丸め誤差の極限分布

志村 隆彰 (統計数理研究所（立川市緑町）)

2011 年 12 月 22 日（木）

1 問題と準備

観測データに混入する様々な種類の誤差のうち、丸め誤差は観測精度の限界から必然的に生ずる。本講演では観測値が大きいときの丸め誤差の性質を論ずる。

ある連続型の実数値確率変数に対し、その整数部分を得られる観測値、小数部分を未知の丸め誤差とみなし、実現値の整数部分のみがわかり、小数部分はわからない状況を考える。ともに確率変数である整数部分と小数部分は互いに無関係に定めることが出来るし、指数分布に従う確率変数においては互いに独立である。しかし、多くの分布では自然な従属性があり、それは元の分布の裾の挙動と密接に関係している。主な関心は整数部分が大きくなるときの小数部分の分布の振る舞いにある。

記法、分布族の定義等を準備する。上限点 $x_F = \sup\{x : F(x) < 1\}$ が無限大の分布 F のみ扱い、その裾 $F((x, \infty))$ を $\bar{F}(x)$ と書く。裾の性質（重い軽い）により分布族を定義していく。

最初に、$\mathcal{L}(\infty)$ で任意の $k > 0$ に対して、$\lim_{x\to\infty} \bar{F}(x+k)/\bar{F}(x) = 0$ が成り立つ分布の全体を表す。 正規分布、レイリー分布、パラメーターが 1 より大きいワイブル分布などがこれに属する。次に、ある指数 $\gamma > 0$ があって、任意の $k \in \mathbf{R}$ に対して、$\lim_{x\to\infty} \bar{F}(x+k)/\bar{F}(x) = e^{-\gamma k}$ を満たすとき分布 F は指数的裾を持つといい、その全体からなる分布の族を $\mathcal{L}(\gamma)(\gamma > 0)$ と書く。指数分布をはじめ、ガンマ分布、逆正規分布、ロジスティック分布、双曲分布などがその例である。最後に、分布 F が長い裾 (long-tailed) を持つとは、任意の $k \in \mathbf{R}$（実は ある $k \neq 1$ で十分）に対して、$\lim_{x\to\infty} \bar{F}(x+k)/\bar{F}(x) = 1$ を満たすときをいい、その全体を $\mathcal{L}$ と書く。その部分族で、ハザード関数 $h(t) = p(t)/\bar{F}(t)$ ($p(t)$ は F の密度関数（分布の絶対連続性を仮定）) が 任意の $k \in \mathbf{R}$ に対して、$\lim_{t\to\infty} h(t+k)/h(t) = 1$ を満たす分布全体を $\mathcal{L}_0$ で表す。対数正規分布、コーシー分布、パレート分布、F 分布などが例である（名のつくような $\mathcal{L}$ の分布はほぼ $\mathcal{L}_0$ であり、劣指数的な分布でもある）。[1]

普通に見られる分布のハザード関数は初期不良を考慮するとバスタブ型になったりもするが、この場合も含め、多くは遠方で単調である。更に、$\lim_{t\to\infty} h(t) = \infty, \gamma, 0$ ならば、それぞれ $F \in \mathcal{L}(\infty), \mathcal{L}(\gamma), \mathcal{L}$ であるから、ほとんどの連続分布はこれらのいずれかに属すると言っても過言ではないであろう。

[1] $\mathcal{L}, \mathcal{L}(\gamma)$ は普通に用いられる記号だが、$\mathcal{L}(\infty)$ と $\mathcal{L}_0$ はそうではない。

2 結果

平均 γ^{-1} の指数分布の小数部分の分布を $Fe(\gamma)$ で、$[0,1]$ 上の一様分布を $U(0,1)$ でそれぞれ表わす。確率変数 X の整数部分を $[X]$、小数部分を $\{X\}$ としたとき、$F_n(x) = P(\{X\} \leq x|[X] = n)$ の $n \to \infty$ のときの挙動に関して以下のことが言える。

$$F_n(x) = \begin{cases} 0, & x < 0, \\ F[n, n+x]/F[n, n+1), & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

定理 1 *(i)* $F \in \mathcal{L}(\infty)$ ならば、 $F_n \to \delta_0 (n \to \infty)$ となる。

(ii) $F \in \mathcal{L}(\gamma)(\gamma > 0)$ ならば、 $F_n \to Fe(\gamma)(n \to \infty)$ となる。

(iii) $F \in \mathcal{L}_0$ ならば、$F_n \to U(0,1)(n \to \infty)$ となる。

(iii) のみ条件がきついのは、一般の $\mathcal{L}$ では次が成り立つからである.

定理 2 任意の $F \in \mathcal{L}$ と任意の $[0,1]$ 上の分布 G に対して, $F_G \in \mathcal{L}$ を $\lim_{x\to\infty} \bar{F}_G(x)/\bar{F}(x) = 1$ かつ F_G の小数分布が G となるように取ることが出来る。

一定の条件下で逆も成り立つ。

定理 3 F のハザード関数 $h(t)$ が $t \to \infty$ のとき収束するか、無限へ発散すると仮定すれば次が成り立つ。

(i) $F \in \mathcal{L}(\infty)$ と $F_n \to \delta_0$ は同値である。

(ii) $F \in \mathcal{L}(\gamma)$ と $F_n \to Fe(\gamma)$ は同値である。

(iii) $F \in \mathcal{L}$ と $F_n \to U(0,1)$ は同値である.

講演ではより詳しい結果や研究動機についても述べる予定である。

参考文献

Shimura, T.(2011). Limit distribution of roundoff error (submitted).

志村 隆彰 (2011). 丸め誤差の極限分布について, 統計数理研究所共同研究リポート極値理論の工学への応用 (8), **261**, 85-98.