

不完全な気象資料の統計処理*

鈴木 栄一 児玉 東子**

要 旨

欠測値を含んだ気象資料を統計的に処理して平均や標準偏差を求め、観測された値とともに生かしてつかう方法を解説する。

1. はしがき

気象観測に従事されている地方現場の方から、次のようなことを聞かれたことがある。

“台風がきてある観測所のロビンソン風力計がある時刻以後故障し、そのため一部欠測した。こんなとき欠測時間をも入れて平均風速とか全風程とか又は風速変動(風の息に関係する)を示す標準偏差などを推定する方法はないか”。

“自記雨量計が何らかの原因である時刻以降調子が悪くなり、どうもよみ取れないが、今までにない降水量が観測されているらしい。これら読取不能の部分も入れて総降水量又は平均降水量を推定するにはどうすればよいか”。

事実伊勢湾台風の際、名古屋地方気象台管内の観測所にはロビンソンの故障が多かった***。

これに類した問題は気象の分野に可成り多いように思われる。これに対する統計的解決法としてはいくつか考えられるが、その中で比較的役立ちそうに思われる Censored sampling theory (監査済み標本の理論)の一部とその応用可能性を具体的実例を中心にのべてみる。元来破壊試験などの中から生れたこの方法は案外便利であり、実際筆者らが台風時における電柱倒壊と風圧との関係を調べる際、こうした推定値を利用したことが

あるのでここに一応報告するが、以下のべることは論文や研究報告といったものでなく、いくつかの具体例を伴った解説的紹介である。

2. Censored sampling theory とは

1950年以降、統計関係のいくつかの雑誌に censored sample という題目の論文が数多く見られ、1つの流行(?)の感がある。censored sample とは意味からいえば不完全標本又は不充分標本とでも訳すべきものであって、つまり得られた標本の中に、数値として興えられずに、その存在範囲のみがわっているようなものがあるという意味で、標本として不完全または不充分なものだからである。

Censored sample 又は truncated sample についての詳しい総合報告は千葉大学の村上正康氏[†]によりなされている。しかしここでそれを詳述敷衍する必要も余裕もないので、吾々の問題に当面役立ちそうなことだけについて考え方と結果の要点をあげよう。

Gupte, A.K. (1952) はある母集団から得られた大きさ n の任意標本について2つの型の censored sample を考えている。

(i) 観測がすべてある値 (truncation point といわれる) 以下又は以上の領域でだけしか得られていない場合。(第1型)

(ii) n コの標本中 k コ ($k < n$) の標本は数値として与えられているが、残り $n-k$ コについては k コの標本のいづれよりも大きいか又はいづれよりも小さいかということだけがわかっている場合(第2型)

気象資料には本質的に第1型、第2型ともあり得る

* On the Statistical Treatment of Incomplete Weather Data

** E. Suzuki & T. Kodama, 気象研究所予報研究部

*** 伊勢湾台風気象概報 名古屋地方気象台刊 昭和35年2月

が、ロビンソン風力計は何時どんなところではどんな条件で故障したり、とばされたりするかわからないし、自記雨量計はいつ故障するかわからない。その意味で truncation point は一定でなく、ときと場合でかわる変数と見られ、一定の truncation point を仮定する第1型よりも第2型について考えた方が妥当であろう。

Gupta は平均 μ 、分散 σ^2 である正規分布 $N(\mu, \sigma^2)$ からの大きさ n の標本のうち k コ ($0 < k < n$) の標本については1つ1つ数値としてわかっているが、残りの $n-k$ コについては k コ中の最大値 $x_{(k)} = \text{Max } x_i$ ($i=1, 2, \dots, k$) より大きいことだけしかわかっていない場合について、このような不完全な資料から未知の μ, σ^2 の最尤推定値 μ^*, σ^{*2} とその推定値の良否の尺度である漸近分散行列 (Asymptotic dispersion matrix) を求める方法を研究した。

