# 正規分布の標準偏差に対する簡便不偏推定法

統計数理研究所　鈴　木　義 一 郎

(1979年4月　受付)

## Note on Unbiased Estimation of the Standard Deviation of Normal Distribution

Giitiro Suzuki
(The Institute of Statistical Mathematics)

Unbiased estimation of the standard deviation of the normal distribution based on the sample standard deviation

$$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}\,,\qquad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

has been thoroughly explored in a series of articles [1]～[8].

In this note, we propose the estimate

$$S_0{}^*=\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/g_n}\,,$$

$$g_n=n-\frac{3}{2}+\frac{n+1}{4(n-1)(2n+1)}\,,$$

which is a simple and very accurate approximation to the exactly unbiased estimate

$$S_0=\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/k_n}\,,$$

$$k_n=2\left[\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right]^2 .$$

Gurland and Tripathi [4] derived another approximation

$$S_0{}^{**}=\sqrt{\Sigma(X_i-\bar{X})^2/t_n}\,,$$

$$t_n=n-\frac{3}{2}+\frac{1}{8(n-1)}\,,$$

using the asymptotic expansion for log $k_n$. Our result is based on rather heuristic method. From Figure 1 and Table 1, it will be seen that our result is more accurate than [4].

## 1. はじめに

最近，いろんな関数キーのついている電卓が安価に手に入るようになった．そのような関数キーの一つに，偏差平方和をデータ数 $n$ で割ってから平方根をとったもの，$n-1$ で割って平方根をとったものがある．しかも前者に標本標準偏差，後者には母標準偏差という呼称がつけられたりしている．正しい統計の知識が身につかないまゝ計算法だけが先行している昨今の風潮を，垣間みた感じがする．

ともあれデータ数 $n$ が大きければ両者の値がそう違わないが，$n$ の小さいときが問題である．

そもそも偏差平方和を $n-1$ で割るのは，母分散の不偏推定を得るためだった．それの平方根をとってみても，分散の平方根である標準偏差に対して不偏にはならない．母集団に正規性を仮定できる場合には，偏差平方和を $n-1.5$ で割ってから平方根をとると，不偏推定量により近いものが得られる．いずれにせよ標準偏差の推定を行う目的の場合は，$n-1$ で割る根拠は無いといってよい．それなら単純にデータ数 $n$ で割るほうがすっきりしている．

正規性の仮定の下では，標準偏差の不偏推定を得るための修正係数がガンマ関数を用いて表現できる．この小稿では，そのような係数に対する単純な有理式を用いた近似式(3)を提示する．

## 2.　標準偏差の不偏推定に対する修正係数

$X_1$, $X_2$, ･･, $X_n$ を平均 $\mu$，標準偏差 $\sigma$ の正規分布からの標本とする．

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

と置いて

$$V_0=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

を考えれば，これは分散 $\sigma^2$ に対する不偏推定量となるが，$\sqrt{V_0}$ は標準偏差 $\sigma$ の不偏推定ではない．$\mu$ が未知の場合

$$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

が $\sigma$ の最尤推定量であるが，無論不偏ではない．そこで

$$a_n=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\frac{n}{2}}$$

$$a_n'=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\frac{n-1}{2}}=a_n\sqrt{\frac{n-1}{n}}$$

と置いて

$$S_0=a_nS=a_n\sqrt{V_0}$$

を考えると，これが $\sigma$ の不偏推定量で分散は

$$\mathrm{V}\{S_0\}=\left[a_n^2\frac{n-1}{n}-1\right]\sigma^2$$

$$=\frac{\sigma^2}{2n}+O\left(\frac{1}{n^2}\right)$$

で与えられる．

$\sigma$ に対する不偏推定量を得るための修正係数に関する簡単な歴史を，Jarrett [6] が報じている．Pearson [7] は $1/a_n$ に対する4桁の表を，Treloar [8] は $a_n$ に対する5桁の表を与えた．これらの表の存在に気づかずに，Holtzman [5] は $a_n'$ に対する表を，Cureton [3] は $k_n=n/a_n^2=(n-1)/a_n'^2$ に対する表を，さらに Bolch [1] も $a_n$ 及び $a_n'$ に対する表を与えた．

修正係数に対して，ガンマ関数を用いない近似式を最初に与えたのが Brugger [2] である．Cureton の与えた表からもすぐ見当つくことであるが

$$a_n = \sqrt{n/(n-1.5)}$$
$$a_n' = \sqrt{(n-1)/(n-1.5)}$$

で算出したものが，$n$ が6以上なら結構よい近似を与える．Gurland-Tripathi [4] は

$$t_n = n - \frac{3}{2} + \frac{1}{8(n-1)} \tag{1}$$

が $k_n(=n/a_n^2)$ に対するかなり良い近似を与えることを示した．この $t_n$ は

$$\log k_n = \log(n-1) - \frac{1}{2(n-1)} + \frac{1}{12(n-1)^3} + \cdots\cdot$$

の展開式の指数部の第2項までをとって得られたものである．

ところが図1からみてもわかるように，(1) で与えられる $t_n$ の値は正確な $k_n$ の値より一様に小さい．そこで

$$k_n = n - \frac{3}{2} + \frac{1}{n-1}\cdot\frac{1}{8-\dfrac{1}{\delta_n}} \tag{2}$$

