正規母集団からとられた標本に於ける正又は負の偏差の和の分布に就いて

塩　谷　　実

(1954 年 4 月受付)

On the Distribution of the Sum of the Positive or Negative Deviations from the Mean in the Sample drawn from the Normal Population.

Minoru SHIOTANI

ABSTRACT: Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be an ordered sample from a normal population with mean zero and variance σ^2. Let $\bar{\bar{x}}$ be the mean of the observations smaller than the sample mean $\bar{x}$ and $\bar{\bar{\bar{x}}}$ the mean of the observations larger than $\bar{x}$. In this paper we consider the statistics based on $U_n = \bar{\bar{\bar{x}}} - \bar{\bar{x}}$ and $T_n = \bar{x} - \bar{\bar{x}}$ in order to estimate the population standard deviation σ. Their sampling distribution can be obtained as follows: We, first, obtaine the conditional distributions of U_n and T_n under the condition which the number of the observatisns smaller than $\bar{x}$ is ν and, then, take the average of them with respect to ν. Conditional statistics $U_n(\nu)$ and $T_n(\nu)$ are associated with the sum of the negative deviations from the sample mean, $\sum_{i=1}^{\nu}(x_i - \bar{x})$, as shown below. The author has obtained the sampling distribution of $\sum_{1}^{\nu}(x_i - \bar{x})$ and, from it, the distribution of $T_n(\nu)$, $U_n(\nu)$, T_n and U_n are obtained easily. Moreover, the efficiencies, in using the statistics U_n, T_n to estimate the population standard deviation σ, have calculated and tabulated in Table 3 and 4 with the efficiencies of range and mean deviation. Also, the coefficients of unbiasedness of our estimation are tabulated in Table 2 and 4.

Institute of Statistical Mathematics

1. 緒言. 平均 m, 分散 σ^2 を持つ正規母集団から取られた大さ n の nandom sample を x_1', x_2', $\cdots$, x_n' とする. 標本平均

(1)
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i'$$

からの偏差 $x_i' - \bar{x}$, $i=1, \cdots, n$ のうち正のものを例へば $n-\nu$ 箇, 負のものを ν 箇とし, 夫々

$$x_1'' - \bar{x},\ x_2'' - \bar{x},\ \cdots\cdots,\ x''_{n-\nu} - \bar{x},$$
$$x_1''' - \bar{x},\ x_2''' - \bar{x},\ \cdots\cdots,\ x_\nu''' - \bar{x}$$

で表はす. 我々は此等の偏差の和

(2)
$$\sum_{i=1}^{n-\nu}(x_i'' - \bar{x})\ ;\quad \sum_{i=1}^{\nu}(x_i''' - \bar{x})$$

に関係する統計量の分布に就いて考へよう．標本平均より小さい標本値の平均 $\bar{\bar{x}}$ を計算し，母集団標準偏差 σ の推定に

$$T_n = \bar{x} - \bar{\bar{x}} \tag{3}$$

を用ひる事が考へられる．更に $\bar{x}$ より大きい標本値の平均を一般に $\bar{\bar{\bar{x}}}$ で表はせば

$$U_n = \bar{\bar{\bar{x}}} - \bar{\bar{x}} \tag{4}$$

を σ の推定に用ひる事も考へられる．(3),(4) に於ける統計量の分布を求めるのに次の様に考へる．$\bar{x}$より小さい標本値が ν 箇である時の (3),(4) を夫々

$$T_n(\nu) = \bar{x} - \bar{\bar{x}}(\nu) \tag{5}$$

$$U_n(\nu) = \bar{\bar{\bar{x}}}(n-\nu) - \bar{\bar{x}}(\nu) \tag{6}$$

で表はせば，我々は先づ $T_n(\nu), U_n(\nu)$ の標本分布を求め，然る後 ν に就いて平均化することに依り T_n, U_n の分布を求める事が出来る．所で $T_n(\nu), U_n(\nu)$ を書き直すと

$$T_n(\nu) = -\frac{1}{\nu}\sum_{i=1}^{\nu}(x_i''' - \bar{x}) \tag{5'}$$

$$U_n(\nu) = -\frac{n}{n-\nu}\frac{1}{\nu}\sum_{i=1}^{\nu}(x_i''' - \bar{x}) \tag{6'}$$

であるから結局 $\sum_{i=1}^{\nu}(x_i''' - \bar{x})$ の分布が求まればよいことがわかる．亦平均偏差

$$W_n = \frac{1}{n}\sum_{i=1}^{n}|x_i' - \bar{x}| \tag{7}$$

の分布も上述の偏差の和の分布を求める事に帰着する．偏差の負であるものが ν 箇の時の (7) は

$$\begin{aligned} W_n(\nu) &= \frac{1}{n}\sum_{i=1}^{\nu}(\bar{x} - x_i''') + \frac{1}{n}\sum_{i=1}^{n-\nu}(x_i'' - \bar{x}) \\ &= \frac{2}{n}\sum_{i=1}^{n-\nu}(x_i'' - \bar{x}) \\ &= -\frac{2}{n}\sum_{i=1}^{\nu}(x_i''' - \bar{x}) \end{aligned} \tag{7'}$$

