不偏分散の分散の推定

KAZOON

2016 年 11 月 15 日

1　はじめに

[1] では，標本 $\{X_j \mid 0 \le j \le n-1\}$ に対して，不偏分散

$$V := \frac{1}{n-1}\sum_{j=0}^{n-1}(X_j - \overline{X})^2, \tag{1}$$

$$\overline{X} := \frac{1}{n}\sum_{j=0}^{n-1} X_j \tag{2}$$

の分散が

$$\left\langle (V-\sigma^2)^2 \right\rangle = \frac{1}{n}\left\langle (x-\langle x\rangle)^4 \right\rangle - \frac{n-3}{n(n-1)}\left\langle (x-\langle x\rangle)^2 \right\rangle^2 \tag{3}$$

となることを示し，これを標本から推定する際に，

$$\left\langle (V-\sigma^2)^2 \right\rangle \sim \frac{1}{(n-1)(n^2-3n+3)}[n^2K' - (n^2-3)V^2], \tag{4}$$

$$K' := \frac{1}{n}\sum_{j=0}^{n-1}(X_j - \overline{X})^4 \tag{5}$$

またはその近似である

$$\left\langle (V-\sigma^2)^2 \right\rangle \sim \frac{1}{n}[K' - V^2] \tag{6}$$

を用いるとよいと主張している．

しかし，式 (4) を求める際に，標本 4 次モーメント K' の期待値が

$$\langle K' \rangle = \frac{(n-1)(n^2-3n+3)}{n^3}\left\langle (x-\langle x\rangle)^4 \right\rangle + \frac{3(n-1)(2n-3)}{n^3}\left\langle (x-\langle x\rangle)^2 \right\rangle^2 \tag{7}$$

で表される事実を利用している一方で，不偏分散 V の 2 乗の期待値が $\left\langle (x-\langle x\rangle)^2 \right\rangle^2$ にはならない事実を無視しているため，式 (4) の右辺は不偏推定量になっておらず，(4) を何のために導出したのか，意味不明になっている．本文書では，適当な n の有理関数で重み付けした K' と V^2 の和で，不偏分散の分散の不偏推定量

$$\left\langle (V-\sigma^2)^2 \right\rangle \sim \frac{1}{n(n-2)(n-3)}[n^2K' - (n^2-3)V^2] \tag{8}$$

