「差分型」の場合のθの間隔の議論の補足（暫定α版）

統計数理研究所　伊庭幸人

平成 17 年 10 月 31 日

「マルコフ連鎖モンテカルロ法の基礎」（統計科学のフロンティア１２巻）の 4.7 節「積分の道と交換の道（上級篇）」p.85 では

> パラメータθの間隔をどのくらいあけたらよいかという問題についても，差分型の積分法とレプリカ交換モンテカルロ法に並行関係がありそうなことは，いままでの議論から推察できる．

と書きましたが，4.4 節はともかく 4.5 節の議論のほうからこれを導くのはかなりギャップがあるので，もう少し詳しい説明を書いてみました．

1　おおまかな議論

以下から出発する．

$$\log \frac{P(\boldsymbol{x}|\theta_{i+1})}{P(\boldsymbol{x}|\theta_i)} \simeq \left. \frac{d}{d\theta} \log P(\boldsymbol{x}|\theta) \right|_{\theta=\theta_i} \triangle\theta + \frac{1}{2} \left. \frac{d^2}{d^2\theta} \log P(\boldsymbol{x}|\theta) \right|_{\theta=\theta_i} (\triangle\theta)^2$$

ここで $\triangle\theta = \theta_{i+1} - \theta_i$ である．まず，定義

$$P(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta_i)}{Z(\theta_i)} \tag{1}$$

を用いると，

$$\log\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)} \simeq \quad \frac{d}{d\theta}\log P(\boldsymbol{x}|\theta)\Big|_{\theta=\theta_i}\triangle\theta \quad +\frac{1}{2}\frac{d^2}{d^2\theta}\log P(\boldsymbol{x}|\theta)\Big|_{\theta=\theta_i}(\triangle\theta)^2 \quad -\log\frac{Z(\theta_i)}{Z(\theta_{i+1})} \tag{2}$$

と書ける．ここで，

$$\mathbb{E}_{\theta_i}\left(\frac{d}{d\theta}\log P(\boldsymbol{x}|\theta)\Big|_{\theta=\theta_i}\right)=0 \tag{3}$$

が容易に示せるので，

$$\mathbb{E}_{\theta_i}\left(\log\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)}\right) \simeq \quad \frac{1}{2}\ \mathbb{E}_{\theta_i}\left(\frac{d^2}{d^2\theta}\log P(\boldsymbol{x}|\theta)\Big|_{\theta=\theta_i}\right)(\triangle\theta)^2 \quad -\log\frac{Z(\theta_i)}{Z(\theta_{i+1})}$$

となる．これは $I(\theta)$ をフィッシャー情報量

$$I(\theta)= \quad -\mathbb{E}_{\theta_i}\left(\frac{d^2}{d^2\theta}\log P(\boldsymbol{x}|\theta)\Big|_{\theta=\theta_i}\right) \tag{4}$$

とすると

$$\mathbb{E}_{\theta_i}\left(\log\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)}\right) \simeq \quad -\frac{1}{2}I(\theta)(\triangle\theta)^2 \quad -\log\frac{Z(\theta_i)}{Z(\theta_{i+1})} \tag{5}$$

とも書ける．$\log(f(\boldsymbol{x}|\theta_{i+1})/f(\boldsymbol{x}|\theta_i))$ が近似的に正規分布に従うとすると，式 (5)) の右辺を指数の肩に載せたものが比 $f(\boldsymbol{x}|\theta_{i+1})/f(\boldsymbol{x}|\theta_i)$ の最頻値（分布のピークを与える $\boldsymbol{x}$ の値）の近似値となるわけである．

一方，式 (1) より比 $f(\boldsymbol{x}|\theta_{i+1})/f(\boldsymbol{x}|\theta_i)$ の期待値の対数は

$$\log\mathbb{E}_{\theta_i}\left(\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)}\right)=\sum_{\boldsymbol{x}}P(\boldsymbol{x}|\theta_i)\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)}=-\log\frac{Z(\theta_i)}{Z(\theta_{i+1})} \tag{6}$$

となる．式 (6) の右辺を指数の肩にあげたものが比 $f(\boldsymbol{x}|\theta_{i+1})/f(\boldsymbol{x}|\theta_i)$ の期待値である．

「分布の形があまりゆがまない」条件を「最頻値と期待値の比が 1 とあまり違わない」と考えると，式 (5) と式 (6) を比較して，

$\frac{1}{2}I(\theta)(\triangle\theta)^2$ があまり大きくならない

となるが，これは交換法で交換のおこる条件と（係数を除いて）同じである．

以上は普通の計算であるが，計算上の注意点は f が規格化されていない量であることを忘れないことである．

2　対数正規分布

「分布のゆがみ」などといわずに，直接に比 $f(\boldsymbol{x}|\theta_{i+1})/f(\boldsymbol{x}|\theta_i)$ の分散を計算してみたいという人もいると思う．その場合に，ポイントになるのが，対数正規分布の性質である．

いま z が正規分布 $N(m,\sigma^2)$ に従う確率変数だとする．このとき，その定数倍 αz の期待値 αm，分散 $\alpha^2\sigma$ で，標準偏差と期待値の比（変動係数）は σ/m となり α によらない．これらはもちろん当たり前である．

ところが，$\exp(\alpha z)$ となると，まったく違うことになる．$\exp(\alpha z)$ の変動係数は α に依存し，α が大きくなると，その指数関数で発散することが簡単にわかる．具体的に計算すると，$\exp(\alpha z)$ の期待値を $m_{\exp}$，分散を $\sigma^2_{\exp}$ として，

$$m_{\exp} = \exp\left(\alpha m + \alpha^2\sigma^2/2\right) \tag{7}$$

$$\sigma^2_{\exp} = \exp(2\alpha m + \alpha^2\sigma^2) \times \left(\exp(\alpha^2\sigma^2) - 1\right)$$

なので，変動係数は

$$\sqrt{\frac{\sigma^2_{\exp}}{m^2_{\exp}}} = \sqrt{\exp(\alpha^2\sigma^2) - 1}$$

となり，$\alpha^2\sigma^2$ の指数関数で増加する．これが「差分型」の場合に刻みを大きくするとある時点で計算精度がぼろぼろになる理由と考えられる．

式 (2) で，分散の計算には定数項は関係がなく，$\triangle\theta$ の 2 次の項の寄与はつぎの次数になるので，最初の項だけをとると，

$$\log\frac{f(\boldsymbol{x}|\theta_{i+1})}{f(\boldsymbol{x}|\theta_i)} \simeq \quad \frac{d}{d\theta}\log P(\boldsymbol{x}|\theta)\bigg|_{\theta=\theta_i} \times \triangle\theta \quad + 定数$$

なので，$\alpha^2\sigma^2$ が大きくならないというのに相当する条件は

$$\left.\frac{d}{d\theta}\log P(\boldsymbol{x}|\theta)\right|_{\theta=\theta_i} \tag{8}$$

の分散に $(\triangle\theta)^2$ を乗じたものがあまり大きくならないということである．

式 (3) とフィッシャー情報量 (4) が，

$$I(\theta) = \mathbb{E}_{\theta_i}\left(\left.\left(\frac{d}{d\theta}\log P(\boldsymbol{x}|\theta)\right)^2\right|_{\theta=\theta_i}\right)$$

とも書けることを用いると，式 (8) の分散は $I(\theta_i)$ に等しく，

$I(\theta_i)(\triangle\theta)^2$ があまり大きくならない

という条件が（係数を除いて）再び得られる．