と書けるからである．W_n の分布は既に H. J. Godwin に依り求められて居る．[1] 筆者は T_n, U_n の分布と同時に W_n の分布も H. Godwin とは独立に負又は正の偏差の和に着目して求めたので紹介する．亦此等の分布を利用して T_n, U_n を σ の推定に用ひる時の効率を調べる．更に $\bar{x}$ より小さい標本値が ν 箇である確率，$P[N=\nu]$，の値を知つて居れば，母集団が正規分布である時の或種の randomness の検定に役に立つから，$P[N=\nu]$ の数表を与へる．

2. 負の偏差の和の分布

母集団平均を $m=0$ としても一般性を失はない．at random にとられた n 箇の標本値を大さの順に並べて $x_1 \leqq x_2 \leqq \cdots\cdots \leqq x_n$ とする．本節に於て負の偏差の和の分布を求めるが，正の偏差の和の分布も正規分布の対称性から全く同じである．即ち $\bar{x}$ より小さい標本値が ν 箇である時の

$$\sum_{i=1}^{\nu}(x_i - \bar{x})$$

の分布を求める．N で $\bar{x}$ より小さい標本値の箇数を表はす確率変数とすれば，$N=\nu$ といふ条件の下に於ける $x_1, x_2 \cdots\cdots x_n$ の同時分布は

$$(8)\qquad f(x_1,\cdots\cdots,x_n|\nu)=\frac{1}{P[N=\nu]}n!\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right]$$

$$(x_1\leqq x_2\leqq\cdots\cdots\leqq\bar{x}\leqq x_{\nu+1}\leqq\cdots\cdots\leqq x_n)$$

である．此れより x_n を $x_n=n\bar{x}-x_1-\cdots\cdots-x_{n-1}$ に依り $\bar{x}$ に移し，更に

$$(9)\qquad \begin{cases} y_i=x_i-\bar{x}, & (i=1,\cdots\cdots n-1)\\ y_n=\bar{x} \end{cases}$$

に依り変換し，y_n を integrate out すれば

$$(10)\qquad f(y_1,\cdots\cdots,y_{n-1}|\nu)=\frac{1}{P[N=\nu]}\sqrt{n}\,n!\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{n-1}\exp\left[-\frac{1}{2\sigma^2}\left\{y_1^2+\cdots\cdots+y^2_{n-1}+(y_1+\cdots\cdots+y_{n-1})^2\right\}\right]$$

となり変域は

$$(11)\qquad D_1\begin{cases} y_{n-2}\leqq y_{n-1}\leqq-\frac{1}{2}(y_1+\cdots\cdots+y_{n-2})\\ y_{n-3}\leqq y_{n-2}\leqq-\frac{1}{3}(y_1+\cdots\cdots+y_{n-3})\\ \qquad\vdots\\ y_{\nu+1}\leqq y_{\nu+2}\leqq-\frac{1}{n-\nu-1}(y_1+\cdots\cdots+y_{\nu+1})\\ 0\leqq y_{\nu+1}\leqq-\frac{1}{n-\nu}(y_1+\cdots\cdots+y_\nu)\\ y_{\nu-1}\leqq y_\nu\leqq 0\\ \qquad\vdots\\ y_1\leqq y_2\leqq 0\\ -\infty<y_1\leqq 0 \end{cases}$$

となる．更に $y_1,\cdots\cdots,y_{n-1}$ を次の変換に依り $\xi_1,\cdots\cdots,\xi_{n-1}$ に移す．

$$(12)\qquad \begin{cases} \sqrt{2\cdot1}\;\sigma\xi_1=-y_1+y_2\\ \sqrt{3\cdot2}\;\sigma\xi_2=-y_1-y_2+2y_3\\ \qquad\vdots\\ \sqrt{\nu(\nu-1)}\;\sigma\xi_{\nu-1}=-y_1-y_2-\cdots\cdots-y_{\nu+1}+(\nu-1)y_\nu\\ \sqrt{\nu}\;\sigma\xi_\nu=-y_1-y_2-\cdots\cdots\cdots-y_\nu\\ \sqrt{\frac{n-\nu-1}{n-\nu}}\;\sigma\xi_{\nu+1}=-y_{\nu+1}-\frac{y_1+\cdots\cdots+y_\nu}{n-\nu}\\ \sqrt{\frac{n-\nu-2}{n-\nu-1}}\;\sigma\xi_{\nu+2}=-y_{\nu+2}-\frac{y_1+\cdots\cdots+y_{\nu+1}}{n-\nu-1}\\ \qquad\vdots\\ \sqrt{\frac{2}{3}}\;\sigma\xi_{n-2}=-y_{n-2}-\frac{y_1+\cdots\cdots+y_{\nu+2}}{3}\\ \sqrt{\frac{1}{2}}\;\sigma\xi_{n-1}=-y_{n-1}-\frac{y_1+\cdots\cdots+y_{n-2}}{2} \end{cases}$$