それによると、 k コの標本についての標本平均 \bar{x} 、標本分散 s^2 を計算し、 k コの標本中の最大値 $x_{(k)}$ (最大順序統計量又は k 番目の順序統計量) と \bar{x} との差 $d = x_{(k)} - \bar{x}$ から

$$\psi = s^2 / (s^2 + d^2), \quad p = k/n$$

をもとめ、 ψ, p から Gupta の表を用いて z をよみとり

$$\sigma^* = d/z$$

によって σ の最尤推定ができる。これより μ^* は

$$\mu^* = \bar{x} + (\sigma^{*2} - s^2) / \alpha$$

となる。また、 (μ^*, σ^{*2}) による (μ, σ^2) の推定の良さを示す尺度である (μ^*, σ^{*2}) の漸近分散行列は尤度関数

$$L(\mu, \sigma^2) = \frac{n!}{(k-1)!(n-k)!} \left(\frac{\sigma}{\sqrt{2\pi}} \right)^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^k (x_i - \mu)^2 \right] \cdot \left\{ \int_{x_{(k)}}^{\infty} \exp\left[-\frac{\sum (x - \mu)^2}{2\sigma^2} \right] dx \right\}^{n-k}$$

の log の母数 μ, σ^2 に関する2階偏微分を要素とする行列の逆であることは周知の通りであって、Gupta は $n \rightarrow \infty$ のとき $\eta_k = (x_k - \mu) / \sigma$ は

$$\int_{-\infty}^{\hat{\eta}} \phi(t) dt = p, \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

なる $\hat{\eta}$ に収束すること、および、

$$\lim_{n \rightarrow \infty} -\frac{\sigma^2}{n} \frac{\partial^2 \log L}{\partial \mu^2} = p + \phi(\hat{\eta})(A - \hat{\eta}) = v_{11}$$

$$\lim_{n \rightarrow \infty} -\frac{\sigma^2}{n} \frac{\partial^2 \log L}{\partial \mu \partial \sigma} = -\phi(\hat{\eta}) + \hat{\eta} \phi(\hat{\eta})(A - \hat{\eta})$$

$$= v_{12} = v_{21}$$

$$\lim_{n \rightarrow \infty} -\frac{\sigma^2}{n} \frac{\partial^2 \log L}{\partial \sigma^2} = 2p - p\hat{\eta}\phi(\hat{\eta}) + \hat{\eta}^2 \phi(\hat{\eta})(A - \hat{\eta}) = v_{22}$$

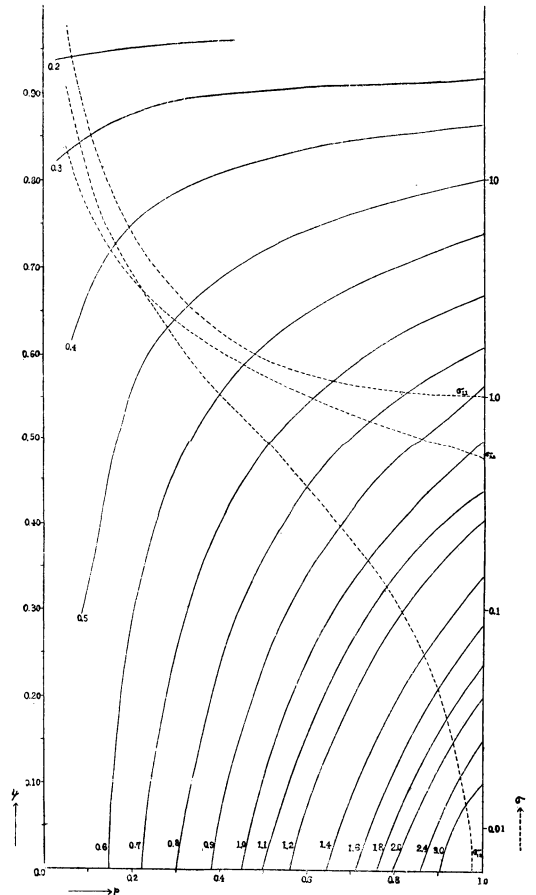
$$A = \phi(t) / \int_{\hat{\eta}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

