指数分布確率変数列の最大値

December 3, 2016

Akio Arimoto

互いに独立な確率変数列 $X_1, X_2, \cdots$ で各々の分布が指数分布、すなわち

$$P\left(X_k \le x\right) = \begin{cases} 1 - e^{-x} & (x \ge 0) \\ 0 & (x < 0) \end{cases} \qquad k = 1, 2, \cdots$$

とする。最大値をあたえる確率変数 $M_n = \max\limits_{1 \le k \le n} \left\{X_k\right\}$ とおくと、M_n の分布関数は容易にわかるように

$$P\left(M_n \le x\right) = \begin{cases} \left(1 - e^{-x}\right)^n & (x \ge 0) \\ 0 & (x < 0) \end{cases}$$

となる。$M_n - \log n$ の分布は $n \to \infty$ のとき、Gumbell 分布に行くことがわかる。

実際、 $P\left(M_n - \log n \le x\right) = \begin{cases} \left(1 - \dfrac{e^{-x}}{n}\right)^n & (x \ge -\log n) \\ 0 & (x < -\log n) \end{cases}$

で $n \to \infty$ のとき、各点 x で Gunbell 分布の分布関数 $e^{-e^{-x}}$ に収束することは、

$\lim\limits_{n \to \infty} \left(1 + \dfrac{a}{n}\right)^n = e^a$ という式からわかる。分布関数の収束は法則収束と呼ばれる。

これに関して、法則収束するための十分条件は、その特性関数が各点で収束して、その収束先が原点で連続であることである。ところで我々は、Gumbell 分布の特性関数 $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} d\left(e^{-e^{-x}}\right) = \int_{-\infty}^{\infty} e^{itx} e^{-x} e^{-e^{-x}} dx = \int_0^{\infty} x^{it} e^{-x} dx = \Gamma\left(1 - it\right)$ を知っている。ここでガンマ関数 $\Gamma(z)$ はガウス平面の右半平面では解析的 (holomorphic)であるから $\Gamma\left(1 - it\right)$ は t について連続どころか C^{∞} でさえある。

したがって分布の収束をいうには確率変数$M_n-\log n$の特性関数$\varphi_n(t)$が$n\to\infty$

各点t で$\varphi(t)$ に収束することを言えばよい。$\varphi_n(t)$を計算しよう。

$$P(M_n-\log n\le x)=\begin{cases}\left(1-\dfrac{e^{-x}}{n}\right)^n & (x\ge -\log n)\\ 0 & (x<-\log n)\end{cases}$$

より$M_n-\log n$の密度関数$f_n(x)$ は、微分して

$$f_n(x)=\begin{cases}e^{-x}\left(1-\dfrac{e^{-x}}{n}\right)^{n-1} & (x\ge -\log n)\\ 0 & (x<-\log n)\end{cases}$$

である。したがって、$\varphi_n(t)=\displaystyle\int_{-\infty}^{\infty}e^{itx}f_n(x)dx=\int_{-\log n}^{\infty}e^{itx}e^{-x}\left(1-\frac{e^{-x}}{n}\right)^{n-1}dx$

において積分と lim の交換が許される条件を確かめて、

$\displaystyle\lim_{n\to\infty}\varphi_n(t)=\int_{-\infty}^{\infty}e^{itx}e^{-x}\lim_{n\to\infty}\left(1-\frac{e^{-x}}{n}\right)^{n-1}dx=\varphi(t)$となり証明終了となるわけだが、ここ

では、ガンマ関数の無限積という観点ではどうなっているか調べてみよう。複

素関数論の教科書から$\displaystyle\Gamma(z)=\lim_{n\to\infty}\frac{n^z n!}{z(z+1)\cdots(z+n)}$ を見つけることができるの

で Gumbell 分布の特性関数は$\displaystyle\varphi(t)=\lim_{n\to\infty}\frac{n^{1-it}\ n!}{(1-it)(2-it)\cdots((n+1)-it)}$ということ