此の変換の Jacobian は $(-1)^{n-1}\sigma^{-(n-1)}\sqrt{n-\nu}$ であり．且つ

$$y_1^2+\cdots\cdot y_\nu^2=\sigma^2(\xi_1^2+\cdots+\xi_\nu^2)$$

y_{n-1}

$$(13)\quad y_1^2+\cdots\cdot+y^2_{n-1}+(y_1+\cdots\cdot+y_{n+1})^2=y_1^2+\cdots\cdot+y_\nu^2+\frac{1}{n-\nu}(y_1+\cdots\cdot+y_\nu)^2$$

$$+\frac{n-\nu-1}{n-\nu}\left(y_{\nu+1}+\frac{y_1+\cdots\cdot+y_\nu}{n-\nu}\right)^2+\frac{n-\nu-1}{n-\nu-2}\left(y_{\nu+2}+\frac{y_1+\cdots\cdot+y_{\nu+1}}{n-\nu-1}\right)^2$$

$$+\cdots\cdots\cdots$$

$$+\frac{3}{2}\left(y_{n-2}+\frac{y_1+\cdots\cdot+y_{n-3}}{3}\right)^2+\frac{2}{1}\left(y_{n-1}+\frac{y_1+\cdots\cdot+y_{n-2}}{2}\right)^2$$

であるから変換後の $\xi_1,\cdots\cdot\xi_{n-1}$ の同時分布は

$$(14)\quad f(\xi_1,\cdots,\xi_{n-1}|\nu)=\frac{1}{P[N=\nu]}\left(\frac{n}{n-\nu}\right)^{\frac{1}{2}}n!\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1}$$

$\xi^2_{\nu+1}$

$$\exp\left[-\frac{1}{2}\left\{\xi_1^2+\cdots\cdot+\xi^2_{\nu-1}+\xi^2_{\nu-1}+\cdots\cdot+\xi^2_{n-1}+\frac{n}{n-\nu}\xi_\nu^2\right\}\right]$$

となる．変域は D_1 及び (12) より

$$(15)\quad D_2\begin{cases}0\leqq\xi_{n-i}\leqq\sqrt{\dfrac{i+2}{i}}\,\xi_{n-i-1} & (i=1,\cdots,n-\nu-2)\\ 0\leqq\xi_{\nu+1}\leqq\sqrt{\dfrac{\nu}{(n-\nu)(n-\nu-1)}}\,\xi_\nu & \\ 0\leqq\xi_\nu<+\infty & \\ 0\leqq\xi_{\nu-1}\leqq\dfrac{1}{\sqrt{\nu-1}}\,\xi_\nu & \\ 0\leqq\xi_j\leqq\sqrt{\dfrac{j+2}{j}}\,\xi_{j+1} & (j=\nu-2,\ \nu-3\cdots,1)\end{cases}$$

となる．負の偏差の和の分布は (14) に於て $\xi_1,\cdots,\xi_{\nu-1},\xi_{\nu+1},\cdots,\xi_{n-1}$ を消去して ξ_ν の密度函数を求めればよい．今

$$(16)\quad Q_{\nu-1}(x)=\nu!\left(\frac{1}{\sqrt{2\pi}}\right)^{\nu-1}\int_0^x\int_0^{\sqrt{\frac{\nu}{\nu-2}}\xi_{\nu-1}}\cdots\cdots\int_0^{\sqrt{\frac{4}{2}}\xi_2}\int_0^{\sqrt{\frac{3}{1}}\xi_2}$$

$$\exp\left[-\frac{1}{2}(\xi_1^2+\xi_2+\cdots\cdot+\xi^2_{\nu-1})\right]d\xi_1d\xi_2\cdots\cdot d\xi_{\nu-1}$$

$$(17)\quad Q_{n-\nu-1}(x)=(n-\nu)!\left(\frac{1}{\sqrt{2\pi}}\right)^{n-\nu-1}\int_0^x\int_0^{\sqrt{\frac{n-\nu}{n-\nu-2}}\xi_{\nu+1}}\cdots\cdots\cdots\int_0^{\sqrt{\frac{4}{2}}\xi_{n-1}}\int_0^{\sqrt{\frac{3}{1}}\xi_{n-2}}$$

$$\exp\left[-\frac{1}{2}(\xi^2_{n-1}+\xi^2_{n-2}+\cdots\cdot+\xi^2_{\nu+1})\right]d\xi_{n-1}d\xi_{n-2}\cdot\cdot d\xi_{\nu+1}$$

とおけば ξ_ν の密度函数は

$$(18)\quad f(\xi_\nu)=\frac{1}{P[N=\nu]}\binom{n}{\nu}\frac{1}{\sqrt{2\pi}}\left(\frac{n}{n-\nu}\right)^{\frac{1}{2}}Q_{\nu-1}\left(\frac{1}{\sqrt{\nu-1}}\xi_\nu\right)\cdot Q_{n-\nu-1}\left(\sqrt{\frac{\nu}{(n-\nu)(n-\nu-1)}}\xi_\nu\right)$$

$$\cdot\exp\left[-\frac{n}{n-\nu}\xi^2\right]$$

$\frac{1}{2}\frac{\nu}{n-\nu}\xi_\nu^2$

となる．此処に定義した$Q_r(x)$は Godwin の定義した函数 $G_r(x)$ と次の関係があることは容易に分る．[1] 即ち

$$
(19)\quad \begin{cases} Q_{\nu-1}(x) = \sqrt{\nu}\left(\dfrac{1}{\sqrt{2\pi}}\right)^{\nu-1} G_{\nu-1}(\sqrt{\nu(\nu-1)}\,x) \\ Q_{n-\nu-1}(x) = \sqrt{n-\nu}\left(\dfrac{1}{\sqrt{2\pi}}\right)^{n-\nu-1} G_{n-\nu-1}(\sqrt{(n-\nu-1)(n-\nu)}\,x) \end{cases}
$$

