L\to\infty} \frac{\sum_{l}^{L} \mathbb{I}(z^{(l)} \leq a)\, \tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_{l}^{L} \tilde{p}(z^{(l)})/q(z^{(l)})}
\end{aligned}
$$

以上より、極限においては真の分布に正確に従うサンプルであることがわかった。

もし、分布 $p(z)$ についてのモーメントが必要な場合、$E[f(z)] \simeq \sum_{l=1}^{L} w_l f(z^{(l)})$ として近似できる。これは次の式よりわかる。

$$
\begin{aligned}
E[f(z)] &= \int f(z)\, p(z)\, dz \\
&= \frac{\int f(z)\, [\tilde{p}(z)/q(z)]\, q(z)\, dz}{\int [\tilde{p}(z)/q(z)]\, q(z)\, dz} \\
&\simeq \frac{\sum_{l}^{L} f(z^{(l)})\, [\tilde{p}(z^{(l)})/q(z^{(l)})]}{\sum_{l}^{L} \tilde{p}(z^{(l)})/q(z^{(l)})} \\
&= \sum_{l}^{L} f(z^{(l)})\, w_l \qquad \left(\because\ w_l = \frac{\tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_{m} \tilde{p}(z^{(m)})/q(z^{(m)})}\right)
\end{aligned}
$$

11.1.6 サンプリングとEMアルゴリズム

ここまでは、主に期待値などのモーメントを求めるためにサンプリング法を適用してきた。サンプリング法の使用用途は、最尤解を求めるといった頻度論的パラダイムにおいても利用することができる。この節では、EMアルゴリズムの一部にサンプリング法を適用する例を学び、後半ではこれを完全ベイズアプローチに拡張した、**IPアルゴリズム**を学ぶ。

おさらいとして、EMアルゴリズムの一般的なステップは次のようなものであった。（156頁）

✤ 一般の **EM** アルゴリズム ✤

観測変数 $\mathbf{X}$ と潜在変数 $\mathbf{Z}$ の同時分布 $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ が与えられていて，パラメータ $\boldsymbol{\theta}$ で支配されているとする．アルゴリズムの目的は尤度関数 $p(\mathbf{X}|\boldsymbol{\theta})$ を $\boldsymbol{\theta}$ について最大化することである．

1. パラメータの初期値 $\boldsymbol{\theta}^{\text{old}}$ を選ぶ．
2. **E** ステップ．$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ を計算する．
3. **M** ステップ．次式で与えられる $\boldsymbol{\theta}^{\text{new}}$ を計算する．

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}). \tag{9.32}$$

ただし

Z の事後分布についての、完全データ対数尤度の期待値、

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}). \tag{9.33}$$

4. 対数尤度関数またはパラメータ値のいずれかについて，収束条件が満たされているか調べ，満たされていなければ

$$\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}} \tag{9.34}$$

を実行し，ステップ 2 に戻る．

（9.33）式は期待値の形となっており、サンプリング法により $Q(\theta, \theta^{old})$ の値を近似できる。事後分布の推定分布 $p(Z \mid X, \theta^{old})$ から生成したサンプル $\{Z^{(l)}\}$ を用いて、次のように表される。

$$Q(\theta, \theta^{old}) \simeq \frac{1}{L}\sum_{l=1}^{L} \ln p(Z^{(l)}, X \mid \theta) \qquad (11.29)$$

この近似を用いたEMアルゴリズムは、**モンテカルロEMアルゴリズム**と呼ばれる。L=1の場合は特別に**確率的EM（stochastic EM）**とも呼ばれる。

（感想）
ここでのサンプリングは、事後分布から「サンプル値を得られる」ことを前提としているので、棄却サンプリングやSIRなどを利用することができる。重点サンプリングを用いる場合は、(11.29) の代わりに重みつき和（11.22式）のような式になるはずである。

EMアルゴリズムではパラメータ θ の最尤解を得ることができる。これを完全なベイズアプローチに拡張し、θ の事後分布を得る問題を考える。この問題は **IP アルゴリズム**と呼ばれる、サンプリング法を利用した手法によって解くことができる。

EMアルゴリズムとIPアルゴリズムの大きな違いは次のようにまとめられる。概要を理解した上で、細かい手続きについては教科書を読み合わせることとする。

EM アルゴリズム	IPアルゴリズム
Eステップ： 事後分布 $P(Z \mid X, \theta^{old})$ を解析的に求める。	Iステップ： 事後分布 $P(Z \mid X)$ に従うサンプル $\{Z^{(l)}\}$ を生成する。
Mステップ： 完全データ対数尤度の期待値 $Q(\theta, \theta^{old}) = E[\ln P(X, Z \mid \theta)]$ を最大にする θ に更新する	Pステップ： θ の事後分布 $P(\theta \mid X)$ を、サンプリング法により更新する。

11.1 節の範囲はストーリー性に乏しく、なおかつ多数の手法が登場したため、内容をおさらいして知識を定着させたい。下記に 11.1.1 - 11.1.6 までの節の概要をまとめた。最後にこれを確認して終わりとする。

11.1.1 標準的な分布
よく知られた分布からサンプルを生成する手法を学んだ。
キーワード：変換法、逆関数法、Box-Muller 法

11.1.2 棄却サンプリング
複雑な分布からサンプルを生成するナイーブな手法を学んだ。
キーワード：提案分布、受理率

11.1.3 適応的サンプリング
提案分布を初期に設定する必要がなく、動的に提案分布を更新するサンプリング法を学んだ。
キーワード：なし

11.1.4 重点サンプリング
真の分布からのサンプリングを諦める代わりに、積分を直接サンプルにより近似する手法を学んだ。サンプルを棄却するのではなく、重み付けによって評価する。
キーワード：重要度重み

11.1.5 SIR
重要度重みを使って、真の分布に近似的に従うサンプルを得る手法を学んだ。棄却サンプリングと異なり、定数kの設定が必要ない。
キーワード：なし

11.1.6 サンプリングとEMアルゴリズム
サンプリング法をEMアルゴリズムに適応する手法を学んだ。
キーワード：モンテカルロEMアルゴリズム、IPアルゴリズム