が得られることを示す．

2　計算

2.1　不偏分散の 2 乗の期待値

$$
\begin{aligned}
\left\langle V^2 \right\rangle &= \left\langle \left(\frac{1}{n-1} \sum_{j=0}^{n-1} \left(X_j - \frac{1}{n} \sum_{k=0}^{n-1} X_k \right)^2 \right)^2 \right\rangle && (9)\\
&= \frac{1}{(n-1)^2} \sum_{j=0}^{n-1} \sum_{j'=0}^{n-1} \left\langle \left(X_j - \frac{1}{n} \sum_{k=0}^{n-1} X_k \right) \left(X_j - \frac{1}{n} \sum_{l=0}^{n-1} X_l \right) \left(X_{j'} - \frac{1}{n} \sum_{k'=0}^{n-1} X_{k'} \right) \left(X_{j'} - \frac{1}{n} \sum_{l'=0}^{n-1} X_{l'} \right) \right\rangle && (10)\\
&= \frac{1}{n^4(n-1)^2} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \sum_{j'=0}^{n-1} \sum_{k'=0}^{n-1} \sum_{l'=0}^{n-1} \langle (X_j - X_k)(X_j - X_l)(X_{j'} - X_{k'})(X_{j'} - X_{l'}) \rangle && (11)\\
&= \frac{1}{n^4(n-1)^2} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \sum_{j'=0}^{n-1} \sum_{k'=0}^{n-1} \sum_{l'=0}^{n-1} \langle X_j^2 X_{j'}^2 - X_j^2 X_{j'} X_{k'} - X_j^2 X_{j'} X_{l'} + X_j^2 X_{k'} X_{l'} - X_j X_k X_{j'}^2 \\
&\quad + X_j X_k X_{j'} X_{k'} + X_j X_k X_{j'} X_{l'} - X_j X_k X_{k'} X_{l'} - X_j X_l X_{j'}^2 + X_j X_l X_{j'} X_{k'} + X_j X_l X_{j'} X_{l'} \\
&\quad - X_j X_l X_{k'} X_{l'} + X_k X_l X_{j'} X_{k'} - X_k X_l X_{j'} X_{l'} + X_k X_l X_{k'} X_{l'} \rangle && (12)\\
&= \frac{1}{n^2(n-1)^2} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \sum_{j'=0}^{n-1} \left(\left\langle X_j^2 X_k^2 \right\rangle - 2 \left\langle X_j^2 X_k X_l \right\rangle + \langle X_j X_k X_l X_{j'} \rangle \right) && (13)\\
&= \frac{1}{(n-1)^2} \sum_{j=0}^{n-1} \left[\langle X_j \rangle^4 + \sum_{k \neq j} \left\langle X_j^2 X_k^2 \right\rangle \right] - \frac{2}{n(n-1)^2} \sum_{j=0}^{n-1} \left[\left\langle X_j^4 \right\rangle + \sum_{k \neq j} \left\langle X_j^2 X_k^2 \right\rangle + 2 \sum_{k \neq j} \left\langle X_j^3 X_k \right\rangle \right. \\
&\quad \left. \sum_{k \neq j} \sum_{l \neq j,k} \left\langle X_j^2 X_k X_l \right\rangle \right] + \frac{1}{n^2(n-1)^2} \sum_{j=0}^{n-1} \left[\left\langle X_j^4 \right\rangle + 4 \sum_{k \neq j} \left\langle X_j^2 X_k^2 \right\rangle + \sum_{k \neq j} \left\langle X_j^3 X_k \right\rangle \right. \\
&\quad \left. 6 \sum_{l \neq j} \sum_{k \neq j,k} \left\langle X_j^2 X_k X_l \right\rangle + \sum_{k \neq j} \sum_{l \neq j,k} \sum_{j' \neq j,k,l} \langle X_j X_k X_l X_{j'} \rangle \right] && (14)\\
&= \frac{n}{(n-1)^2} \left(\left\langle x^4 \right\rangle + (n-1) \left\langle x^2 \right\rangle \langle x \rangle^2 \right) - \frac{2}{(n-1)^2} \left(\left\langle x^4 \right\rangle + (n-1) \left\langle x^2 \right\rangle^2 + 2(n-1) \left\langle x^3 \right\rangle \langle x \rangle \right. \\
&\quad \left. (n-1)(n-2) \left\langle x^2 \right\rangle \langle x \rangle^2 \right) + \frac{1}{n(n-1)^2} \left(\left\langle x^4 \right\rangle + 4(n-1) \left\langle x^3 \right\rangle \langle x \rangle + 3(n-1) \left\langle x^2 \right\rangle^2 \right. \\
&\quad \left. 6(n-1)(n-2) \left\langle x^2 \right\rangle \langle x \rangle^2 + (n-1)(n-2)(n-3) \langle x \rangle^4 \right) && (15)\\
&= \frac{1}{n} \left(\left\langle x^4 \right\rangle - 4 \left\langle x^3 \right\rangle \langle x \rangle \right) + \frac{n^2 - 2n + 3}{n(n-1)} \left\langle x^2 \right\rangle^2 + \frac{(n-2)(n-3)}{n(n-1)} \left(\langle x \rangle^4 - 2 \left\langle x^2 \right\rangle \langle x \rangle^2 \right). && (16)
\end{aligned}
$$

ここで,

$$
\begin{aligned}
\left\langle (x - \langle x \rangle)^2 \right\rangle^2 &= \left(\left\langle x^2 \right\rangle - \langle x \rangle^2 \right)^2 && (17)\\
&= \left\langle x^2 \right\rangle^2 - 2 \left\langle x^2 \right\rangle \langle x \rangle^2 + \langle x \rangle^4, && (18)\\
\left\langle (x - \langle x \rangle)^4 \right\rangle &= \left\langle x^4 - 4x^3 \langle x \rangle + 6x^2 \langle x \rangle^2 - 4x \langle x \rangle^3 + \langle x \rangle^4 \right\rangle && (19)\\
&= \left\langle x^4 \right\rangle + 4 \left\langle x^3 \right\rangle \langle x \rangle + 6 \left\langle x^2 \right\rangle \langle x \rangle^2 - 3 \langle x \rangle^4, && (20)
\end{aligned}
$$

だから，

$$\left\langle V^2 \right\rangle = \frac{1}{n}\left\langle (x-\langle x\rangle)^4 \right\rangle + \frac{n^2-2n+3}{n(n-1)}\left\langle (x-\langle x\rangle)^2 \right\rangle^2. \tag{21}$$

2.2　中心化 4 次モーメントと分散の 2 乗の不偏推定量

K' と V^2 の期待値が，それぞれ中心化 4 次モーメントと分散の 2 乗の線形和となっていることから，逆に K' と V^2 の適当な線形和を考えることで，中心化 4 次モーメントと分散の 2 乗の不偏推定量を得ることができる．