亦 F. E. Grubbs が $(x_n-\bar{x})/\sigma$ の分布函数として求めた $F_n(x)$ との間には

$$
(20)\quad \begin{cases} Q_{\nu-1}(x) = F_\nu\left(\sqrt{\dfrac{\nu-1}{\nu}}\,x\right) \\ Q_{n-\nu-1}(x) = F_{n-\nu}\left(\sqrt{\dfrac{n-\nu-1}{n-\nu}}\,x\right) \end{cases}
$$

の関係がある．[2]

(18) に於ける ξ_ν の分布より統計量 $T_n(\nu)$, $U_n(\nu)$, $W_n(\nu)$ の分布は容易に求まり夫々次の様になる．

$$
(21)\quad f\{T_n(\gamma)\} = \frac{1}{P[N=\nu]}\binom{n}{\nu}\frac{1}{\sqrt{2\pi}\sigma}\left(\frac{n\nu}{n-\nu}\right)^{\frac{1}{2}}\cdot Q_{\nu-1}\left(\sqrt{\frac{\nu}{\nu-1}}\frac{T_n(\nu)}{\sigma}\right) \cdot Q_{n-\nu-1}\left(\frac{\nu}{\sqrt{(n-\nu)(n-\nu-1)}}\frac{T_n(\nu)}{\sigma}\right)\cdot \exp\left[-\frac{1}{2\sigma^2}\frac{n\nu}{n-\nu}\cdot T_n^2(\nu)\right]
$$

$$
(22)\quad f\{U_n(\nu)\} = \frac{1}{P[N=\nu]}\binom{n}{\nu}\frac{1}{\sqrt{2\pi}\sigma}\left(\frac{\nu(n-\nu)}{n}\right)^{\frac{1}{2}}\cdot Q_{\nu-1}\left(\sqrt{\frac{\nu}{\nu-1}}\frac{n-\nu}{n}\frac{U_n(\nu)}{\sigma}\right) \cdot Q_{n-\nu-1}\left(\sqrt{\frac{n-\nu}{n-\nu-1}}\frac{\nu}{n}\frac{U_n(\nu)}{\sigma}\right)\cdot \exp\left[-\frac{1}{2\sigma^2}\frac{\nu(n-\nu)}{n}U_n^2(\nu)\right]
$$

$$
(23)\quad f\{W_n(\nu)\} = \frac{1}{P[N=\nu]}\binom{n}{\nu}\frac{1}{\sqrt{2\pi}\sigma}\left(\frac{n^3}{4\nu(n-\nu)}\right)^{\frac{1}{2}}\cdot Q_{\nu-1}\left(\frac{n}{2\sqrt{\nu(\nu-1)}}\frac{W_n(\nu)}{\sigma}\right) \cdot Q_{n-\nu-1}\left(\frac{n}{2\sqrt{(n-\nu)(n-\nu-1)}}\frac{W_n(\nu)}{\sigma}\right)\cdot \exp\left[-\frac{1}{2\sigma^2}\frac{n^3}{4\nu(n-\nu)}W_n^2(\nu)\right]
$$

特に $W_n(\nu)$ の分布に於て (19) に依り $Q_r(x)$ の代りに $G_r(x)$ を入れて少し整理すれば，Godwin の結果に一致する事がわかる．

前節で述べた如く $N=\nu$ なる条件を消去した T_n, U_n, W_n の分布は

$$
(24)\quad \begin{aligned} f(T_n) &= \sum_{\nu=1}^{n-1} f(T_n(\nu))P[N=\nu] \quad , \quad f(U_n) = \sum_{\nu=1}^{n-1} f(U_n(\nu))P[N=\nu] \\ f(W_n) &= \sum_{\nu=1}^{n-1} f(W_n(\nu))P[N=\nu] \end{aligned}
$$

として求まるのである．

3. $P[N=\nu]$ の計算

前節 (24) に見る如く T_n, U_n, W_n の分布には $P[N=\nu]$ の値は直接関係がないが，此の値を知つておれば母集団が正規分布に従つて居る事がわかつて居る場合取られる標本の randomness を検するのに役に立つから此処で求めておく．品質管理を行ふ時抜き取る標本毎に $\bar{x}$ より小さい値の数を記録し，その数があまり少ないとかあまり多い場合が屡々現はれるならば，正規分布の仮定が正しくないか，工程に外的な擾乱が働いて居るのではないかと疑つて見るのである．

$$(25)\quad \begin{aligned} P[N=\nu] &= P_r\{y_1 \leqq \cdots \leqq y_\nu \leqq 0 \leqq y_{\nu+1} \leqq \cdots \leqq y_{n-1} \leqq -y_1 - \cdots - y_{n-1}\} \\ &= P_r[(\xi_1, \cdots, \xi_{n-1}) \in D_2] \end{aligned}$$

