

## Abstract

このドキュメントは標本数の変化に対する最尤推定値の依存性について調べた記録です。得られた結論の一つとして、標本数を増加させても、対数尤度自体は鋭い分布にはならないことがわかります (確率分布の母数を推定した分布は真の母数を中心に、標本数を増やすほど鋭い分布になりますが、標本数をいくら増やしても、もともとの分布自体 (例えば標本数を変化させても分布自体の母数である平均や分散) は変化していないことから分かります)。

構成は、初めに正規分布を例に、パラメータ (統計学では母数) である 1 次のモーメントの推定値  $\hat{\mu}$  の平均値、および分散について調べます。正規分布を例にした場合は母数の推定値の分散は標本数に対して平方根の逆数で減少することが導かれます。

次に、ギブス分布を用いて、正規分布の場合と同様の議論を繰り返し、母数の推定値の分散が標本数に対してどのように変化するかを調べます。結果は、一時のモーメントは標本数無限の極限で真のパラメータに一致することが直接確認できましたが、パラメータに関して陽に解くことができないことが原因で、2 次のモーメントについての標本数の依存性は依然として分からないです (一次のモーメントの標本数依存性を調べた方法は、汎用的ではなく、2 次のモーメントを調べる場合には用いることができませんでした)。

## 1 母数推定値の標本数依存性について

### 1.1 正規分布の場合

簡単な例として、正規分布について母数の推定値の標本数の依存性を調べます。正規分布 (1 次元) は以下のように与えられるとします。

$$x \sim \mathcal{N}(x|\mu_0, \sigma_0^{-2}) \quad (1)$$

ただし、 $\mu_0, \sigma_0^{-2}$  は確率分布の真のパラメータ (母数) であるとし、両方とも未知であるとしします。従って、標本数  $N$  だけ、標本が得られた場合の対数尤度関数は以下ようになります。

$$l(\mu, \sigma^{-2}) = \sum_{n=1}^N \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\pi\sigma^2}(x^{(n)} - \mu)^2\right)\right) \quad (2)$$

この尤度関数をパラメータの極大値について解きます。解は簡単に求めることができ以下のような表現の推定解になります。

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N x^{(n)}, \quad \hat{\sigma} = \frac{1}{N-1} \sum_{n=1}^N (x^{(n)} - \hat{\mu}) \quad (3)$$

分散の規格因子が  $N$  出ない理由は、平均の推定に自由度が 1 減っているためです。

以下では、母数の推定値である  $\hat{\mu}$  に対しての標本数依存性を調べます。真の母数の値と、推定結果  $\hat{\mu}_N$  (標本数  $N$  依存性について調べるので、以降では頭に  $N$  の添え字を書きます) との差を以下のように定義します。

$$\delta_\mu := \hat{\mu}_N - \mu_0 \quad (4)$$

標本は確率分布  $\mathcal{N}(x|\mu_0, \sigma_0^{-2})$  に従って現れるので、母数の推定値と真の母数との差は以下ようになります。

$$\langle \delta_\mu \rangle_x = \left\langle \frac{1}{N} \sum_{n=1}^N x^{(n)} - \mu_0 \right\rangle \quad (5)$$

$$= \frac{1}{N} \sum_{n=1}^N \int_{\mathbf{R}} dx^{(n)} x^{(n)} \mathcal{N}(x^{(n)}|\mu_0, \sigma_0^{-1}) - \mu_0 \quad (6)$$

$$= \mu_0 - \mu_0 = 0 \quad (7)$$

次に母数の推定値の分散についての統計値を求めます

$$\langle \delta_\mu^2 \rangle_x = \langle (\frac{1}{N} \sum_{n=1}^N x^{(n)} - \mu_0)^2 \rangle \quad (8)$$

$$= \frac{1}{NN'} \sum_{n=1}^N \sum_{n'=1}^N \langle x^{(n)} x^{(n')} \rangle - \mu_0^2 \quad (9)$$

$$= \frac{1}{NN'} \sum_{n,n'=1|n' \neq n}^N \langle x^{(n)} x^{(n')} \rangle + \frac{1}{NN'} \sum_{n=1}^N \langle x^{(n)^2} \rangle - \mu_0^2 \quad (10)$$

$$= \frac{1}{NN'} \sum_{n,n'=1|n' \neq n}^N \mu_0^2 + \frac{1}{NN'} \sum_{n=1}^N \langle x^{(n)^2} \rangle - \mu_0^2 \quad (11)$$

$$= \mu_0^2 (\frac{N(N-1)}{N^2} - 1) + \frac{1}{N^2} \sum_{n=1}^N \int_{\mathbf{R}} dx^{(n)} x^{(n)^2} \mathcal{N}(x^{(n)} | \mu_0, \sigma_0^{-1}) \quad (12)$$

$$= -\frac{\mu_0^2}{N} + \frac{\sigma_0^2 + \mu_0^2}{N} = \frac{\sigma_0^2}{N} \quad (13)$$

