第Ⅳ部門　　確率密度関数がべき級数のときの最尤法の適用性に関する一考察 −二次方程式の場合−

京都大学大学院　学生員　○益井　大樹
京都大学大学院教授　フェロー　細田　尚

１．はじめに

関数形が既知である確率密度関数を，観測値から推定する方法の 1 つとして最尤法が挙げられる．これまで，正規分布や Weibull 分布の確率密度関数への適用はなされているが，関数形がべき級数のときの検討はあまり行われていないように考えられる．そこで本研究では，まず二次関数の場合に最尤法を適用して関数のパラメータを計算する手法について検討する．得られた知見を下に，今後，確率密度関数に極大値が二つ以上存在する場合へのべき級数の適用性を検討していきたいと考えている．

２．二次方程式の確率密度関数

確率密度関数 $f(y)$ を以下の二次方程式で表す．

$$f(y) = a_0 + a_1 y + a_2 y^2 \tag{1}$$

ここで，$f(y) = 0$ の解をそれぞれ y_l，y_r $(y_l < y_r)$ とすると以下で表される．

$$f(y_l) = 0 \tag{2}$$

$$f(y_r) = 0 \tag{3}$$

$$\int_{y_l}^{y_r} \left(a_0 + a_1 y + a_2 y^2\right) dy = 1 \tag{4}$$

これらより，a_0，a_1，a_2 は以下の式で表される．

$$a_0 = -\frac{6 y_r y_l}{(y_r - y_l)^3} \tag{5}$$

$$a_1 = \frac{6(y_r + y_l)}{(y_r - y_l)^3} \tag{6}$$

$$a_2 = -\frac{6}{(y_r - y_l)^3} \tag{7}$$

３．観測値のデータが 2 個の場合

簡単のため，データが 2 個の場合を考える．2 個の観測値を y_1，y_2 $(y_1 \leq y_2)$ とすると，尤度関数 $L(y_l, y_r)$ は次式となる．

$$L(y_l, y_r) = \left(a_0 + a_1 y_1 + a_2 y_1{}^2\right)\left(a_0 + a_1 y_2 + a_2 y_2{}^2\right) \tag{8}$$

$\dfrac{\partial L(y_l, y_r)}{\partial y_l} = 0$，$\dfrac{\partial L(y_l, y_r)}{\partial y_r} = 0$ を解くと，$y_l < y_1 \leq y_2 < y_r$ であることから y_l，y_r は次式で与えられる．

$$(y_l, y_r) = \left(\frac{(1+\sqrt{3})y_1 + (1-\sqrt{3})y_2}{2}, \frac{(1-\sqrt{3})y_1 + (1+\sqrt{3})y_2}{2}\right) \tag{9}$$

Hiroki MASUI and Takashi HOSODA
masui.hiroki.86u@st.kyoto-u.ac.jp

このとき，$\dfrac{y_l + y_r}{2} = \dfrac{y_1 + y_2}{2}$となり，観測値の平均と求めた確率密度関数の平均値が一致することが分かる．しかし，分散に関しては，観測値が$\dfrac{(y_1 - y_2)^2}{4}$であるのに対し，求めた確率密度関数は$\dfrac{(y_l - y_r)^2}{4} = \dfrac{3(y_1 - y_2)^2}{4}$となり，両者が一致しないことも示すことができる．

４．観測値のデータが 3 個の場合

3 個の観測値を y_1，y_2，y_3 $(y_1 \le y_2 \le y_3)$とすると，尤度関数$L(y_l, y_r)$は以下となる．

$$L(y_l, y_r) = \left(a_0 + a_1 y_1 + a_2 {y_1}^2\right)\left(a_0 + a_1 y_2 + a_2 {y_2}^2\right)\left(a_0 + a_1 y_3 + a_2 {y_3}^2\right) \tag{10}$$

観測値 y_1，y_2，y_3を用いて$\dfrac{\partial L(y_l, y_r)}{\partial y_l} = 0$，$\dfrac{\partial L(y_l, y_r)}{\partial y_r} = 0$を解くと解の表示式を得ることができるが，複雑なので省略する．導かれた式に観測値を代入して解を求めてみると以下のようになる．

$y_1 = 1$，$y_2 = 2$，$y_3 = 4$を与えると，$y_l < y_1 \le y_2 \le y_3 < y_r$であることから，以下の結果を得た．

$$(y_l, y_r) = (0.131, 4.760) \tag{11}$$

このとき，観測値の平均値2.333と求めた確率密度関数の平均値2.446は近い値となるが一致しない．また，観測値の分散1.556は求めた確率密度関数の分散5.357と一致しない．すなわち分散だけでなく平均値についても最尤法とモーメント法は一致しないことが分かる．

５．観測値のデータが 4 個の場合

前章と同様の手順で, 4 個の観測値に具体的な値$y_1 = 1$，$y_2 = 4$，$y_3 = 5$，$y_4 = 7$を与えて解を求めると，$y_l < y_1 \le y_2 \le y_3 \le y_4 < y_r$であることから，

$$(y_l, y_r) = (0.199, 8.353) \tag{12}$$

このことから，観測値の平均値4.25は最尤法の平均値4.276と一致しないものの近い値をとることが分かる．また，観測値の分散4.688は最尤法の分散16.622と一致しない．

６．観測値のデータが n 個の場合

観測値の個数が 2 個，3 個，4 個の場合に，$\dfrac{\partial L(y_l, y_r)}{\partial y_l} = 0$，$\dfrac{\partial L(y_l, y_r)}{\partial y_r} = 0$は規則性を持ち，これらを解くことで，煩雑にはなるもののy_l，y_rの一般解を求めることができることを示した．また，観測値の平均値と最尤法で計算した平均値はデータの個数の増加とともに近づく可能性が示された．観測値が n 個の場合も同様の手順で一般解を求めることができることを確認している．

７．おわりに

本研究では，確率密度関数がべき級数で表される場合の最も単純な場合として二次方程式を取り上げ，最尤法の適用性について考察した．その結果，観測値の個数の増加とともに平均値に関してはモーメント法と最尤法の結果が近づくことを示唆した．本研究で得られた知見を下に，4 乗の項を考慮した場合，すなわち確率密度関数に極大値が二つ存在する可能性がある場合についての解析手順や解の一意性について検討したいと考えている．