とすると、漸近分散行列は

$$\frac{\sigma^2}{n} \cdot [v_{ij}]^{-1} = [\sigma_{ij}]$$

となることを導びき、各 $p = 0.05 \sim 0.99$ (0.05刻み) に対する $\sigma_{11}, \sigma_{12} = \sigma_{21}, \sigma_{22}$ の表を作成した。吾々は内挿の便を計るため、 z と σ_{ij} を求める図表を作成した。

(第1図)



第1図 ψ と p を使って、 $Z, \sigma_{11}, \sigma_{12}$ および σ_{22} を求める図

次に最も簡単な指数分布からの Censored sample について考える。母数 θ を含む指数分布の密度関数は

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

で、これから大きさ $n = m + r$ コの標本をとったとき、 m コは数値がわかっていて、大きさの順にならべられ、

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(m)}$$

となったが r コ ($r > 0$) はいづれも $x_{(m)}$ より大きい、数値は分らなかったとする。 θ の最尤推定 θ^* は尤度関数

$$L(\theta) = \prod_{i=1}^m (\theta e^{-\theta x_i}) \cdot \left[\int_{x_{(m)}}^{\infty} \theta e^{-\theta x} dx \right]^r$$

の log を θ で偏微分として 0 とおくことにより、

$$\theta^*(m+r) = m(rx_{(m)} + \sum_{i=1}^m x_{(i)})^{-1}$$

となる。また $\theta^*(m+r)$ の漸近分散は前と同様にして

$$\begin{aligned} \text{Var.} \{ \theta^*(m+r) \} &= \theta^2 (1 - e^{-\theta x_{(m)}})^{-1} (m+r)^{-1} \\ &\approx \{ \theta^*(m+r) \}^2 (1 - e^{-\theta^*(m+r) x_{(m)}})^{-1} (m+r)^{-1} \end{aligned}$$

で、 $m+r$ コ、 m コによる θ の推定として

$$\theta^*(m) = \{ \bar{x}(m) \}^{-1} = \left(\frac{\sum_{i=1}^m x_{(i)}}{m} \right)^{-1} \quad (\text{既知})$$

$$\theta^*(m+r) = \{ \bar{x}(m+r) \}^{-1} = \left(\frac{\sum_{i=1}^{m+r} x_i}{m+r} \right)^{-1} \quad (\text{未知})$$

$\bar{x}(m)$, $\bar{x}(m+r)$ はそれぞれ m コ、 $m+r$ コの標本平均の意なる記号を用いると、平均値としては

$$\bar{x}(m+r) = \bar{x}(m) + \frac{r}{m} x_{(m)} \quad (r \geq 0)$$

となり、未知領域での平均は

$$\begin{aligned} \bar{y}_r &= \frac{1}{r} \sum_{i=1}^r x_{(m+i)} = \bar{x}(m+r) + x(m) \\ &= \bar{x}(m) + \frac{m+r}{m} x_{(m)} \quad (r > 0) \end{aligned}$$

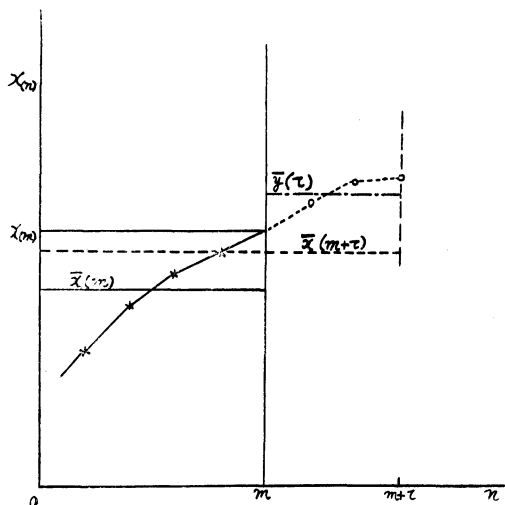
となる。 $x_{(m)}$ はもちろん、 m の単調非減少関数であって以上の関係を示すと第2図のごとくなる。なお $\bar{x}(m+r)^{-1}$ の分散は明らかに

$$\begin{aligned} \text{Var.} \{ \bar{x}(m+r)^{-1} \} &= 1 / (m+r) \bar{x}(m+r)^2 \\ &\{ 1 - \exp \{ -\bar{x}(m) / \bar{x}(m+r) \} \} \end{aligned}$$