2.2.1　中心化 4 次モーメント

$$\frac{n^3}{3(n-1)(2n-3)}\left\langle K' \right\rangle - \frac{n(n-1)}{n^2-2n+3}\left\langle V^2 \right\rangle = \frac{n^2(n-2)(n-3)}{3(2n-3)(n^2-2n+3)}\left\langle (x-\langle x\rangle)^4 \right\rangle. \tag{22}$$

よって，

$$\begin{aligned}\left\langle (x-\langle x\rangle)^4 \right\rangle &= \frac{3(2n-3)(n^2-2n+3)}{n^2(n-2)(n-3)}\left(\frac{n^3}{3(n-1)(2n-3)}\left\langle K' \right\rangle - \frac{n(n-1)}{n^2-2n+3}\left\langle V^2 \right\rangle\right) && (23)\\ &= \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)}\left\langle K' \right\rangle - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)}\left\langle V^2 \right\rangle. && (24)\end{aligned}$$

これより，中心化 4 次モーメントの不偏推定量

$$\left\langle (x-\langle x\rangle)^4 \right\rangle \sim \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)}K' - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)}V^2 \tag{25}$$

を得る．

2.2.2　分散の 2 乗

同様に，

$$n\left\langle V^2 \right\rangle - \frac{n^3}{(n-1)(n^3-3n+3)}\left\langle K' \right\rangle = \frac{n^2(n-2)(n-3)}{(n-1)(n^2-3n+3)}\left\langle (x-\langle x\rangle)^2 \right\rangle^2. \tag{26}$$

よって，

$$\begin{aligned}\left\langle (x-\langle x\rangle)^2 \right\rangle^2 &= \frac{(n-1)(n^2-3n+3)}{n^2(n-2)(n-3)}\left(n\left\langle V^2 \right\rangle - \frac{n^3}{(n-1)(n^3-3n+3)}\left\langle K' \right\rangle\right) && (27)\\ &= \frac{(n-1)(n^2-3n+3)}{n(n-2)(n-3)}\left\langle V^2 \right\rangle - \frac{n}{(n-2)(n-3)}\left\langle K' \right\rangle. && (28)\end{aligned}$$

これより，分散の 2 乗の不偏推定量

$$\left\langle (x-\langle x\rangle)^2 \right\rangle^2 \sim \frac{(n-1)(n^2-3n+3)}{n(n-2)(n-3)}V^2 - \frac{n}{(n-2)(n-3)}K'. \tag{29}$$

を得る．

2.3　不偏分散の分散の不偏推定量

式 (25) と式 (29) を使って式 (3) の各項を推定すれば，全体の不偏推定量となる．すなわち，

$$\left\langle (V-\sigma^2)^2 \right\rangle = \frac{1}{n}\left\langle (x-\langle x\rangle)^4 \right\rangle - \frac{n-3}{n(n-1)}\left\langle (x-\langle x\rangle)^2 \right\rangle^2 \tag{3}$$

$$\begin{aligned} &\sim \frac{1}{n}\left(\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)}K' - \frac{3(n-1)(2n-3)}{n(n-2)(n-3)}V^2\right) \\ &\quad - \frac{n-3}{n(n-1)}\left(\frac{(n-1)(n^2-3n+3)}{n(n-2)(n-3)}V^2 - \frac{n}{(n-2)(n-3)}K'\right) \end{aligned} \tag{30}$$

$$= \frac{1}{n(n-2)(n-3)}\left[n^2K' - (n^2-3)V^2\right] \tag{8}$$

3　注意点

式 (8) は不偏推定量であるが，特に n が小さい場合，負の値を返す場合がある．不偏分散の分散は正の値しか取らないため，ある意味で不合理である．$\left\langle (x-\langle x\rangle)^4 \right\rangle$ と $\left\langle (x-\langle x\rangle)^2 \right\rangle^2$ は，それぞれ K' や V^2 などを推定量として用いれば，推定量の非負性は保証される．しかし，$\left\langle (V-\sigma^2)^2 \right\rangle$ は本質的にこれらの差に関する量であるため，単純には非負性は保証されない．

