g(一様分布) からの三次関数 g の最尤推定

hsjoihs

1. 概要 1

区間 [0,1] 上の連続一様分布に従う確率変数 K を導入し、三次関数 $g(x) = a + bx + cx^2 + dx^3$ を用いて $X = g(K)$ と書ける確率変数を考える。ただし、g は単調増加するものとするので、$d > 0,\ c^2 - 3bd < 0$ を前提とする。このとき、パラメータの最尤推定をしてみる。

2. 3 パラメータ

2.1. 概要

区間 $[0,1]$ 上の連続一様分布に従う確率変数 K を導入し $X = A(K-B)^3 + C(K-B)$ と書ける確率変数を考える。ただし、$A > 0, C > 0, 0 < B < 1$ とする。B は X が 0 以下になる確率を表す。このとき、パラメータの最尤推定をしてみる。

2.2. 累積分布関数と確率密度関数

とりあえず、累積分布関数と確率密度関数を求める。$G(x) = A(x-B)^3 + C(x-B)(0 < x < 1)$ の逆関数が累積分布関数であるので、$x = A(y-B)^3 + C(y-B)$ である。ここで $k = \sqrt{\dfrac{3A}{4C}}$ とおくと

$$\frac{3k}{C}x = 4k^3(y-B)^3 + 3k(y-B)$$

ここで $k(y-B) = \sinh\lambda$ とおくと $\dfrac{3k}{C}x = 4\sinh^3\lambda + 3\sinh\lambda = \sinh 3\lambda$ であり、これは

$\dfrac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3} = \lambda$ と書ける。

ゆえに、累積分布関数は $y = F(x) = B + \dfrac{1}{k}\sinh\dfrac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3} (G(0) < x < G(1))$ であり、確率密度関数はこれを微分して

$$f(x) = \frac{1}{k}\frac{d}{dx}\sinh\frac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3} = \frac{1}{k}\cosh\frac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3}\cdot\frac{d}{dx}\frac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3}$$

$$= \frac{1}{3k}\cosh\frac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3}\cdot\frac{1}{\sqrt{1+\left(\dfrac{3k}{C}x\right)^2}}\cdot\frac{3k}{C}$$

$$= \frac{1}{C}\cosh\frac{\operatorname{arsinh}\left(\dfrac{3k}{C}x\right)}{3}\cdot\frac{1}{\sqrt{1+\left(\dfrac{3k}{C}x\right)^2}}$$

つまり、$f(x)$ の式の形の中に露わには B が現れない。（B は定義域を定める役割のみを果たす。）

以後、$D = \dfrac{3k}{C}$ なるパラメータを導入することとする。

2.3. 最尤推定

3 つのパラメータ B, C, D を推定することを考える。$X_1, X_2 ... X_n$ の n 個のデータがあったとき、このデータが生成される確率密度の ln は、データが全て $G(0) < x_i < G(1)$ を満たす場合は

$$J(C,D) = \sum_{i=1}^{n}\ln\left(\frac{1}{C}\cosh\frac{\operatorname{arsinh}(Dx_i)}{3}\cdot\frac{1}{\sqrt{1+(Dx_i)^2}}\right)$$

$$= -\sum_{i=1}^{n} \ln C + \sum_{i=1}^{n} \ln\left(\cosh\frac{\operatorname{arsinh}\left(Dx_i\right)}{3} \cdot \frac{1}{\sqrt{1+\left(Dx_i\right)^2}}\right)$$

$$\frac{\partial J}{\partial C} = -\sum_{i=1}^{n} \frac{1}{C}$$