であることは明らかで (18) 式の $P[N=\nu]$ 以外の函数を ξ_ν に就いて 0 から ∞ まで積分した値である．前節で述べた $G_r(x)$, $F_r(x)$ の数表があるから，此等を利用して数値計算を行ふ事が出来る．又次の様にしても計算することが出来る．(14) 式の $\xi_1, \xi_2 \cdots, \xi_{n-1}$ を次の Polar Transformation に依り $r, \theta_2, \cdots \theta_{n-2}$ に変換する．

$$(26)\quad \begin{cases} \xi_1 = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_3\sin\theta_2 \\ \xi_2 = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_3\cos\theta_2 \\ \xi_3 = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\cos\theta_3 \\ \vdots \\ \xi_{\nu-1} = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_\nu\cos\theta_{\nu-1} \\ \sqrt{\dfrac{n}{n-\nu}}\,\xi_\nu = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_{\nu+1}\cos\theta_\nu \\ \xi_{\nu-1} = r\sin\theta_{n-1}\sin\theta_{n-2}\cdots\sin\theta_{\nu+2}\cos_{\nu+1} \\ \vdots \\ \xi_{n-2} = r\sin\theta_{n-1}\cos\theta_{n-2} \\ \xi_{n-1} = r\cos\theta_{n-1} \end{cases}$$

此の変換に依つて

$$(27)\quad P[N=\nu] = K_n \int \cdots \int_{D_3} \sin^{n-3}\theta_{n-1}\sin^{n-4}\theta_{n-2}\cdots\sin^2\theta_4\sin\theta_3\, d\theta_{n-1}\cdots d\theta_3 d\theta_2$$

を得る．茲に

$$(28)\quad K_n = n!\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} 2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)$$

であり，範囲 D_3 は (15), (26) より

$$(29)\quad D_3 \begin{cases} \dfrac{\pi}{2} \geqq \theta_{n-i} \geqq \tan^{-1}\left(\sqrt{\dfrac{i}{i+2}}\sec\theta_{n-i-1}\right), & (i=1,2,\cdots,n-\nu-2) \\ \dfrac{\pi}{2} \geqq \theta_{\nu+1} \geqq \tan^{-1}\left(\sqrt{\dfrac{n(n-\nu-1)}{\nu}}\sec\theta_\nu\right) & \\ \tan^{-1}\left(\sqrt{\dfrac{n-\nu}{n(\nu-1)}}\sec\theta_{\nu-1}\right) \geqq \theta_\nu \geqq 0 & \\ \tan^{-1}\left(\sqrt{\dfrac{j+2}{j}}\sec\theta_{j-1}\right) \geqq \theta_j \geqq 0 & (j=\nu-1,\nu-2,\cdots 3) \\ \dfrac{\pi}{3} \geqq \theta_2 \geqq 0 & \end{cases}$$

此の式は n の低い所では簡単に計算出来るが，n の大きい所では Bell Relay Computor に依つて計算することが出来る様である．

扨て $P[N=\nu]$ は母集団分布の対称性から

$$(30)\quad P[N=\nu] = P[N=n-\nu]$$

が成立して居るから計算の箇数を半減させる事が出来る．計算に当つては

$$\sum_{\nu=1}^{n-1} P[N=\nu]=1 \tag{31}$$

をチェックとして使用する．第1表は以上述べた様にして求めた $P[N=\nu]$ の値である．

第 1 表　$P[N=\nu]$

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	13	19	20
1	1.0000	0.5000	0.1755	0.0489	0.0115	0.0024	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.5000	0.6490	0.4511	0.2232	0.0880	0.0294	0.0086	0.0023	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3		0.0000	0.1755	0.4511	0.5306	0.4096	0.2397	0.1147	0.0469	0.0170	0.0056	0.0017	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
4			0.0000	0.0489	0.2232	0.4096	0.4610	0.3766	0.2441	0.1323	0.0622	0.0260	0.0099	0.0034	0.0010	0.0003	0.0001	0.0000	0.0000
5				0.0000	0.0115	0.0880	0.2397	0.3766	0.4134	0.3501	0.2431	0.1440	0.0748	0.0348	0.0148	0.0056	0.0021	0.0006	0.0001
6					0.0000	0.0024	0.0294	0.1147	0.2441	0.3501	0.3779	0.3283	0.2396	0.1516	0.0850	0.0431	0.0197	0.0084	0.0035
7						0.0000	0.0004	0.0086	0.0469	0.1323	0.2431	0.3283	0.3505	0.3101	0.2351	0.1565	0.0933	0.0502	0.0249
8							0.0000	0.0001	0.0023	0.0170	0.0622	0.1440	0.2396	0.3101	0.3282	0.2945	0.2300	0.1595	0.0996
9								0.0000	0.0000	0.0006	0.0056	0.0260	0.0748	0.1516	0.2351	0.2945	0.3097	0.2813	0.2249
10									0.0000	0.0000	0.0001	0.0017	0.0099	0.0348	0.0850	0.1565	0.2300	0.2813	0.2940
11										0.0000	0.0000	0.0000	0.0004	0.0034	0.0148	0.0431	0.0933	0.1595	0.2249
12											0.0000	0.0000	0.0000	0.0001	0.0010	0.0056	0.0197	0.0502	0.0996
13												0.0000	0.0000	0.0000	0.0000	0.0003	0.0021	0.0084	0.0249
14													0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0035
15														0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
16															0.0000	0.0000	0.0000	0.0000	0.0000
17																0.0000	0.0000	0.0000	0.0000
18																	0.0000	0.0000	0.0000
19																		0.0000	0.0000

