平均の推定に於ける近似変動係数の利用

平　野　勝　臣

(1971年11月　受付)

Using Some Approximately Known Coefficient of Variation in Estimating Mean

Katuomi Hirano

Let μ and σ^2 be the unknown mean and variance of a population distribution, respectively, and let $v_0=\sigma/\mu$ be a true coefficient of variation of its population. Searls (1964) proposed the estimator $\bar{x}'$ for the mean μ that made use of a known coefficient of variation v_0 which is a true value. But in many cases what we can know, in practice, is the estimated or approximate value of coefficient of variation $v=v_0\alpha$ $(\alpha>0)$.

In this paper we consider an estimator $\hat{x}$, which substitutes v for v_0 in the estimator $\bar{x}'$, and give the range of α that $\hat{x}$ is more precise than the usual estimator (that is, the sample mean).

The Institute of Statistical Mathematics

§0　内容の概略

推定問題に関するベイズ流の思想は事前情報の利用の原理に基づいている．このノートでは，ある事前情報（母集団の近似変動係数値が既知である）を用いて，平均に対する推定量を与え，この推定量が平均に対する通例の推定量（標本平均）を用いたときよりも精度が上がっていることを示し（真の変動係数値が既知のとき Searls を参照），且つ，その事前情報に含まれる誤差がどの程度までならば，通例の推定量より正確であるのか，その範囲を示す．

§1.　問題の背景

未知な平均μと未知な分散σ^2をもつ母集団分布に対して，$v_0=\sigma/\mu$をこの母集団分布の真の変動係数とする．Searls [1] は，真の変動係数v_0を既知としたときに（つまり，これを事前情報とする）平均μの推定方法を示した．その方法は次の通りであった．

$x_1, x_2, \cdots x_n$を長さnの標本とし

$$\bar{x}' = w\sum_{i=1}^{n} x_i \tag{1}$$

と定義する．但し，wは定数で$\bar{x}'$の2乗平均誤差$E(\bar{x}'-\mu)^2$が最小となるように選ぶ．即ち，それを計算して求めると

$$MSE(\bar{x}') \equiv E(\bar{x}'-\mu)^2 = n w^2\sigma^2 + \mu^2(1-nw)^2 \tag{2}$$

となり，この (2) の値をwの2次函数とみて最小となるwの値を計算すれば

$$w = \frac{1}{n+v_0^2} \tag{3}$$

を得る．これを (2) に代入して，$\bar{x}'$の2乗平均誤差を求めると

$$MSE(\bar{x}') = \frac{\sigma^2}{n+v_0^2} \tag{4}$$

となる．平均μに対する通例の推定量は標本平均$\bar{x}=\sum_{i=1}^{n} x_i/n$で，$\bar{x}$に関する$\bar{x}'$の相対効率

$(REF(\bar{x}')$ と書く) を

$$REF(\bar{x}')=\frac{MSE(\bar{x})}{MSE(\bar{x}')}$$

と定義すると，標本平均 $\bar{x}$ は μ の不偏推定量であるから $MSE(\bar{x})=Var(\bar{x})=\sigma^2/n$ となることに注意すれば (4) から

$$REF(\bar{x}')=\frac{n+v_0^2}{n}=1+\frac{v_0^2}{n} \tag{5}$$

を得て，Searls は $v_0=1,2,\cdots,5$ と標本数 $n=5, 10, 50, 100, 500$ に対して $REF(\bar{x}')$ を求め表で与えた.

真の変動係数値が得られるとすれば，大体，平均 μ と分散 σ^2 がわかっていて，平均の必要はない．我々が得る変動係数は推定された変動係数値であろうし，実際に多くの場合，誤差を含んだ近似変動係数値である．更に，真の変動係数 v_0 が得られたとしても，その値が3以上の数値をとることは，ほとんどあり得ないであろう（理論的にはいくらでも大きい値をとり得る)．むしろ，1より小さい値について考える方が有用であると思われる．以上のことをふまえて，我々が事前情報として入手できる変動係数 v は $v=v_0\alpha\,(\alpha>0)$ とし，真の値 v_0 に対して，$100\,\alpha\%$ の誤差を含んだものとして考えることにする（勿論，$\alpha=1$ のとき，v は真の変動係数となる).

