正規母集団からの順序統計量期待値の一つの漸近的性質

山内, 二郎
東京大学工学部応用物理教室

https://doi.org/10.15017/12721

出版情報：統計科学研究. 1 (3), pp.1-5, 1956-09. Research Association of Statistical Sciences

寄　書

正規母集団からの順序統計量期待値の一つの漸近的性質

山内二郎（東大工学部応用物理教室）

1. はしがき

正規母集団からの標本から，その標準偏差の不偏推定量を求めるのに，標本の大きさが大きいときMosteller[1]が順序統計量を用いることを研究しているし，Dixon-Masseyの著書[2]にも載っている．最近のように諸所に統計機械が備えられ，この方法を利用することが容易になっているので，実施される時期になったと考える．この場合に二つの問題を考えておく必要がある．

一つは用いるデータの数をきめて効率の最も高い分位点が定められるが，果して何番目のデータを撰らび出せばよいかという問題，もう一つは撰んだデータに対して，不偏推定量とするための係数をどう調整したらよいかという問題である．このために数年前日本数学会年会の席上発表した順序統計量の期待値についての漸近的性質が役立つと考えて，それをもとにして論ずることとした．

2. 正規母集団からの順序統計量の一つの漸近的性質

平均値0，標準偏差1の正規母集団からの順序統計量として一般性を失わない．便宜のため大きさnの標本を大きさの小さい方から順にならべ

$$x_1 < x_2 < x_3 < \cdots < x_{n-1} < x_n$$

とし，

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\ exp\ \left(-\frac{x^2}{2}\right)$$

$$\lambda = \int_{-\infty}^{x} \varphi(x)\,dx$$

と書くと，上から m 番目 x_{n-m+1} の密度分布は周知のとおり

$$dF = \frac{n!}{(n-m)!(m-1)!} \lambda^{n-m} (1-\lambda)^{m-1} \varphi(x)\, dx$$

である．Gumbel [3] はこのモードの値を μ_m とし

$$\alpha_m = \frac{n}{m} \varphi(\mu_m)$$

とすると，$n \to \infty$ のとき x_{n-m+1} の期待値

$$\bar{x}_{n-m+1} = \mu_m + \frac{1}{\alpha_m} A_m$$

$$A_m = \log m - \sum_{\nu=1}^{m-1} \frac{1}{\nu} + C$$

C : Euler 定数 $= 0,577'22\cdots$

を求めている．

$\bar{x}_{n-m+1} > x$ である確率を $\bar{\lambda}_m$ と書くと

$$\bar{\lambda}_m = \lambda_m + \int_{\mu_m}^{\mu_m + \frac{1}{\alpha_m} A_m} \varphi(x)\, dx$$

であるから，$x = \mu_m + y\alpha_m^{-1}$ を入れると

$$\bar{\lambda}_m = \lambda_m + \frac{\varphi_m}{\alpha_m} \int_0^{A_m} exp \left(-y - \frac{y^2}{2\alpha_m^2} \right) dy$$

$$= \lambda_m + \frac{m}{n} \left\{ 1 - exp(-A_m) - \frac{1}{\alpha_m^2} \left[1 - e^{-A_m} \left(1 + A_m + \frac{A_m^2}{2} \right) \right] \right.$$

$$\left. + \frac{4!}{2^2 2! \alpha_m^4} \left[1 - e^{-A_m} \left(1 + A_m + \frac{A_m^2}{2!} + \cdots + \frac{A_m^4}{4!} \right) \right] - + \cdots \right\}$$

すなわち

$$\lim_{n \to \infty} n(1 - \bar{\lambda}_m) = m - m \left\{ 1 - exp(-A_m) \right\}$$

$$= m\, exp(-A_m)$$

$m = 1$: $exp(-C) = 0,56146$

$m = 2$: $exp(1-C) = 1,526\,21$

$m = 3$: $exp(1.5-C) = 2,516\,31$

m が大きいときは

$$m\, exp(-A_m) = \left(m - \frac{1}{2} \right) exp\ D_m$$

とかくと，

$$D_m = \sum_{\nu=1}^{m-1} \frac{1}{\nu} - C - \log\left(m-\frac{1}{2}\right)$$

$$= -\frac{1}{m} + \sum_{\nu=1}^{m} \frac{1}{\nu} - C - \log\left(m-\frac{1}{2}\right)$$

$$= -\frac{1}{m} - \log\left(m-\frac{1}{2}\right) + \log m + \frac{1}{2m} - \frac{B_1}{2m^2} + \frac{B_3}{4m^4} - \frac{B_5}{6m^6} + \int_m^{\infty} \frac{B_6(x)}{x^8}\,dx$$