4. 効　率

母集団標準偏差 σ の推定に，T_n, U_n を用ひる時の効率を調べよう．此のため此等の分散を求めよう．今 $E(x|\nu)$ で $N=\nu$ なる時の x の数学的平均を表はし，$E(x)$ で此の条件を消去した時のものを表はす．更に ν に就いての平均を E_ν の記号で表はし，x の分散は $D^2(x)$ と書く．しかる時 T_n, U_n の分散は

$$(33)\qquad \begin{aligned} D^2(T_n)&=E_\nu E(T_n^2|\nu)-[E_\nu E(T_n|\nu)]^2 \\ D^2(U_n)&=E_\nu E(U_n^2|\nu)-[E_\nu E(U_n|\nu)]^2 \end{aligned}$$

となり．従つて $E(T_n|\nu), E(T_n^2|\nu)$ 等を計算すればよい．

扨て平均偏差 W に関しては

$$(33)\qquad \begin{aligned} E(W_n)&=\left[\frac{2}{\pi}\frac{n-1}{n}\right]^{\frac{1}{2}}\sigma \\ E(W_n^2)&=\frac{2}{\pi}\frac{n-1}{n^2}\left\{\frac{\pi}{2}+\sqrt{n(n-2)}+\sin^{-1}\frac{1}{n-1}\right\}\cdot\sigma^2 \end{aligned}$$

なる結果がある．[3] 所で此等の導出には T_n, U_n に於ける様に $N=\nu$ なる点が含まれて居ない為その儘利用する事が出来ない．但し次の様に計算のチェックに使用する事が出来る．

$$(34)\qquad \begin{aligned} E(W_n)&=\frac{2}{n}E_\nu[\nu E(T_n|\nu)] \\ E(W_n^2)&=\frac{4}{n^2}E_\nu[\nu^2E(T_n^2|\nu)] \end{aligned}$$

所で

$$(35)\qquad U_n(\nu)=\frac{n}{n-\nu}T_n(\nu)$$

であるから $E(T_n^i|\nu)$ $(i=1,2)$ を求めれば充分である．(21) の $T_n(\nu)$ の分布に於て $Q_r(x)$ を Grubbs の函数 $F_r(x)$ に書き直す時 $E(T_n^i|\nu)$ $(i=1,2)$ は

$$(36)\qquad E(T_n^i|\nu)=\frac{\sigma^i}{P[N=\nu]}\binom{n}{\nu}\frac{1}{\sqrt{2\pi}}\left(\frac{n\nu}{n-\nu}\right)^{\frac{1}{2}}\int_0^\infty x^iF_\nu(x)F_{n-\nu}\left(\frac{\nu}{n-\nu}\right)\cdot\exp\left[-\frac{1}{2}\frac{n\nu}{n-\nu}x^2\right]dx$$

となるが，$F_r(x)$ の数表を利用して数値計算することが出来る．又 $P[N=\nu]$ の時と同様に極変換を利用した時には

$$(37)\qquad E(T_n^i|\nu)=\frac{\sigma^i}{\nu^{\frac{i}{2}}}E(\xi_\nu^i|\nu)$$

だから $\nu=2,3,\cdots,n-1$ に対して

$$(38)\qquad E(T_n^i|\nu)=\frac{\sigma^i}{P[N=\nu]}L_{i,\nu}\int\cdots\int_{D_3}\sin^{n-3+i}\theta_{n-1}\sin^{n-4+i}\theta_{n-2}\cdots\sin^{\nu-1+i}\theta_{\nu+1}\sin^{\nu-2}\theta_\nu\cdot \cos^i\theta_\nu\sin^{\nu-3}\theta_{\nu-1}\cdots\sin^2\theta_4\sin\theta_3 d\theta_{n-1}\cdots d\theta_2.$$

$\nu=1$ に対しては

$$(39)\qquad E(T_n^i|1)=\frac{\sigma^i}{P[N=1]}L_{i,1}\int\cdots\int_{D_3}\sin^{n-3+i}\theta_{n-1}\sin^{n-4+i}\theta_{n-2}\cdots\sin^{2+i}\theta_4\sin^{1+i}\theta_3 \sin^i\theta_2\,d\theta_{n-1}\cdots d\theta_2.$$

となる．茲に D_3 は勿論 (29) に定義された領域であり，且つ

(40)　$$L_{i,\nu}=n!\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1}\left(\frac{n-\nu}{n\nu}\right)^{\frac{i}{2}}2^{\frac{n-3+i}{2}}\Gamma\left(\frac{n-1+i}{2}\right)\quad(\nu=1,2,\cdots,n-1)$$

である.

(36) より

(41)　$$E(T_n^i|\,n-\nu)=\left(\frac{\nu}{n-\nu}\right)^i E(T_n^i|\,\nu)$$

が成立して居る. 従つて U_n に対しては

(42)　$$E(U_n^i|\,n-\nu)=E(U_n^i|\,\nu)$$

が示される. 斯くして我々は U_n, T_n の平均, 分散を計算する事が出来る.

扨て効率の話に戻らう. 此処では次の二つの場合が考へられる.

(i) σ を推定する場合 $\bar{x}$ より小さい観測値が何箇あるかに特に留意しないで, 全体的に不偏推定を行ふ.

