標本最大値のカーネル型分布推定について

鳥取大学大学院工学研究科 社会基盤工学専攻 森山 卓

1. 序

本講演では独立同分布 F に従う $\{X_1, X_2, \dots, X_n\}$ に基づく、確率変数列 $\{X_{n+1}, \dots, X_{n+m}\}$ の最大値 $Z_{(m)} := \max\{X_{n+1}, \dots, X_{n+m}\}$ の確率分布推定について考える。理論的な興味になるのは $m \rightarrow \infty$ の場合であり、この意味での漸近理論として極値統計学が発展してきた。極値統計学における Fisher-Tippett-Gnedenko theorem に基づくと、典型的には一般化極値分布が標本最大値の近似（漸近）分布として当てはめられる。ここで定理の主張は次の通りである。

$$\lim_{m \rightarrow \infty} P[a_m^{-1}(Z_{(m)} - b_m) \leq x] = G(x)$$

（ただし、累積分布関数 G は非退化）を満たす実数列の組 (a_m, b_m) ($a_m > 0$) が存在するならば、次が成り立つ：

$$G(x) = \exp(-(1 + \gamma x)^{-1/\gamma}), \quad 1 + \gamma x > 0, \quad \gamma \in \mathbb{R}. \quad (1)$$

一方で漸近 $m \rightarrow \infty$ の成立が認められない場合、本近似の精度が課題となる。そこで大標本 $n \rightarrow \infty$ を条件として課すと共に、 m を n の多項式オーダーで発散する場合における確率分布の推定精度について議論する。本講演では、Hall (1982)^[1] に基づく次の仮定を置く：

$$\exists \alpha > 0, \exists \beta > \frac{1}{2}, \exists C > 0, \exists B > 0, \quad \text{s.t.} \quad 1 - F(x) = Cx^{-\alpha}\{1 + Bx^{-\beta} + o(x^{-\beta})\} \quad \text{as } x \rightarrow \infty.$$

2. 標本最大値のカーネル型分布推定について

確率変数列が独立同分布 F に従うことから、ノンパラメトリック推定法が（少なくとも m が固定の場合には）標本最大値の確率分布推定の自然な候補に思える。 $Z_{(m)}$ の確率分布が F^m であることから、本講演ではカーネル型分布推定量 $\hat{F}$ に基づく $\hat{F}^m$ について検討する。ここで、

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

であり、 K は連続な累積分布関数の性質を満たすものとする。 h は一般に $h \rightarrow 0$ 及び $nh \rightarrow \infty$ ($n \rightarrow \infty$) を満たすものが用いられる。実際に $\hat{F}^m$ は一定の条件のもとで、ある意味での一致性を持つことが示される。本講演では漸近収束レートを紹介し、一般化極値分布のフィッティングとの比較についても議論を行う。

3. 参考文献

[1] Hall, P. (1982). On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. Ser. B 44, 37-42.