B_1, B_3, B_5 はベルヌーイ数，最后の項は m が 4 でも 8D で 6 単位，$m=5$ で 8D で 1 単位，m がさらに大きいとさらに小さいからこの項を省略し

$$D_m = -\log\left(1-\frac{1}{2m}\right) - \frac{1}{2m} - \frac{B_1}{2m^2} + \frac{B_3}{4m^4} - \frac{B_5}{6m^6}$$

$$= \frac{1}{24m^2} + \frac{1}{24m^3} + \frac{23}{960m^4} + \frac{1}{160m^5}$$

$m=4$ の場合 8D で 0.003 354 57 を得て，7D の正しい値は 0.003 354 6 である．

したがって次のような結果が得られた．

$$\lim_{n\to\infty} n\left(1-\bar{\lambda}_m\right) = \left(m-\frac{1}{2}\right) \exp D_m$$

$$= m - \frac{1}{2} + \frac{1}{24m} + \frac{1}{48m^2} + \frac{23}{5760m^3} - \frac{17}{3840m^4} - \frac{11471}{40960.81m^5}$$

オ 1 表は $m \leq 20$ の場合の 4D の計算値である．

第 1 表　$\lim_{n\to\infty} n\left(1-\bar{\lambda}_m\right)$ の値

m		m		m	
1	0.5615	8	7.5055	15	14.5029
2	1.5262	9	8.5049	16	15.5027
3	2.5163	10	9.5044	17	16.5025
4	3.5118	11	10.5040	18	17.5024
5	4.5092	12	11.5036	19	18.5023
6	5.5075	13	12.5033	20	19.5021
7	6.5064	14	13.5031		

たとえば $m=1$ のときは，最大値（最小値）についての分位を与えるが，n が大きくても α_m があまり大きくならないために正しい値から少しはずれる．Tippett [4] による値によって $n(1-\bar{\lambda}_m)$ を計算すると才2表のようになり，0,5615より少し大きい．

第　2　表

n	Tippettの値 $\bar{x}_n$	$n(1-\bar{\lambda}_n)$
100	2.5076	0.6078
500	3.0367	0.5978
1000	3.2414	0.5923

3. 分位に応ずる順位

分位 $1-\bar{\lambda}_m$，$\bar{\lambda}_m$ を与えたときそれに応ずる m を撰らぶには，前節の結果から $n(1-\bar{\lambda}_m)$ がほゞ $m-\frac{1}{2}$ に等しいことから，$n(1-\bar{\lambda}_m)$ を整数にするときにこれを切りあげることにしておき，$\bar{\lambda}_m$ の方は中央に対して対称にとれば，かなり近いものを採ることになる．

二つのデータから標準偏差を推定する最も効率の高い不偏推定量としてMosteller は $1-\bar{\lambda}_m = 0,068$ にえらぶと述べている．詳しく計算すると 0,06916 であるが，かりに $n=840$ とすれば

$$840 \times 0.06916 = 58.09$$

であるから，小さい方から59番目，大きい方から 782 番目をとる．

もう一つの問題である不偏推定量とするための係数としては別に研究を要するが，それができるまでは前節により

$$1-\bar{\lambda}_m = \frac{1}{840}\left(59-\frac{1}{2}+\frac{1}{24\times 59}+\cdots\right)$$

$$= 0.0696429$$

$$\bar{\lambda}_m = 0.9303571$$

$$\bar{x}_{n-m+1} = 1.4785$$

$$\hat{\sigma} = \frac{x(\lambda=0.9304)-x(\lambda=0.0696)}{2.957}$$

とすればよいと考える．しかしこの問題はこれで解決されているとはいえない．

参　考　文　献

[1] F. Mosteller : *ASM*, 17 (1946), 377.

[2] W. J. Dixon - F. J. Massey : "*Introduction to Statistical Analysis*" 1951 (McGraw-Hill Book Co.)

[3] E. J. Gumbel : *ASM*, 15 (1944), 414 - 422.

[4] L. H. C. Tippett : *Biometrika*, 17 (1925), 364 - 387.
