Ewens 標本抽出公式の分割の個数の分布への二項分布に基づく近似

鹿児島大学名誉教授　大和　元

自然数の集合 $\{1,2,\cdots,n\}$ の確率分割として良く知られた Ewens 標本抽出公式について、分割の個数 K_n の確率関数 (pf) 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、$\theta>0$, $\theta^{[n]}=\theta(\theta+1)\cdots(\theta+n-1)$ ([2] 等)。K_n は独立なベルヌーイ確率変数の和 $K_n=1+L_n$ a.s. $(L_n=\xi_2+\cdots+\xi_n)$ としても表される。但し、$P(\xi_j=1)=p_j, P(\xi_j=0)=1-p_j$ ($p_j=\theta/(\theta+j-1)$: $j=2,3,\cdots,n$)。昨年、 $\theta\log n\to\lambda(>0)$ の場合に限り K_n は Shifted Poisson distribution $1+Po(\lambda)$ へ法則収束すること、更に、K_n の分布 $\mathcal{L}(K_n)$ への近似として、Shifted Poisson distribution と Charlier polynomial による展開に基づく近似を紹介した 。今回は、$\mathcal{L}(L_n)$ (従って、$\mathcal{L}(K_n)$) への近似として、二項分布に基づく近似を述べるが、ここにはその中の 2 つを記している。なお、$\lfloor x\rfloor$ は x に近い整数を表し、p_j の和と二乗和を次とおく；

$$\lambda_{n-1}=\sum_{i=2}^{n}p_j,\ \lambda_{2,n-1}=\sum_{i=2}^{n}p_j^2\ 。$$

(i) 平均と分散が L_n のそれに略等しい二項分布 $B((n-1)',p')$ を、$\mathcal{L}(L_n)$ への近似とし ([1])、従って、$\mathcal{L}(K_n)$ への近似として Shifted binomial distribution $1+B((n-1)',p')$ を用いる；

$$(n-1)'=\lfloor\lambda_{n-1}^2/\lambda_{2,n-1}\rfloor,\ p'=\lambda_{n-1}/(n-1)'\ 。$$

(ii) Krawtchouk polynomial による展開 ([3], [5]) を用いる：$\bar{p}_{n-1}=\lambda_{n-1}/(n-1)$, $\gamma_2(\bar{p}_{n-1})=\lambda_{2,n-1}-(n-1)\bar{p}_{n-1}^2$ とし、$g_B(x;n,p)$ を $B(n,p)$ の pf とする。 Signed pf g_B^* ([3], [4]) を次とおく $(x=0,1,\cdots,n-1)$;

$$g_B^*(x;n-1,\bar{p}_{n-1})=g_B(x;n-1,\bar{p}_{n-1})-\frac{\gamma_2(\bar{p}_{n-1})}{2}\Delta^2 g_B(x;n-3,\bar{p}_{n-1}),$$

$$\Delta^2 g_B(x;n-3,p)=\frac{g_B(x;n-1,p)}{(n-1)(n-2)p^2(1-p)^2}\Big\{x^2-\big[1+2(n-2)p\big]x+(n-1)(n-2)p^2\Big\}\ 。$$

$\mathcal{L}(K_n)$ の pf への近似には、 g_B^* を右へ shift し $g_B^*(x-1;n-1;\bar{p}_{n-1})$ $(x=1,\cdots,n)$ を用いる。

$\mathcal{L}(K_n)$ への近似として、これら二項分布に基づく近似を比較すると共に、Shifted Poisson distribution 及び Charlier polynomial による展開を用いた近似、更に正規近似との比較を与える。応用として、(i) を基にした K_n の pf の推定を述べる。

参考文献

[1] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson approximation.* Clarendon Press, Oxford.

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

[3] Roos, B. (2001). Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Theory of Probability and Its Applications*, **45**, 258–272.

[4] 竹内 啓 (1975). 確率分布の近似、教育出版.

[5] Takeuchi, K. and Takemura, A. (1987). On sum of 0-1 random variables I. univariate case. *Ann. Inst. Statist. Math.*, **39**, 85–102.

[6] Yamato, H. (2017). Shifted Binomial approximation for the Ewens sampling formula. *Bulletin of Informatics and Cybernetics*, **49**, 81–88.