POISSON APPROXIMATION FOR SUM OF BERNOULLI RANDOM VARIABLES AND ITS APPLICATION TO EWENS SAMPLING FORMULA

鹿児島大学名誉教授　大和　元

$n(\geq 1)$ 個の自然数 $\{1, 2, \cdots, n\}$ の確率分割として良く知られた Ewens sampling formula (ESF) について、分割の個数 K_n の確率分布は $P(K_n = k) =\mid s(n,k) \mid \theta^k/\theta^{[n]}(k = 1, 2, \ldots, n)$ で与えられる。ここで、$s(n, k)$ は第 1 種スターリング数、$\theta > 0$, $\theta = \theta(\theta+1)\cdots(\theta+n-1)$。$K_n$ は独立なベルヌーイ確率変数の和 $K_n = \xi_1 + \xi_2 \cdots + \xi_n$ としても表される。但し、$P(\xi_j = 0) = (j-1)/(\theta+j-1)$,　$P(\xi_j = 1) = (\theta)/(\theta+j-1)$　$(j = 1, 2, \cdots)$。この事から、K_n は漸近正規性をもつ事が知られている。(例えば、Johnson et al. (1997), chap. 41)。K_n の平均 $\mu_n = E(K_n) = \theta[\psi(\theta+n) - \psi(\theta)]$ 、分散 $V(K_n) = \theta[\psi(\theta+n) - \psi(\theta)] + \theta^2[\psi'(\theta+n) - \psi'(\theta)]$ の digamma ft. ψ, trigamma ft. ψ' は R に含まれており、R により良い正規近似が得られる。一方、ポアソン近似については、Arratia et al. (2000) により、ESF を含むモデル Logarithmic combinatorial structure で、詳細に求められている。

以上の ESF の研究とは別に、独立なベルヌーイ確率変数 $X_1, X_2, \cdots, X_n$ の和 S_n については次が示されている：$P(X_j = 1) = p_j$, $P(X_j = 0) = 1 - p_j$ とし、$\lambda_n = \sum_{j=1}^n = p_j$, $\lambda_{2,n} = \sum_{j=1}^n = p_j^2$ とおく。$\lambda_n \to \lambda(>0)$, $\lambda_{2,n} \to 0$ $(n \to \infty)$ の場合に限り S_n はポアソン分布 $Po(\lambda)$ に法則収束する (Wang (1993))。ESF については、確率 1 で $\xi_1 = 1$ であるから、$K_n = 1 + L_n$ $(L_n = \xi_2 \cdots + \xi_n)$ として、L_n へこの法則収束の結果を用いて次を得た。

[**命題**] ESF について： $\theta \log n \to \lambda(>0)$ の場合に限り、L_n はポアソン分布 $Po(\lambda)$ へ法則収束する。このとき、K_n は Shifted ポアソン分布 $1 + Po(\lambda)$ へ法則収束する。

このことから、ESF の K_n の確率への近似として、Shifted ポアソン分布 $1 + Po(\mu_n)$ が考えられる。更には、Charlier polynomial $C_k(x, \lambda)$ of degree k $\big(C_k(x,\lambda) = d^k p(x,\lambda)/d\lambda^k\big/(p(x,\lambda)$, $p(x,\lambda) = e^{-\lambda}\lambda^k/x!$、竹内 (1975), Zacharovas and Hwang (2010)$\big)$ を用いると、$P(K_n = m)$ へのより良い近似が得られる；

$$1 + p(m, \mu_n)\Big(1 - \frac{\mu_{2,n}}{2} C_2(m, \mu_n)\Big)。$$

但し、$\mu_{2,n} = \theta^2[\psi'(\theta+1) - \psi'(\theta+n)]$, $C_2(x, \lambda) = \big[x^2 - (2\lambda+1)x + \lambda^2\big]/\lambda^2$。これらの近似は、Arratia et al. (2000) から得られる ESF の K_n の確率のポアソン分布に依る近似よりも良い。

参考文献

[1] Arratia, R., Barbour, A.D. and Tavaré, S. (2000). The number of components in logarithmic combinatorial structure. *Annals of Applied Probability*, **10**, 331–361.

[2] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997). *Discrete multivariate distributions*, John Wiley & Sons, New York.

[3] 竹内 啓 (1975). 確率分布の近似、教育出版.

[4] Wang, Y.H. (1993). On the number of successes in independent trials. *Statistica Sinica*, **3**, 295–312.

[5] Zacharovas, V. and Hwang, H.-K. (2010). A Charlier-Parseval approach to Poisson approximation and its applications. *Lithuania Math. J.*, **50**, 88–119.