§2. 考　察

母集団分布からの標本 $x_1, x_2, \cdots, x_n$ に対して

$$\tilde{x}=w\sum_{i=1}^{n}x_i$$

と定義する．我々が入手できる事前情報は真の変動係数 v_0 に対して $100\,\alpha\%$ の誤差を含んだ $v=v_0\alpha$ だけであると仮定してあるので，(3) において v_0 の代わりに v を代入すると

$$w=\frac{1}{n+v^2}$$

となるので，$\tilde{x}$ の2乗平均誤差 $MSE(\tilde{x})$ は

$$MSE(\tilde{x})=\frac{n\sigma^2+\mu^2v^4}{(n+v^2)^2}$$

となる．それ故，標本平均 $\bar{x}$ に関する $\tilde{x}$ の相対効率 $REF(\tilde{x})$ は

$$REF(\tilde{x})\equiv\frac{MSE(\bar{x})}{MSE(\tilde{x})}=\frac{\sigma^2(n+v^2)^2}{n(n\sigma^2+\mu^2v^4)}=\frac{\sigma^2(n+v_0^2\alpha^2)^2}{n(n\sigma^2+\mu^2v_0^4\alpha^4)} \tag{6}$$

となる．ここで我々が興味をもつのは，通例の推定量である標本平均 $\bar{x}$ よりも $\tilde{x}$ の方がより正確である α の範囲である．その α の範囲は (6) から α に関する次の不等式を満す α の範囲に等しい：

$$REF(\tilde{x})=\frac{\sigma^2(n+v_0^2\alpha^2)^2}{n(n\sigma^2+\mu^2v_0^4\alpha^4)}\geqq 1. \tag{7}$$

不等式 (7) を解くと次の様な α の範囲を得る：

i) $v_0^2>n$ のとき

不等式 (7) は $\alpha^2>-2n/(v_0^2-n)$ となり，任意の α に対して成り立つ.

ii) $v_0^2=n$ のとき

$REF(\tilde{x})=\alpha^2+1$ となり，不等式 (7) は任意の α に対して成り立つ.

iii) $v_0^2<n$ のとき

$$\sqrt{\frac{2n}{n-v_0^2}} > \alpha > 0.$$

注意 $\sqrt{2n/(n-v_0^2)}$ は n に関して単調に減少しその極限は $\lim_{n\to\infty}\sqrt{2n/(n-v_0^2)}=\sqrt{2}$ で，任意の n に対して，$(\sqrt{2n/(n-v_0^2)},\ 0)\supset(\sqrt{2},\ 0)$ で $\sqrt{2}>\alpha>0$ は α の最も狭い範囲である．

以上の α の範囲を表1に与えた．つまり表1の α の範囲は $\bar{x}$ よりも $\tilde{x}$ の方が正確である範囲を与えている．

表1　不等式 (7) を満す α の範囲

v_0	sample size n 5	10	20	50	100	500	1,000
0.01	0~1.414	0~1.414	0~1.414	0~1.414	0~1.414	0~1.414	0~1.414
0.05	0~1.415	0~1.414	0~1.414	0~1.414	0~1.414	0~1.414	0~1.414
0.10	0~1.417	0~1.415	0~1.415	0~1.414	0~1.414	0~1.414	0~1.414
0.15	0~1.417	0~1.416	0~1.415	0~1.415	0~1.414	0~1.414	0~1.414
0.20	0~1.420	0~1.417	0~1.416	0~1.415	0~1.414	0~1.414	0~1.414
0.25	0~1.423	0~1.419	0~1.416	0~1.415	0~1.415	0~1.414	0~1.414
0.30	0~1.427	0~1.421	0~1.417	0~1.415	0~1.415	0~1.414	0~1.414
0.35	0~1.432	0~1.423	0~1.419	0~1.416	0~1.415	0~1.414	0~1.414
0.40	0~1.437	0~1.426	0~1.420	0~1.416	0~1.415	0~1.414	0~1.414
0.45	0~1.444	0~1.429	0~1.421	0~1.417	0~1.416	0~1.415	0~1.414
0.50	0~1.451	0~1.432	0~1.423	0~1.418	0~1.416	0~1.415	0~1.414
0.55	0~1.459	0~1.434	0~1.425	0~1.419	0~1.416	0~1.415	0~1.414
0.60	0~1.468	0~1.440	0~1.427	0~1.419	0~1.417	0~1.415	0~1.414
0.65	0~1.478	0~1.445	0~1.429	0~1.420	0~1.417	0~1.415	0~1.415
0.70	0~1.489	0~1.450	0~1.432	0~1.421	0~1.418	0~1.415	0~1.415
0.75	0~1.501	0~1.456	0~1.435	0~1.422	0~1.418	0~1.415	0~1.415
0.80	0~1.514	0~1.462	0~1.437	0~1.423	0~1.419	0~1.415	0~1.415
0.85	0~1.529	0~1.468	0~1.440	0~1.425	0~1.419	0~1.415	0~1.415
0.90	0~1.545	0~1.475	0~1.444	0~1.426	0~1.420	0~1.415	0~1.415
0.95	0~1.562	0~1.483	0~1.447	0~1.427	0~1.421	0~1.415	0~1.415
1.00	0~1.581	0~1.491	0~1.451	0~1.429	0~1.421	0~1.416	0~1.415
1.10	0~1.624	0~1.508	0~1.459	0~1.432	0~1.423	0~1.416	0~1.415
1.20	0~1.676	0~1.529	0~1.468	0~1.435	0~1.425	0~1.416	0~1.415
1.30	0~1.738	0~1.551	0~1.478	0~1.439	0~1.426	0~1.417	0~1.415
1.40	0~1.814	0~1.577	0~1.489	0~1.443	0~1.428	0~1.417	0~1.416
1.50	0~1.907	0~1.606	0~1.501	0~1.447	0~1.430	0~1.417	0~1.416
2.00	0~3.162	0~1.826	0~1.581	0~1.474	0~1.443	0~1.420	0~1.417