になる。$\varphi_n(t)$がこの無限積の有限部分になっているのではないか。実際計算を

継続して、この真偽を確かめていこう。

$u=e^{-x}$ と変数変換すると上の式は、$\displaystyle\varphi_n(t)=\int_0^n u^{-it}\left(1-\frac{u}{n}\right)^{n-1}du$

となるが、さらに$u=n\tau$ と変数変換すると、$\varphi_n(t)=n^{1-it}\int_0^1 \tau^{-it}(1-\tau)^{n-1}d\tau$

が得られる。α が複素数でも$(z^{\alpha})'=\alpha z^{\alpha-1}$であることに注意して部分積分すると

$$\int_0^1 \tau^{-it}(1-\tau)^{n-1}d\tau=\frac{\tau^{-it+1}}{1-it}(1-\tau)^{n-1}\Big|_0^1+\frac{n-1}{1-it}\int_0^1 \tau^{-it+1}(1-\tau)^{n-2}\,d\tau=\frac{n-1}{1-it}\int_0^1 \tau^{-it+1}(1-\tau)^{n-2}\,d\tau$$

さらにこの部分積分を繰り返し繰り返し実行すると

$$=\cdots=\ \frac{(n-1)!}{(1-it)(2-it)\cdots(n-it)}$$

となるので、$\varphi_n(t)=\dfrac{n^{1-it}(n-1)!}{(1-it)(2-it)\cdots(n-it)}$

がわかり、無限積の意味でも$\varphi_n(t)\to\varphi(t)$ が言えた。しかし、なぜこのような回り道をする必要があるのか？$\varphi_n(t)$ は$\dfrac{k}{(k-it)},k=1,2\cdots$ という関数の積で書けている。いま独立な確率変数X,Y があるとき$X+Y$ の特性関数は$Ee^{it(X+Y)}=Ee^{itX}Ee^{itY}$ という積になる。つまり、$\varphi_n(t)$は何か独立な確率変数の和の特性関数になっているのではないかという疑問が起きる。平均$\alpha^{-1}>0$ の指数分布の密度関数は$\alpha e^{-\alpha x}$ で、特性関数は$\varphi(t)=\alpha\int_0^\infty e^{itx}e^{-\alpha x}dx=\dfrac{\alpha}{\alpha-it}$ となり、$\alpha=k$ と置いてやれば上の関数が出てくる。したがって次の定理を得る。

互いに独立な確率変数列 $X_1, X_2, \cdots$ で各々の分布が指数分布、すなわち

$$P(X_k \le x) = \begin{cases} 1-e^{-x} & (x \ge 0) \\ 0 & (x < 0) \end{cases} \qquad k = 1, 2, \cdots$$

を持つとき、

$S_n = X_1 + \dfrac{X_2}{2} + \cdots + \dfrac{X_n}{n}$ の分布は

$M_n = \max\limits_{1 \le k \le n} \{X_k\}$ の分布に等しい

証明） $S_n - \log n$ の特性関数を計算する。

$\varphi_{S_n - \log n}(t) = Ee^{it(S_n - \log n)} = n^{-it} Ee^{itX_1} Ee^{it\frac{X_2}{2}} Ee^{it\frac{X_3}{3}} \cdots Ee^{it\frac{X_n}{n}}$ において、

$Ee^{it\frac{X_k}{k}} = \displaystyle\int_0^\infty e^{it\frac{x}{k}} e^{-x} dx = \frac{k}{k-it}$ となるので、

$\varphi_{S_n - \log n}(t) = n^{-it} \dfrac{n!}{(1-it)(2-it)\cdots(n-it)}$ を得て、これは $M_n - \log n$ の特性関数

$\varphi_n(t) = \dfrac{n^{1-it}(n-1)!}{(1-it)(2-it)\cdots(n-it)}$ に一致している。■

系；

互いに独立な確率変数列 $X_1, X_2, \cdots$ で各々の分布が平均 1 の指数分布

であるとき、$S_n = X_1 + \dfrac{X_2}{2} + \cdots + \dfrac{X_n}{n}$ とおくと

$S_n - \log n$ の漸近分布は Gumbell 分布である。