（ＪＡＰＬＡ 2010/8/5-7)

『標本分散』と『不偏分散』

統計数理研究所（名誉教授）　鈴木義一郎

初心者に統計の講義をしていて、よく質問をされるのに、「分散を推定するのに、平均からの偏差平方値をデータ数ｎでなくｎ－１で割るのは何故か？」という問題がある。

小生のさしあたっての答えは「どちらでもよい」である。データ数ｎが大きければ大差ないし、またｎが小さければ、分散を推定しようとすることなどずーずーしいのである。実は区間推定などに必要なのは分散ではなく標準偏差の方である。ｎ－１で割って平方根をとっても標準偏差の不偏推定ではない！だから、「ｎやｎ－１ではなく、ｎ－1.5 で割り平方根をとって推定値とせよ」と推奨している。

$X_1, X_2, \cdots\cdots\cdots, X_n$ を、平均が μ 、分散が σ^2 の分布からの標本とするとき

標本平均：$M = \dfrac{X_1 + X_2 + \cdots\cdots\cdots + X_n}{n}$

偏差平方和：$Q = (X_1 - M)^2 + (X_2 - M)^2 + \cdots\cdots\cdots + (X_n - M)^2$

標本分散と不偏分散：$V = Q/n \quad , \quad s^2 = Q/(n-1)$

に対して

$$E\{M\} = \mu \ , E\{Q\} = (n-1)\sigma^2 \ , E\{s^2\} = \sigma^2$$

である。ところで

$$E\{s\} > \sigma$$

であるから、$s = \sqrt{s^2}$ はσの不偏推定ではない。ただし正規分布 $N(\mu, \sigma^2)$ の場合には

$$s_0 = \sqrt{s_0^{\,2}} \quad ; s_0^{\,2} = Q/k(n) \, , \, k(n) = 2\left\{\Gamma\left(\frac{n}{2}\right) / \Gamma\left(\frac{n-1}{2}\right)\right\}^2$$

がσの不偏推定となる。

```
    kn=:3 :'+:*:(!&<:-:y)%!&<:-:y-1'
1      2      3      4      5
0 0.63662 1.5708 2.54648 3.53429
    c,:kn c=:6+i.5
        6      7      8      9      10
4.52707 5.52233 6.51899 7.51651 8.51459
```

s_0が正規分布の標準偏差の不偏推定になることを、ノーマル・チップの実験を用いて確かめてみよう。

n_tip=:(M=:50+i:5),:f=:n,20,|.n=:1 3 7 12 17	100個のノーマル・チップ

i:5 _5 _4 _3 _2 _1 0 1 2 3 4 5	]M=:50+i:5 45 46 47 48 49 50 51 52 53 54 55
]f=.n,20,|.n=:1 3 7 12 17 1 3 7 12 17 20 17 12 7 3 1 <:+/¥f 0 3 10 22 39 59 76 88 95 98 99 (}:,:}.)<:+/¥f 0 3 10 22 39 59 76 88 95 98 3 10 22 39 59 76 88 95 98 99	(0 3 belong k)#k=:i.100 1 2 3 (3 10 belong k)#k 4 5 6 7 8 9 10 (10 22 belong k)#k 11 12 13 14 15 16 17 18 19 20 21 22 …………………………
belong=.(]>{.@[)*.]<:{:@[change=:3 :0 (M=.50+i:5),:f=.n,20,|.n=:1 3 7 12 17 L=.(}:,.}.)<:+/\f belong=.(]>{.@[)*.]<:{:@[if.y=0 do.{.M else.(L belong"1 y)#}.M end.)	(88 95 belong k)#k 89 90 91 92 93 94 95 (95 98 belong k)#k 96 97 98 (98 99 belong k)#k 99
n_tip 45 46 47 48 49 50 51 52 53 54 55 1 3 7 12 17 20 17 12 7 3 1 10 10$,change"0 i.100 45 46 46 46 47 47 47 47 47 47 47 48 48 48 48 48 48 48 48 48 48 48 48 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 52 52 52 52 52 52 52 52 52 52 52 52 53 53 53 53 53 53 53 54 54 54 55	(,.M);' *'{~0<f#"0 M =:45+i:5 45 | * 46 | *** 47 | ******* 48 | ************ 49 | ***************** 50 | ******************** 51 | ***************** 52 | ************ 53 | ******* 54 | *** 55 | *