よって、母数の推定値の標本数依存性は

$$\langle \delta_\mu \rangle_x = 0, \langle \delta_\mu^2 \rangle_x = \frac{\sigma_0^2}{N} \quad (14)$$

従って、真の確率分布が  $\mathcal{N}(x | \mu_0, \sigma_0^{-2})$  である時、その母数の推定値  $\hat{\mu}$  の分散は  $\frac{\sigma_0^2}{N}$  となる。

## 1.2 Gibbs 分布の場合

この章では、正規分布において、母数の推定値の標本数依存性を調べた議論を Gibbs 分布に対して用います。一般的な Gibbs 分布ではなく Ising モデルをエネルギー関数とする場合で調べます

正規分布の場合と同様にして、注目すべき量は  $\langle \delta_\mu \rangle, \langle \delta_\mu^2 \rangle$  です。

確率分布の形は以下のように与えられているとします。

$$p(x|J_0) = \frac{e^{-E(x|J_0)}}{Z(J_0)}, \quad E(x|J_0) = -J_0 \sum_{i=1}^d x_i x_{i+1}, \quad Z(J_0) = \sum_{x:all} e^{-E(x|J_0)} \quad (15)$$

変数の次元は  $d$  次元として、パラメータ数は 1 つであるとします。これより、標本数が  $N$  とき、対数尤度関数は以下のような式になります。

$$l(J) = - \sum_{n=1}^N E(x^{(n)}) - N \log(Z(J)) \quad (16)$$

パラメータの推定値は尤度関数の極大値を与えるパラメータであるため、パラメータ微分した式について  $= 0$  となるパラメータが最尤推定解です。

$$\frac{\partial l(J)}{\partial J} = - \sum_{n=1}^N (\sum_{i=1}^d x_i^{(n)} x_{i+1}^{(n)}) + N \frac{\sum_{x:all} (\sum_{i=1}^d x_i x_{i+1}) e^{-J(\sum_{i=1}^d x_i x_{i+1})}}{\sum_{x:all} e^{-J(\sum_{i=1}^d x_i x_{i+1})}} \quad (17)$$

式を見やすくするため、一部をまとめます。

$$g(x) := \sum_{i=1}^d x_i x_{i+1}, \quad x^N := \{x^{(n)}\}_{n=1}^N \quad (18)$$

$$\frac{1}{N} \sum_{n=1}^N g(x^{(n)}) = \sum_{x:all} g(x) e^{-\hat{J}(x^N)g(x)} Z(\hat{J}(x^N)) \quad (19)$$

上式の両辺を、真の確率分布で平均を取った量を考えます。すなわち

$$\left\langle \frac{1}{N} \sum_{n=1}^N g(x^{(n)}) \right\rangle_{x^N} = \left\langle \sum_{x:all} g(x) e^{-\hat{J}(x^N)g(x)} Z(\hat{J}(x^N)) \right\rangle_{x^N} \quad (20)$$

$$l.h.s = \left\langle \frac{1}{N} \sum_{n=1}^N g(x^{(n)}) \right\rangle_{x^N} \quad (21)$$

$$= \frac{1}{N} \sum_{n=1}^N \langle g(x^{(n)}) \rangle_{x^N} \quad (22)$$

$$= \langle g(x^{(n)}) \rangle_{x^N} \quad (23)$$

$$= \sum_{x:all} g(x) e^{-Jg(x)} Z(J_0) \quad (24)$$

$$r.h.s = \left\langle \sum_{x:all} g(x) e^{-\hat{J}(x^N)g(x)} Z(\hat{J}(x^N)) \right\rangle_{x^N} \quad (25)$$

$$= \sum_{x:all} g(x) \langle e^{-\hat{J}(x^N)g(x)} Z(\hat{J}(x^N)) \rangle_{x^N} \quad (26)$$