(ii) 標本毎に $\bar{x}$ より小さい観測値の箇数に着目し, その都度推定を不偏に行ふ.

勿論 (ii) の場合は N の値が ν 箇であつたと言ふ智識を利用して居るから推定の効率は良くなる. 先づ (i) の場合を考へよう.

(43)　$$E(T_n)=d_n\sigma\quad,\quad E(U_n)=k_n\sigma$$

とおけば T_n/d_n, U_n/k_n は σ の不偏推定子になる. $\bar{x}$ より小さい観測値が何箇あるか留意しないわけであるから全体的に σ の不偏推定を行ふには此の形が用ひられる. 此の様な推定方式を採用する時の, 標本標準偏差に基いて推定する時に対する (相対的) 効率は次の様になる. 即ち分散

(44)
$$D^2(T_n/d_n)=E\left\{\frac{T_n}{d_n}-E\left(\frac{T_n}{d_n}\right)\right\}^2=\frac{1}{d_n^{\,2}}D^2(T_n)$$
$$D^2(U_n/k_n)=E\left\{\frac{U_n}{k_n}-E\left(\frac{U_n}{k_n}\right)\right\}^2=\frac{1}{k_n^{\,2}}D^2(U_n)$$

と, 標本標準偏差を不偏推定子になほしたものの分散

(46)　$$D^2(S_n/C_n)=E\left\{\frac{S_n}{C_n}-E\left(\frac{U_n}{C_n}\right)\right\}^2$$

とを比較すればよい. 但し

(46)　$$S_n=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n}}\quad,\qquad C_n=\sqrt{\frac{2}{n}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}=E(S_n)$$

第2表に, 不偏にするための係数 $\frac{1}{d_n}$, $\frac{1}{k_n}$, 及び $D^2(T_n/\sigma), D^2(U_n/\sigma)$ をまとめておく. 第3表は此等 T_n/d_n, U_n/k_n の効率及び比較のため '範囲', '平均偏差' の $\frac{S_n}{C_n}$ に対する効率を求めたものである.

(ii) の場合は次の様になる. $\bar{x}$ より小さい観測値の箇数に着目し, その都度推定を不偏にするから

(47)　$$E(T_n|\,\nu)=d_{n,\nu}\,\sigma\quad,\quad E(U_n|\,\nu)=k_{n,\nu}\,\sigma$$

とおけば, ν の値に着目して $\frac{T_n(\nu)}{d_{n,\nu}}$, $\frac{U_n(\nu)}{k_{n,\nu}}$ を推定に用ひることになる. 此の場合の全体的な効率は分散

第 2 表

n	$\frac{1}{d_n}$	$D^2\left(\frac{T_n}{\sigma}\right)$	$\frac{1}{k_n}$	$D^2\left(\frac{U_n}{\sigma}\right)$
2	1.772	0.1817	0.8862	0.7268
3	1.364	0.2241	0.6822	0.5919
4	1.302	0.1942	0.6512	0.4597
5	1.289	0.1520	0.6444	0.3528
6	1.282	0.1211	0.6412	0.2852
7	1.278	0.1001	0.6391	0.2371
8	1.275	0.0856	0.6376	0.2036
9	1.273	0.0748	0.6363	0.1787
10	1.271	0.0662	0.6354	0.1585
11	1.269	0.0596	0.6345	0.1430
12	1.268	0.0540	0.6339	0.1296
13	1.267	0.0494	0.6333	0.1193
14	1.266	0.0458	0.6329	0.1100
15	1.265	0.0424	0.6325	0.1023
16	1.264	0.0396	0.6321	0.0957
17	1.264	0.0371	0.6318	0.0897
18	1.263	0.0350	0·6315	0.0845
19	1.262	0.0329	0.6312	0.0803
20	1.262	0.0313	0.6310	0.0756

第 3 表

n	R. E. of $\frac{T_n}{d_n}$	R. E. of $\frac{U_n}{k_n}$	R. E. of Ramge of Total Sample *)	R. E. of Group Ranges *)	R. E. of Mean, Deviation
2	1.000	1.000	1.000		1.000
3	0.655	0.992	0.992		0.992
4	0.541	0.913	0.975		0.964
5	0.523	0.900	0.955		0.946
6	0.525	0.892	0.934		0.934
7	0.528	0.892	0.911		0.925
8	0.530	1.891	0.890		0.919
9	0.531	1.891	0.869		0.914
10	0.531	0.890	0.850		0.910
11	0.532	0.889	0.831		0.907
12	0.533	0.889	0.814	0.830	0.904
13	0.533	0.888	0.797	0.828	0.902
14	0.534	0.888	0.781	0.825	0.900
15	0.535	0.887	0.765	0.821	0.898
16	0.535	0.886	0.753	0.817	0.897
17	0.535	0.886	0.740	0.812	0.895
18	0.535	0.885	0.724	0.806	0.894
19	0.536	0.885	0.711	0.800	0.894
20	0.536	0.885	0.700	0.799	0.892