sample=:[:,[:change"0[:/:~?@$&100	ノーマル・チップからの標本抽出
qsum=:[:+/*:@(-mean=:+/%#) dev=:3 :'v,%:v=:(qsum y)%5 4'	「qsum」は(平均)偏差平方和 「dev」は標本分散、不偏分散、それらの平方根

```
   ]S=:sample 5
48 49 50 50 53
   ]M=:(mean=:+/%#)S
50
   +/*:S-M
14
   qsum S
14
   dev S
2.8 3.5 1.67332 1.87083
```

```
   S;5j2":dev"1 S=:sample"0(5$5)
```

45 50 50 50 50	4.00 5.00 2.00 2.24
46 49 50 52 54	7.36 9.20 2.71 3.03
48 48 51 51 52	2.80 3.50 1.67 1.87
48 48 50 52 53	4.16 5.20 2.04 2.28
45 48 48 50 52	5.44 6.80 2.33 2.61

「5j2": y」は右引数の数値をスペースを５にとり小数点２桁目で四捨五入した値(文字！)を表示。

```
   A;5j2":dev"1 A=:sample"0(10$5)
```

48 49 50 50 53	2.80 3.50 1.67 1.87
47 49 50 50 51	1.84 2.30 1.36 1.52
47 50 50 50 51	1.84 2.30 1.36 1.52
49 50 50 50 52	0.96 1.20 0.98 1.10
47 49 49 51 54	5.60 7.00 2.37 2.65
47 48 50 52 52	4.16 5.20 2.04 2.28
49 50 51 53 53	2.56 3.20 1.60 1.79
48 49 50 51 53	2.96 3.70 1.72 1.92
46 48 49 49 50	1.84 2.30 1.36 1.52
48 50 50 50 53	2.56 3.20 1.60 1.79

‘ ’

```
   B;5j2":dev"1 B=:sample"0(10$5)
```

48 48 49 51 53	3.76 4.70 1.94 2.17
49 49 50 51 53	2.24 2.80 1.50 1.67
49 50 50 51 53	1.84 2.30 1.36 1.52
47 47 48 49 52	3.44 4.30 1.85 2.07
47 48 50 50 51	2.16 2.70 1.47 1.64
47 50 52 54 54	7.04 8.80 2.65 2.97
47 51 52 52 53	4.40 5.50 2.10 2.35
47 50 50 50 52	2.56 3.20 1.60 1.79
48 50 51 52 53	2.96 3.70 1.72 1.92
48 49 50 50 51	1.04 1.30 1.02 1.14

C;5j2":dev"1 C=:sample"0(10$5)

45 47 49 51 51	5.44 6.80 2.33 2.61
48 49 50 52 52	2.56 3.20 1.60 1.79
48 49 50 50 51	1.04 1.30 1.02 1.14
48 49 50 51 52	2.00 2.50 1.41 1.58
46 47 49 50 51	3.44 4.30 1.85 2.07
46 47 50 50 52	4.80 6.00 2.19 2.45
49 49 49 51 53	2.56 3.20 1.60 1.79
50 50 51 51 52	0.56 0.70 0.75 0.84
47 48 48 49 52	2.96 3.70 1.72 1.92
47 49 51 51 51	2.56 3.20 1.60 1.79

‘ ’

D;5j2":dev"1 D=:sample"0(10$5)

50 51 51 52 54	1.84 2.30 1.36 1.52
47 48 49 49 51	1.76 2.20 1.33 1.48
46 49 50 50 51	2.96 3.70 1.72 1.92
50 51 51 52 53	1.04 1.30 1.02 1.14
48 49 49 50 51	1.04 1.30 1.02 1.14
49 50 50 51 53	1.84 2.30 1.36 1.52
49 49 49 51 51	0.96 1.20 0.98 1.10
50 51 52 53 54	2.00 2.50 1.41 1.58
47 48 51 51 55	7.84 9.80 2.80 3.13
50 51 52 52 54	1.76 2.20 1.33 1.48

```
qsum=:[:+/*:@(-mean=:+/%#)
dif=:3 :'|2.00499-%:(qsum y)%5 4 3.53429'
order=:3 :'/:dif y'
```