となるが、これはあまりつかわれなない。

この他の分布 (たとえばポアソン分布、2項分布、ガンマ分布その他) についても全く同じような考え方のもとで理論が展開されるが、当面必要な2つの分布についての結果をあげておくにとどめ、風や降水量のデータ処理について具体的に示そう。

たゞ、Censored sampling theory の応用にはどうしても数値表又は図表が必要であるにも拘わらず、現在これらを総合的にまとめたものが出版されています、どう



第2図 Censored sample の説明図

しても原論文を1つ1つみなければならぬ。この点はやがて何とかされるであろう。

3. 具体例とその検討

(例1) 汐岬の10分間最大風速と瞬間最大風速
伊勢湾台風の時、汐岬の最大風速は欠測がなく完全に記録されている。ここではその1部が仮りに欠測であったとして全体の推定をし、この方法の妥当性を check してみよう。伊勢湾台風がおそった昭和34年、9月26日11時30分から18時までの10分間最大風速 V_m 、瞬間最大風速 V_p の記録は第1表の通りである。 V_m について15時30分までは観測されなかったとし、その後16時~18時は25m/s以上の風は吹いたことは確かだが、数値は不明であったとし、11時30分から18時までの平均風速を推

第1表 汐岬における最大風速の記録

time	V_m (m/s)	V_p (m/s)	
k	11.30	17.0	24.6
	12.00	16.0	24.0
	12.30	16.8	23.2
	13.00	17.8	26.0
	13.30	20.4	30.6
	14.00	20.8	31.2
	14.30	21.0	29.6
	15.00	22.2	34.2
	15.30	23.8	34.6
	n-k	16.00	30.0
16.30		27.2	38.2
17.00		27.8	42.8
17.30		26.3	35.6
18.00		31.0	48.5

定してみる。

$$n=14, k=9 \quad x_{(k)}=x_{(9)}=23.8 \quad \bar{x}=\frac{1}{9}\sum_{i=1}^9 x_i=19.5$$

$$s^2=6.58, \quad d=x_{(9)}-\bar{x}=4.3, \quad p=9/14=0.64$$

$$\psi=s^2/(s^2+d^2)\doteq 0.26$$

であるから第1図より $Z\doteq 1.05$ となる。よって

$$\sigma^*=4.3/1.05\doteq 4.3,$$

$$\mu^*=19.5+(18.49-6.58)/4.3\doteq 22.4$$

が最尤推定値で、 μ^* の漸近分散とその平方根（標準偏差）は

$$\text{Var.}\{\mu^*\}=\sigma_{11}\cdot\sigma^{*2}/n=1.20\times 4.3^2/14\doteq 1.585,$$

$$\sqrt{\text{Var.}\{\mu^*\}}\doteq 1.26$$

全く同様にして σ^* の漸近分散とその平方根は

$$\text{Var.}\{\sigma^*\}=\sigma_{22}\cdot\sigma^{*2}/n=0.92\times 4.3^2/14\doteq 1.195,$$