$$(27)$$

従って、任意の  $x \in \{-1, +1\}^d$  に対して

$$\sum_{x:all} g(x) \{ \langle e^{-\hat{J}(x^N)g(x)} Z(\hat{J}(x^N)) \rangle_{x^N} - e^{-Jg(x)} Z(J_0) \} = 0 \quad (28)$$

$$(29)$$

であるので、これより

$$\langle \hat{J}(x^N) \rangle_{x^N} = J_0 \quad (30)$$

となります。

先ほどの正規分布の時と同様に

$$\langle \hat{J}(x^N) - J_0 \rangle_{x^N} = \langle \delta_J \rangle_{x^N} = 0 \quad (31)$$

$$\langle \delta_J^2 \rangle$$

### 1.3 熱力学的極限の結果の援用

parameter  $J$  の対数尤度解を陽に解いて、確率変数 (スピン変数) として書きたいのですが、前の章の手法では直接解くことができませんでした

しかしながら、一次元で結合定数が一様なモデルに対しては転送行列法で分配関数が簡単に解け、分配関数のパラメータ依存性が分かります。

以下の議論では、一次元、一様な結合定数の場合で最近接相互作用のみのモデル (前章と同じ設定) に対してパラメータを解いて標本数の依存性を調べます。Hamiltonian は以下のように定義します。

$$\mathcal{H}(x|J) = -J \sum_{i=1}^d x_i x_{i+1} \quad (32)$$

分配関数は以下のように与えられます。

$$Z(J)_d = \lambda_1^d + \lambda_2^d, \quad 2\lambda_1 := \sinh(J), \quad 2\lambda_2 = \cosh(J) \quad (33)$$

天下りのですが、以下の結果を援用します。

$$\langle x_k x_l \rangle_d = \frac{1}{Z} \sum_{x:all} x_k x_l e^{\beta \mathcal{H}(J)} \quad (34)$$

$$\langle x_k x_l \rangle_d = \frac{\lambda_1^{d-(l-k)} \lambda_2^{(l-1)} + \lambda_1^{(l-1)} \lambda_2^{d-(l-k)}}{\lambda_1^d + \lambda_2^d} \quad (35)$$

ただし、 $l \langle k$  としています。システムのサイズ  $d$  の極限をとります。

$$\langle x_k x_l \rangle = \lim_{d \rightarrow \infty} \langle x_k x_l \rangle_d = \left( \frac{\lambda_2}{\lambda_1} \right)^{(l-k)} \quad (36)$$

上式と、前章で扱っていた尤度方程式とを組み合わせます。

$$\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^d x_i^{(n)} x_{i+1}^{(n)} = d \frac{1}{N} \sum_{n=1}^N x_i^{(n)} x_{i+1}^{(n)} \quad (37)$$

$$= d \lim_{d \rightarrow \infty} \langle x_i x_{i+1} \rangle_{d, \hat{J}} = d \frac{\cosh(\hat{J})}{\sinh(\hat{J})} \quad (38)$$

$\hat{J}$  についてまとめます

$$\frac{\cosh(\hat{J})}{\sinh(\hat{J})} = \frac{e^{\hat{J}} + e^{-\hat{J}}}{e^{\hat{J}} - e^{-\hat{J}}} \quad (39)$$

$$= \frac{e^{\hat{J}} - e^{-\hat{J}} + 2e^{-\hat{J}}}{e^{\hat{J}} - e^{-\hat{J}}} \quad (40)$$

$$= 1 + \frac{2e^{-\hat{J}}}{e^{\hat{J}} - e^{-\hat{J}}} \quad (41)$$

$$= 1 + \frac{2}{e^{2\hat{J}} - 1} \quad (42)$$

$$(43)$$

標本平均の結果を使います。

$$e^{2\hat{J}} - 1 = \frac{2}{\frac{1}{N} \sum_{n=1}^N x_i^{(n)} x_{i+1}^{(n)} - 1} \quad (44)$$

$$\hat{J} = \frac{1}{2} \log \left( \frac{2N}{\sum_{n=1}^N x_i^{(n)} x_{i+1}^{(n)} - 1} \right) \quad (45)$$

上で得られた結果を、真のパラメータによる平均  $\langle \rangle_{J_0}$  でも成り立つため、パラメータの平均の推定誤差の統計量はゼロになります。