<table>
<tr><td>S
48 49 50 50 53
]Q=:qsum S
14
]s=:%:Q%5 4 3.53429
1.67332 1.87083 1.99027</td>
<td>]d=:|2.00499-s
0.33167 0.134161 0.0147157
dif S
0.33167 0.134161 0.0147157
/:d
2 1 0</td>
<td>order S
2 1 0

「/:d」はｄを昇順に並べたときのインデクスを返す。
ｄの指標２(３番目)が最小であることが分かる。</td></tr>
</table>

order"1 A	order"1 B	order"1 C	order"1 D
2 1 0	0 1 2	0 1 2	2 1 0
2 1 0	2 1 0	2 1 0	2 1 0
2 1 0	2 1 0	2 1 0	2 1 0
2 1 0	1 0 2	2 1 0	2 1 0
0 1 2	2 1 0	1 0 2	2 1 0
0 1 2	0 1 2	0 1 2	2 1 0
2 1 0	0 1 2	2 1 0	2 1 0
2 1 0	2 1 0	2 1 0	2 1 0
2 1 0	2 1 0	2 1 0	0 1 2
2 1 0	2 1 0	2 1 0	2 1 0

（半数以上のケースで、Ｑを 3.53429 で割った推定値の誤差が最小である！）

```
normal=:3 :'_6+(+/?(12,y)$1000)%1000'
k n=:3 :'+:*:(!&<:-:y)%!&<:-:y-1'
dif1=:3 :'|1-%:(qsum y)%5 4 3.53429'
order1=:3 :'/:dif1 y'
```

12 個の$(0,1)$上の一様乱数の和の平均値は６で分散が１である。これから６を引いたもの

は標準正規分布に極めて近い。
また、U,V を２組の独立な一様乱数として

$$X = \sqrt{-2\log(U)}\cos(V)\ ,\ Y = \sqrt{-2\log(U)}\sin(V)$$

と定義すれば、これら２つの変数が独立な正規乱数となることも知られている。

<table>
<tr><td colspan="2">]N1=:normal"1(10, 5)
1.105 0.513 _0.846 0.328 _0.275
_1.112 _0.284 _0.221 _0.949 _0.556
0.421 0.242 0.852 0.404 1.127
0.366 0.026 1.028 1.321 _0.249
1.201 _2.261 _0.63 _1.41 1.383
1.276 0.893 0.95 0.605 _0.678
1.994 _0.674 0.656 _0.083 0.853
_0.662 1.574 2.023 0.273 _0.547
0.765 _0.613 0.828 0.425 2.02
_2.343 0.249 1.887 0.468 _0.423</td><td>]N2=:normal"1(10,5)
0.755 1.54 _0.166 _1.56 _1.208
_0.945 0.932 0.538 0.922 _1.875
1.332 0.178 0.086 _0.336 0.341
0.135 _1.142 1.038 1.55 _0.131
_0.378 0.211 0.28 _0.74 0.047
0.836 _0.281 1.425 _0.843 _1.018
0.354 _1.214 _0.702 _0.218 0.666
0.653 _0.83 1.288 0.678 _0.67
1.11 _0.613 0.235 _1.327 _0.176
1.1 _0.357 _1.594 0.733 _1.319</td></tr>
<tr><td rowspan="3">order1”1 N1
2 1 0
2 1 0
2 1 0
2 1 0
0 1 2
2 1 0
1 2 0
0 1 2
2 1 0
0 1 2</td><td rowspan="3">order1”1 N2
0 1 2
0 1 2
2 1 0
1 0 2
2 1 0
0 1 2
2 1 0
2 1 0
2 1 0
0 1 2</td><td>|:order1”1 normal"1(10,5)
2 0 0 2 0 2 0 2 2 0
1 1 1 1 1 1 1 1 1 1
0 2 2 0 2 0 2 0 0 2</td></tr>
<tr><td>|:order1”1 normal"1(10,5)
2 0 0 2 2 0 2 0 2 2
1 1 1 1 1 1 1 1 1 1
0 2 2 0 0 2 0 2 0 0</td></tr>
<tr><td>|:order1”1 normal"1(10,5)
2 0 2 0 2 2 2 2 0 0
1 1 1 1 1 1 1 1 1 1
0 2 0 2 0 0 0 0 2 2</td></tr>
</table>