$$\sqrt{\text{Var.}\{\sigma^*\}}\doteq 1.09$$

つまり、未知の平均、分散（風速 Vm についての）の推定はそれぞれ

$$22.4\pm 1.26, \quad 4.3\pm 1.09$$

の範囲であろうとなる。一方実際に18時までの資料により計算すると

$$\hat{\mu}=23.4, \quad \hat{\sigma}=4.88$$

となり、充分この範囲に入っている。 Vp に関しても全く同様にしらべてみよう。

$$n=14, k=9, \quad x_{(k)}=x_{(9)}=34.6, \quad \bar{x}=28.7$$

$$s^2=16.91, \quad d=6.1, \quad p=0.64$$

であるから、第1図をつかって

$$Z=0.95, \quad \sigma^*=6.1/0.95\doteq 6.2,$$

$$\mu^*=28.7+\frac{38.44-16.91}{6.1}\doteq 32.3$$

$$\sqrt{\text{Var.}\{\mu^*\}}\doteq 1.82, \quad \sqrt{\text{Var.}\{\sigma^*\}}\doteq 1.59$$

を得た、未知の μ, σ はそれぞれ

$$32.8\pm 1.82, \quad 6.2\pm 1.59$$

の範囲内と推定される。実際に18時までの資料では、

$$\hat{\mu}=33.6, \quad \hat{\sigma}=7.60$$

であるから、これもまづ合っていると見做される。

(例2) 東京の10分間最大風速と瞬間最大風速

伊勢湾台風のとき、9月26日18時から27日9時までの東京10分間最大風速 Vm と瞬間最大風速 Vp の記録は第2表のとおりである。26日23時~27日3時までの Vm, Vp はそれぞれ15m/s以上、30m/s以上であった。30分毎のこの31コの平均はそれぞれ12.0m/s, 24.1m/sであった。 $Vm>15m/s, Vp>30m/s$ の上記範囲が欠測であったとし、残り22コの資料から全体の平均風速を推定する。

第2表 東京における風速の記録

time	$Vm(m/s)$	$Vp(m/s)$
18	5.4	10.8
	6.3	12.0
19	7.0	13.5
	7.6	15.0
20	8.5	16.8
	9.6	19.0
21	10.8	20.9
	11.9	23.0
22	13.1	25.0
	14.3	28.1

23	15.9	30.5
	17.1	33.7
24	18.8	37.0
	19.4	38.0
1	19.2	37.0
	18.5	36.0
2	17.4	34.8
	16.7	32.8
3	15.4	31.0

	14.0	29.0
4	13.0	17.0
	12.2	25.5
5	10.8	24.0
	10.5	22.5
6	9.8	21.0
	9.1	19.3
7	8.8	18.5
	8.3	17.5
8	8.1	16.5
	7.9	15.5
9	8.0	15.0

(i) Vm に関する推定

$$n=31, k=22, \quad x_{(k)}=x_{(22)}=14.3, \quad \bar{x}=9.8,$$

$$d=4.5, \quad p=0.68, \quad \psi=0.228$$

であるから、

$$z=1.1239, \quad \sigma^*=4.004, \quad \mu^*=12.0,$$

$$\sqrt{\text{Var.}\{\mu^*\}}\doteq 0.746$$

となり、実際の $\hat{\mu}=12.0$ とピッタリ合っている。

(ii) Vp に関する推定

$$n=31, k=22, \quad x_{(k)}=29.0, \quad \bar{x}=19.8, \quad d=9.2,$$

$$p=0.68, \quad \psi=0.241$$

であるから、全く同様にして

$$z=1.1091, \quad \sigma^*=8.295, \quad \mu^*=24.4,$$

$$\sqrt{\sqrt{\text{Var.}\{\mu^*\}}}=1.606$$

となり、よく合っている。

また、 Vp, Vm に関する $\hat{\sigma}$ と σ^* もよく合っていることが分る。

(例3) 岐阜の10分間最大風速と瞬間最大風速

伊勢湾台風のとき、9月26日15時から27日5時までの岐阜の30分毎の風速記録を第3表に示す。

第3表 岐阜における風速の記録

time (h)	V_m (m/s)	V_p (m/s)
15	5.7	9.0
	6.3	10.2
16	7.1	11.5
	8.4	13.2
17	9.9	15.2
	11.2	17.0
18	11.5	17.3
	10.5	16.0
19	10.9	18.0
	12.5	21.0
20	14.8	24.6
	18.3	30.0
21	23.4	37.0
	28.9	44.0
22	30.8	44.0
	26.7	40.0
23	19.6	29.5
	13.4	20.0
24	10.8	15.5
	11.5	15.3
1	11.5	15.5
	9.5	14.0
2	9.5	15.0
	11.0	17.0
3	12.5	18.5
	13.0	18.6
4	12.2	18.0
	11.5	16.4
5	10.0	14.8