*) F. E. Grubbs and C. L. Weaver の Table[4] を基に算出したものである．

$$E_\nu E\left(\frac{T_n^2}{d^2_{n,\nu}}\Big|\,\nu\right)-\left[E_\nu E\left(\frac{T_n}{d_{n,\nu}}\Big|\,\nu\right)\right]^2=\sum_{\nu=1}^{n-1}P[N=\nu]\,E\left(\frac{T_n^2}{d^2_{n,\nu}}\Big|\,\nu\right)-1 \tag{48}$$
$$E_\nu E\left(\frac{U_n^2}{k^2_{n,\nu}}\Big|\,\nu\right)-\left[E_\nu E\left(\frac{U_n}{k_{n,\nu}}\Big|\,\nu\right)\right]^2=\sum_{\nu=1}^{n-1}P[N=\nu]\,E\left(\frac{U_n^2}{k^2_{n,\nu}}\Big|\,\nu\right)-1$$

と S_n/C_n の分散 (45) との比較である．上式にも見られる様に此の（ii）の場合にも全体的に不偏推定となつて居る．第4表に各 n と ν の値に対する係数 $\frac{1}{k_{n,\nu}}$ の表をあたへると共に，全体的な率も合せ示しておく．少し考へればわかる事であり，亦第3表にも示されて居る如く，T_n に基く推効定は効率がわるく使用に耐えないから $\frac{1}{d_{n,\nu}}$ の表は作成しなかつた，

5. 結 び

第3表，第4表に見る如く我々が提起した統計量のうち U_n に基くものは甚だ都合が良い．少し計算の手間は必要であるが Range の欠点を補ふものである．特に標本平均より小さい観測値の箇数に着目することは，$\bar{x}(\nu)$ を計算する時必要になることであるから，此れを利用して効率を高める

は容易に出来ることである．$n<10$ 位までは Range を用ひるとしても $n\geq 10$ では我々の U_n, $U_n(\nu)$ に基いて σ を推定する事が必要とならう．大きな n の値に対しては F. Mosteller の定義

第 4 表 $\frac{1}{k_{n,\nu}}$ の表及び効率

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.8862	0.6822	0.5714	0.5000	0.4500	0.411	0.40	0.38	0.3	0.3	0.3	**	**	**	**	**	**	**	**
3		0.6822	0.7044	0.6652	0.6202	0.759	0.543	0.514	0.490	0.47	0.45	0.4	0.4	**	**	**	**	**	**
3			0.5714	0.6652	0.6729	0.656	0.631	0.605	0.580	0.557	0.537	0.52	0.50	0.50	0.5	0.5	**	**	**
4				0.5000	0.6202	0.656	0.660	0.650	0.635	0.617	0.599	0.581	0.564	0.556	0.53	0.52	0.5	0.4	**
5					0.4500	0.579	0.631	0.650	0.652	0.646	0.636	0.623	0.609	0.595	0.582	0.566	0.56	0.53	0.5
6						0.411	0.543	0.605	0.535	0.646	0.648	0.643	0.636	0.626	0.615	0.604	0.591	0.584	0.58
7							0.40	0.514	0.580	0.617	0.636	0.643	0.646	0.641	0.636	0.628	0.619	0.609	0.600
8								0.38	0.490	0.557	0.599	0.623	0.636	0.641	0.642	0.640	0.635	0.629	0.621
9									0.3	0.47	0.537	0.581	0.609	0.626	0.636	0.640	0.640	0.639	0.634
10										0.3	0.45	0.52	0.564	0.595	0.615	0.628	0.635	0.639	0.639
11											0.3	0.4	0.50	0.556	0.582	0.604	0.619	0.629	0.634
12												**	0.4	0.50	0.53	0.566	0.519	0.609	0.621
13													**	**	0.5	0.52	0.56	0.584	0.600
14														**	**	0.5	0.5	0.53	0.558
15															**	**	**	0.4	0.5
16																**	**	**	**
17																	**	**	**
18																		**	**
19																			**
効率	1.00	0.99	0.97	0.96	0.94	0.93	0.92	0.92	0.92	0.91	0.91	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.89

** の所は $P[N=\nu]$ が小さい故実際には必要がないので省略した．

した Quasi Range[5] を用ひる事が出来る.

最後に本問題を研究するにあたり，いろいろ御意見を聞かせて下さつた阪大理学部小川潤次郎先生，及び当研究所の内田良男氏に深く感謝し，また面倒な数値計算を手伝つていただいた高瀬浜子氏にも玆で御礼を申し上げる.

統 計 数 理 研 究 所

参 考 文 献

[1] H. J. GODWIN, "On the distribution of the estimate of mean deviation obtained from samples from a normal population" Biometrika, vol 33, (1945), pp. 254—265

[2] F. E. GRUBBS, "Sample Criteria for testing outlying observations" The Ann. of Math. Stat. vol. 21, (1950), pp. 27～58

[3] R. A. FISHER, "A mathematical examination of the methods determining the accuracy of an observation by the mean error and by the mean square error" Monthly Notice of the Royal Astronomical Society, vol. 80, (1920), pp. 758—770

[4] F. E. GRUBBS and C. L. WEAVER, "The best unbiased estimate of population standard deviation based on group ranges" Journal of the American Statistical Association, vol. 42, (1947), pp. 224—241.

[5] F. MOSTELLER, "On some useful "inefficient" statistics" The Ann. of Math. Stat. vol. 17, (1946), pp. 377—408.

[6] 柴垣和三雄，八桁函数表 I，指数函数，丸善，1949