§3 結　論

注意によると，平均 μ の推定量 $\tilde{x}$ が標本平均 $\bar{x}$ を用いて推定するよりもより正確である α の最も小さい範囲は $\sqrt{2}>\alpha>0$ である．だから $v>v_0$ のとき，近似変動係数が約 41% の誤差を含んでいても，また $v<v_0$ のとき，約 100% の誤差を含んでいても推定量 $\tilde{x}$ を用いる方が正確であるといえる．そして，平均 μ を推定する際に変動係数の事前情報を用いるならば，いつも変動係数を小さめに評価しておけば，このノートで示した推定方法が有効であると結論できる．

Searls は小標本のときにこの推定方法の効率がよいといっているが，それは変動係数値が大きいときに，といいなおすべきであろう．即ち，平均 μ を 0 に近づければ，v_0 はいくらでも

大きくなり，(5) で与えられた $REF(\bar{x}')=1+v_0{}^2/n$*1 は，固定した n に対しいくらでも大きくなる．だが，前にも指摘したように変動係数値が実際問題として，3以上をとるということは，ほとんどあり得ない．相対効率が高いという理由で，推定量 $\tilde{x}$ もしくは $\bar{x}'$ を用いよ，と主張したいならば変動係数値が1より小さいときであり，我々は実際，そのときに相対効率を計算してみて*2，小標本 $n=5\,(10)$ で相対効率は 1.20(1.10) を越えてない．これは同じ精度ならば約83% (約91%) の標本数でよいということである．$n=20$ となると各 α と v_0 に対して相対効率は 1.00〜1.05 の場合がほとんどであった．よって相対効率の高さで，このノートの方法の有効性を主張したいとすれば，せいぜい標本数が10以下で，それも強い主張とはいえない．しかし我々は，むしろ相対効率の高さよりも，かなり大きい巾の誤差を含む近似変動係数でもこの方法が使えるという意味で，この方法が有効であると主張したい．

謝辞　本稿の作成にあたりまして，当研究所の高橋宏一室長に，数値計算にあたりましては，荒畑恵美子さんにお世話になりました．ここに感謝の意を表します．

統計数理研究所

引　用　文　献

[1] Searls, Donald J., The utilization of a known coefficient of variation in the estimation procedure, Journal of the Ameıican Statistical Association, (1964) 59 1225-1226.

---

*1　自明ではあるが，$REF(\bar{x}')$ は n に関する減少関数で，小標本に対して相対効率が高いことに注意

*2　v_0 を0から0.05きざみに1まで，1からは1きざみに10まで与え，α は0.1から0.1きざみに2まで与え，更に $n=5, 10, 20, 50, 100, 500, 1000$ として $REF(\tilde{x})$ を求めた．又，$\alpha=1$ のとき即ち，変動係数が真のとき各 v_0，n に対し $REF(\tilde{x})$ は最大となる．