(i) V_m に関する推定

26日21時から22時30分までの4つの資料は20m/s以上である。これがなかったものとして同様の推定をする。 n コ(全資料)の平均は実際は13.5m/sである。 $V_m > 20$ m/sなる4コのぞいた25コでの平均は11.3m/sである。この4コは欠測と仮定して推定することになる。

$$n=29, k=25, x_{(k)}=19.6, s^2=9.36, d=8.3,$$

$$\bar{x}=11.3, \psi=0.120, p=0.86$$

$$\therefore z=1.8283, \sigma^*=4.551, \mu^*=12.7,$$

$$\sqrt{\text{Var.}\{\mu^*\}} \doteq 0.860$$

だから、13.5は 12.7 ± 0.860 の範囲内にあり、一応よい推定となる。

(ii) V_p に関する推定

26日20時30分から22時30分まで5つの資料(30m/s以上)が欠測と仮定して同様の推定をする。

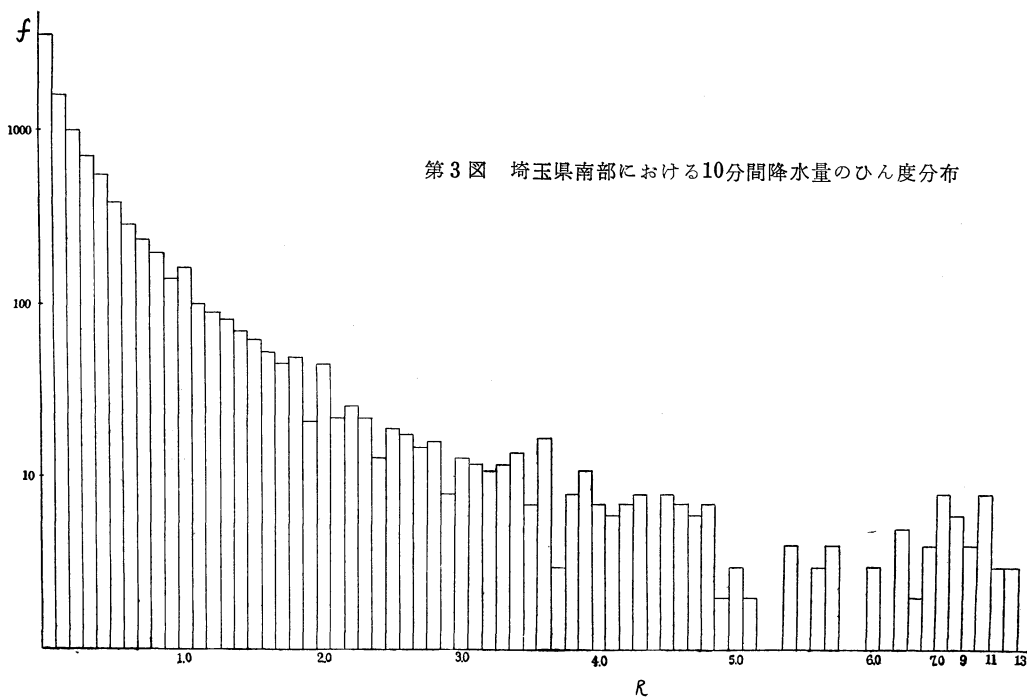
$$n=29, k=24, x_{(k)}=29.5, s^2=17.87, \bar{x}=16.7,$$