3.1　必ず非負な推定量

x_j のみを取り除いたサイズ $n-1$ の標本 $\{x_i \mid 1 \le i \le n \wedge i \ne j\}$ の平均と不偏分散を次のように定義する．

$$m_{n,j} := \frac{1}{n-1}\sum_{i\ne j} x_i, \tag{31}$$

$$V_{n,j} := \frac{1}{n-2}\sum_{i\ne j}(x_i - m_{n,j})^2. \tag{32}$$

このとき，$V_{n,j}$ を変形すると，

$$V_{n,j} = \frac{1}{n-2}\left[\sum_{i\ne j}(x_j - m_n)^2 - (n-1)(m_{n,j}-m_n)^2\right] \tag{33}$$

$$= \frac{1}{n-2}\left[\sum_{i=1}^{n}(x_i - m_n)^2 - (x_j-m_n)^2 - (n-1)\left(\frac{nm_n - x_j}{n-1} - m_n)^2\right)\right] \tag{34}$$

$$= \frac{1}{n-2}\left[(n-1)V - (x_j-m_n)^2 - \frac{(m_n-x_j)^2}{n-1}\right] \tag{35}$$

$$= \frac{1}{n-2}\left[(n-1)V - \frac{n}{n-1}(x_j-m_n)^2\right]. \tag{36}$$

よって，j に関する $V_{n,j}$ の平均は，

$$\frac{1}{n}\sum_{j=1}^{n} V_{n,j} = \frac{1}{n(n-2)}\left[n(n-1)V - \frac{n}{n-1}(n-1)V\right] \tag{37}$$

$$= V \tag{38}$$

であり，$V_{n,j}$ の j に関する偏差の 2 乗和を求めると，

$$\sum_{j=1}^{n}(V_{n,j}-V)^2 = \sum_{j=1}^{n}\frac{1}{(n-2)^2}\left[V^2 - \frac{2n}{n-1}V(x_j-m_n)^2 + \frac{n^2}{(n-1)^2}(x_j-m_n)^4\right] \tag{39}$$

$$= \frac{1}{(n-2)^2}\left[nV^2 - 2nV^2 + \frac{n^3}{(n-1)^2}K'\right] \tag{40}$$

$$= \frac{n}{(n-2)^2}\left[\frac{n^2}{(n-1)^2}K' - V^2\right] \tag{41}$$

となる．これは，$n\to\infty$ において式 (6) に漸近することから，$\left\langle (V-\sigma^2)^2 \right\rangle$ の一致推定量である．また，定義より，有界の標本に対して，常に非負の値を返す．期待値は，式 (25)，(29) より，

$$\left\langle \frac{n}{(n-2)^2}\left[\frac{n^2}{(n-1)^2}K' - V^2\right]\right\rangle = \frac{1}{n-1}\left\langle (x-\langle x\rangle)^4\right\rangle - \frac{n-6}{(n-1)(n-2)}\left\langle (x-\langle x\rangle)^2\right\rangle^2 . \tag{42}$$

このことから，分布の尖度 $\kappa = \left\langle (x-\langle x\rangle)^4\right\rangle / \left\langle (x-\langle x\rangle)^2\right\rangle^2 - 3$ が分かっているならば，全体を

$$\frac{\dfrac{\kappa+3}{n} - \dfrac{n-3}{n(n-1)}}{\dfrac{\kappa+3}{n-1} - \dfrac{n-6}{(n-1)(n-2)}} = \frac{\dfrac{\kappa}{n} + \dfrac{2}{n-1}}{\dfrac{1}{n-1}\left(\kappa + \dfrac{2n}{n-2}\right)} \tag{43}$$

$$= \frac{(n-2)[(n-1)\kappa + 2n]}{n[(n-2)\kappa + 2n]} \tag{44}$$

倍すれば不偏推定量となる．すなわち，

$$\left\langle (V-\sigma^2)^2 \right\rangle \sim \frac{(n-1)\kappa + 2n}{(n-2)[(n-2)\kappa+2n]}\left[\frac{n^2}{(n-1)^2}K' - V^2\right]. \tag{45}$$

例えば正規分布ならば $\kappa = 0$ だから，これを代入して，

$$\left\langle (V-\sigma^2)^2 \right\rangle \sim \frac{1}{n-2}\left[\frac{n^2}{(n-1)^2}K' - V^2\right] \tag{46}$$

となる．非負性を気にする場合はこれを推定量として採用してもいいだろう．κ が真の値と違っていたとしても，不偏性は崩れるが，一致性は崩れない点は安心である．もっとも，(8) が負になるほど n が小さいのならば，どんなものにせよ推定量の分散自体が大きすぎて大して意味のある量にはならないだろうが……．

参考文献

[1] 井田隆: 標本分散の誤差の評価, `http://www.crl.nitech.ac.jp/~ida/research/memo/SampleVariance/sample_variance.pdf`, (2011).