ポアソン系列平滑化の計算法について

柏 木 宣 久

ポアソン系列の平滑化を次の状態空間表現により考える.

$$y_i \sim Po(\lambda_i) \qquad i=1,\dots,T$$
$$\log\lambda_i=\log\lambda_{i-1}+\varepsilon_i \qquad \varepsilon_i\sim \text{i.i.d. } N(0,\sigma^2) \qquad i=2,\dots,T$$

ここに, y_i は観測量, λ_i はポアソン分布の平均, そして T はデータの長さを表わす. この時, 観測量の密度関数は

$$p(y\mid\lambda)=\prod_{i=1}^{T} e^{-\lambda_i}\frac{\lambda_i^{y_i}}{y_i\,!}$$

と書かれ, $\lambda_2,\dots,\lambda_T$ の密度関数は

$$p(\lambda_2,\dots,\lambda_T\mid\lambda_1,\sigma^2)=\prod_{i=2}^{T}\frac{1}{\sqrt{2\pi}\,\sigma\lambda_i}e^{-\frac{1}{2\sigma^2}\left(\log\frac{\lambda_i}{\lambda_{i-1}}\right)^2}$$

と書かれる. ここに, $y=(y_1,\dots,y_T)$, $\lambda=(\lambda_1,\dots,\lambda_T)$ である. 従って, λ の推定値は, 例えば, ある与えられた $\hat{\sigma}^2$ のもとで, λ の事後尤度

$$p(\lambda\mid\hat{\sigma}^2)\propto p(y\mid\lambda)p(\lambda_2,\dots,\lambda_T\mid\lambda_1,\hat{\sigma}^2)p(\lambda_1)$$

による平均として求まる. 一方, σ^2 の推定値は周辺尤度

$$l(\sigma^2)=\int p(y\mid\lambda)p(\lambda_2,\dots,\lambda_T\mid\lambda_1,\sigma^2)p(\lambda_1)d\lambda$$

を最大にするように決めてやればよい.

問題点が2つある. ひとつは $p(\lambda_1)$ のモデリングであり, もうひとつは計算時間の短縮である. データの長さが長い場合, $p(\lambda_1)$ の選択が結果に与える影響はそれほど大きくないと期待される. しかし, データの長さが短い場合, $p(\lambda_1)$ の影響は本質的である. また, 状態空間表現の数値計算は比較的高速に行なえるが, シミュレーション等を試みようとする場合には, さらに高速化を図る必要がある.

先ず第一の問題点であるが, $p(\lambda_1)$ の同定法として, $l(\sigma^2)$ を $p(\lambda_1)$ の汎関数とみなし, これを最大にする $p(\lambda_1)$ を選択する方法が考えられる. そうした方法によった場合, $p(\lambda_1)$ として $\delta(\lambda_1-\hat{\lambda}_1)$ が最も適切であることが分かった. ここに, $\delta(\cdot)$ はディラック分布を表わす.

次に第二の問題点であるが, 計算時間の短縮を実現するには, 状態空間表現の数値計算で最も計算量の多い数値積分を高速化してやればよい. 今, 次のような台形公式による数値積分を考える.

$$\int_a^b f(x)g(x)dx\approx\frac{1}{2}\{f(a)g(a)+f(b)g(b)\}+\sum_i f(x_i)g(x_i)$$

式の形から, $f(x_i)g(x_i)$ の値がほとんど0と見なせる場合については, 右辺の和をとる操作を省略しても, その影響はごく僅かであると期待される. そこで,「省略」という操作によって計算の高速化を試みてみた. ポアソン系列の平滑化の場合, $f(x_i)$, $g(x_i)$ に相当する関数が極端に大きい値をとるのは極めて希であるので, $f(x_i)$ または $g(x_i)$ が密度に換算して 10^{-12} より小さい場合に「省略」するようにした. その結果, 各種推定値にほとんど影響を与えずに, 最大で1/100, 平均的には1/10程度に計算時間の短縮が可能であることを確認した.