$$d=12.8, \psi=0.098, p=0.83$$

$$\therefore z=1.77634, \sigma^*=7.206, \mu^*=19.5,$$

$$\sqrt{\text{Var.}\{\mu^*\}} \doteq 1.371$$

となる。実際は $\hat{\mu}=20.6$ だから 19.5 ± 1.371 の範囲内に入り、一応良好な推定であったことが確かめられた。



第3図 埼玉県南部における10分間降水量のひん度分布

Vp, Vm の各標準偏差についても同様によい推定であることが確かめられている。

以上は主として風についての例であるが、

(a) 正規分布に従うことが大体わかっている変量であること。

(b) p があまり小さくないこと。

が満足されておれば、欠測を生かして統計処理をなし得ることがわかる。

次に分布が正規型でない例として、極めて短時間の降水量を考えてみる。

(例4) 埼玉県南部における10分間降水量

1959年の夏から秋にかけての9回の標記区域における降水量日記々録から読み取った全資料10046コから度数分布を作ってみると第3図のごとく、割合キレイな指数分布をしている³⁾。(縦軸が対数目盛)

このデータのうち10分間降水量3.5mm以上は251回、3.4mm以下は9795回である。一応完全なデータが実際とられたのであるが、応用例としてのべるために、3.5mm以上のデータがなかったものとして、3.4mm以下のデータから、平均雨量を推定してみよう。いま

$$m=9795, r=251, x_{(m)}=3.45$$

とし、 $\bar{x}(m) \approx 0.402$ (3.4mm以下の平均値)だから

$$\begin{aligned} \bar{x}(m+r) &= \bar{x}(m) + \frac{r}{m} x_{(m)} \\ &= 0.402 + \frac{251}{9795} \times 3.45 \approx 0.482 \end{aligned}$$

が全資料10046コの平均の推定である。実際は0.499であったが先づよく合っているとと言える。同様に4mm以

上が欠測、3mm以上が欠測とした場合はそれぞれ、推定が0.487, 0.495となっていづれにしても充分よい推定がなされることがわかる。推定の誤差はオーダーとして0.02~0.025位(正確な計算はしていない)であるから、もちろん推定値±誤差の範囲内に実際の値は入っている。

4. あとがき

以上いくつかの具体的な例からみて、確かにここへの推定法はかなり有用であることがわかるであろう。

いづれにしても推定であるから、そのつもりで用いるべきであるが、欠測であるからといって統計処理ができないと思うのは早計であって、欠測値がある条件を満足している限り、それを生かして利用することが可能なのである。欠測が多ければ、推定の誤差(推定の漸近分散の平方根)も当然大きくなるので統計処理しても意味ないかもしれないが、少なれば活用の方途を見出し得るであろう。

文 献

- 1) 村上正康: Truncate 標本及び Censored 標本による母数推定. 統計科学研究, 第1巻第2号 15~34頁 (1956)
- 2) Gupta, A.K.: Estimation of the mean and standard deviation of anormal population from a censored sample. Biometrika 39 (1952) 260-273
- 3) 今井一郎・鈴木栄一: 相関解析による雨域の構造の研究 (1959) 未印刷

気象学会秋季大会記事 正誤表

誤植のためご迷惑をおかけしたことをお詫びします。

頁	行	誤	正	頁	行	誤	正		
35	右	33	太陽活動持数	太陽活動指数	38	左	23	示度を関数として	示度との関数として
36	左	38	気象予報	気研予報	39	右	36	5832	5822
//	右	17, 23	//	//	40	左	22	各各分野	各分野
//		25	ひきつづき	ひきつづき	40	右	24	南庄の	南北の
//		30	気象予報	気研予報	41	右	4	湿度の視程	湿度と視程
37	左	8	興えられた	与えられた	//		9	冷却済	冷却剤
//		36	気象予報	気研予報	42	左	31	増効果	増雨効果
//	右	8	差でれる	差である	//	右	7	Rについて	Rについて
//	右	10	気象予報	気象庁予報	43	左	9	磯野謙二	磯野謙治
//	右	40	またわ	または	//		17	考察しの	考察して

(345頁につづく)