の形の近似を考えてみる．そのために

$$\delta_n = \left\{8 - \frac{1}{(n-1)\left(k_n - n + \dfrac{3}{2}\right)}\right\}^{-1}$$

をプロットしたのが図2である．ほぼ直線的傾向を示すから，最小二乗直線を求めると

$$(0.5378) + (0.2556)n$$

が得られる（図2の実線）．ところがこの直線，$n$=3〜16 の辺で一様に上すぎる．$n$ の値の小さなところでのずれのほうがむしろ深刻であるから（係数も多少丸めて）

$$(0.25) + (0.25)n$$

という直線（図2の点線）を用いたほうが良さそうである．これを (2) 式の $\delta_n$ に代入して

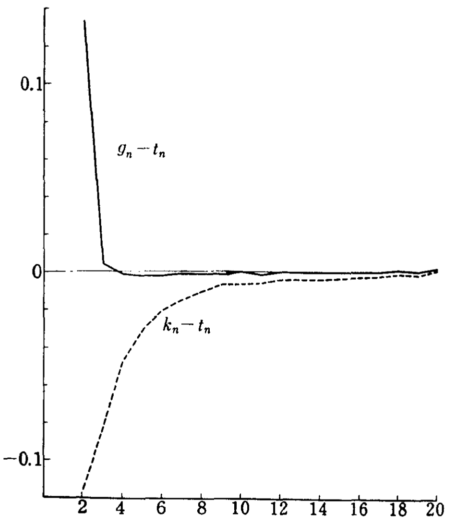


図　1

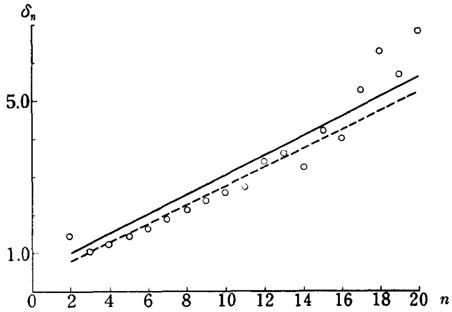


図　2

表 1

| $n$ | $k_n$ | $t_n$ | $g_n$ |
|---|---|---|---|
| 2 | 0.6366 | 0.6250 | 0.6500 |
| 3 | 1.5708 | 1.5625 | 1.5714 |
| 4 | 2.5464 | 2.5417 | 2.5463 |
| 5 | 3.5343 | 3.5312 | 3.5341 |
| 6 | 4.5271 | 4.5250 | 4.5269 |
| 7 | 5.5223 | 5.5208 | 5.5222 |
| 8 | 6.5190 | 6.5179 | 6.5189 |
| 9 | 7.5165 | 7.5159 | 7.5164 |
| 10 | 8.5146 | 8.5139 | 8.5146 |
| 11 | 9.5131 | 9.5125 | 9.5130 |
| 12 | 10.5118 | 10.5114 | 10.5118 |
| 13 | 11.5108 | 11.5104 | 11.5108 |
| 14 | 12.5100 | 12.5096 | 12.5099 |
| 15 | 13.5092 | 13.5089 | 13.5092 |
| 16 | 14.5086 | 14.5083 | 14.5086 |
| 17 | 15.5080 | 15.5078 | 15.5080 |
| 18 | 16.5075 | 16.5074 | 16.5076 |
| 19 | 17.5071 | 17.5069 | 17.5071 |
| 20 | 18.5065 | 18.5066 | 18.5067 |

(3) $$g_n = n - \frac{3}{2} + \frac{n+1}{4(n-1)(2n+1)}$$

という近似式が得られる．これは図1からみても判るように，$n$=2 以外では驚くほど精度が良い．表1に，正確な $k_n$ の値と2つの近似式の値とを掲げておいた．

## 参考文献

[1] Bolch, B.W. (1968) More on unbiased estimation of the standard deviation, *Amer. Statistician*, **22** (3), 27.

[2] Brugger, R.M. (1969) A note on unbiased estimation of the standard deviation, *Amer. Statistician*, **23** (4), 32.

[3] Cureton, E.E. (1968) Unbiased estimation of the standard deviation, *Amer. Statistician*, **22** (1), 22.

[4] Gurland, J. and Tripathi, R.C. (1971) A simple approximation for unbiased estimation of the standard deviation., *Amer. Statistician*, **25** (4), 30–32.

[5] Holtzman, W.H. (1950) The unbiased estimate of the population variance and standard deviation, *Amer. J. Psychology*, **63**, 615–617.

[6] Jarrett, A.E. (1968) A minor exercise in history, *Amer. Statistician*, **22** (3), 25–26.

[7] Pearson, K. (1915) *The Distribution of the Standard Deviation of Small Samples*, Appendix I to papers by "Student" and R.A. Fisher.

[8] Treloar, A.E. (1935) *Outline of Biometric Analysis*, Burgess Publishiug Co., Minneapolis.