天　気
1985年12月
Vol. 32, No. 12
415（稀現象）

気象，気候における稀現象の解析

鈴木栄一*

はじめに

1977～1984年の間にいくつかの気象統計，気候統計の国際会議が開催され，気象極値（meteorological extremes）と気候稀現象（climatic rare events）の統計解析がその都度取上げられ，このテーマに対し強い関心が示されている．

その理由は社会的諸関係が複雑多様化してきたため日常的によくおこる気象現象よりも災害の原因となる異常現象への関心が高まってきたからであろう．

岡本雅典氏の「第2回統計気候学国際集会報告」(1984「天気」Vol. 31, No. 11）はコンパクトで正確なものであるが，紙面の都合でこのテーマに関する研究報告の多くが省略されている．

このリスボン会議での第10セッションで，座長となったI.I. Gringorten と筆者は簡単な summarized memo でこれまでの研究現況と今後の諸問題をあげて総括をした記憶がある．

この memo と，筆者が十数年来まとめてきた「極値予測論」をもとに，予備知識を全くもたない方々に理解して頂けるように解説してみたい．

そのため，主要な問題の提起と用語の説明をまず行うことにする．

1. 問題提起と用語

具体的な問題として，次の3つをあげることができよう．

（a）ある地点の $n=10$ 年間にわたる年最大風速データとして $x_1=47, x_2=50, \cdots, x_{10}=41$（単位 m/s）が与えられた．これにより $N_1=20$ 年および $N_2=50$ 年における年最大風速の極値を予測せよ．

（b）ある地点の積雪深データ $N=150$ 個を収集し，その度数分布を作った．ここから大きさ n の標本を抽出し，その最大値をもとめると，標本のとり方により異なっているが，もとの度数分布と，この最大値のあらわれ方との関係は？

（c）面倒な数式をなるべく用いないで，グラフによりデータを処理する合理的な方法はないか？

この他にも問題はあるが，考察の要点が不明確になるので，一応上記3問題を中心とする．

つぎのことは知っておられる方も多かろうが，この解説で用いられる基本用語を一括しておく．

（i）順序標本（順序統計量）…ある母集団から大きさ n の標本を抽出したとき，これを小さい方から大きい方へと順序にならべたもの．

（ii）極値…順序標本における最大と最小で，気象界では最大を極値とすることになっている．

（iii）稀現象…ある所定の小さい確率でおこる現象をいい，この小さい確率は分析目的に応じ，理論モデルできめておく．

（iv）再現期間…任意の x 以上の値が出現する確率 $P(X \geqq x)$ の逆数で，もしこの確率が0.1なら再現期間 $T=10$ となる．これを recurrence interval ともいう．

（v）確率変数…不確実性をもつ事象の元に割当てられた実数変数で，例えば日最大風速など．

* Eiichi Suzuki, 青山学院大学経済学部．

(vi) 母数…度数分布モデル（つまり確率分布）の形態特性を示すもの．例えば正規分布での期待値μ，分散σ^2やポアソン分布でのλなど．

(vii) パラメータ…変数間の関係式にみられる定数や係数および基本分布から変形，誘導された分布の母数。

(viii) 極値確率紙…横軸に順序標本またはその換算値をとり，縦軸にこの順序標本をこえない確率（非超過確率という）をとったグラフ用紙で，一般に縦軸を不等間隔にしてデータを記入し分布形態を判断するもの．縦軸に再現期間を補足することもある（2重指数型，Weibull 型など）．

(ix) 極限分布…想定された無限母集団から大きさnの標本をとって，$n\to\infty$としたときに得られる何らかの標本統計量（例えば極値の換算変数）の確率分布．

気象界では確率変数と通常の変数を混用したり，母数とパラメータを区別しないこともある．この解説では一応区別しておくことにし，問題（b）をモデル化することからスタートする．なお石原健二（1981）の研究解説も是非読まれることを期待したい．ここでは併読をすすめるため，なるべく上記解説との重複をさけた．

2. もとの分布と標本最大の分布の関係

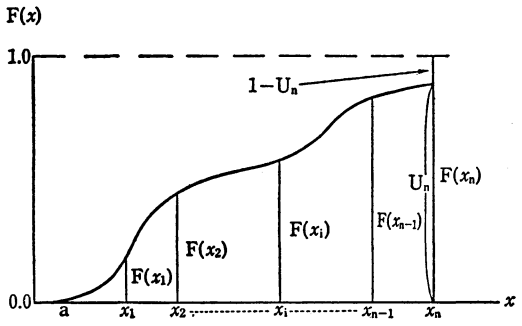
気象界でよく聞かれる素朴な質問の1つは，「なぜ，極値分布はもとの分布に関係なく指数型なのか？」である．この質問に正面から正確に答えた解説的報告は気象界にはほとんどなく，そのため時に誤解があったのはやむを得ない．（b）のようにデータを150個も集めれば，その度数分布図は一応安定した母集団的なものとなり，これをもとの確率分布としてモデル化してよいであろう．そこで，第1図によって，数学的厳密さを少し犠牲にして基本的事項を要約しよう．

積雪深のように連続な確率変数と考えられるものを一般にX（大文字）とし，その確率分布関数を$P(X\leqq x)\equiv F(x)$と書く（小文字xは実変数）．この母集団から大きさnの順序標本

$$\boldsymbol{X}=(X_1, X_2,\cdots, X_n),\quad X_1\leqq X_2\leqq\cdots\leqq X_n$$

を得たとする（n個の標本といわずに大きさnというのは理論上この$\boldsymbol{X}$を上記不等号を満たすn次元確率変数とみるからである）．

そこで，最大順序標本（極値統計量）の確率分布関数を$P(X_n\leqq x_n)\equiv F(x_n)=U_n$とおくと第1図のような状況となる．ただしこの図では分かりやすくするため，数学的厳密さを犠牲にして$\boldsymbol{X}$を1次元化し，$P(X_i\leqq x_i)\equiv F(x_i)$を一括して縦軸で示し$F(a)\equiv 0$とした．（ただし，$n$は可能な下限値）



第1図　順序統計量と対応する分布$F(x)$のモデル．

明らかにnを大きくすると$U_n\to 1$，$1-U_n\to 0$となるが，$n(1-U_n)$がどんな分布をするかについて次の基本定理が成立つ（証明略）．

〔基本定理〕$1-U_n$が連続で，密度関数$g(1-U_n)$をもち，$g(0)=c(>0)$が存在するなら，$n(1-U_n)$は$n\to\infty$のとき指数分布に収束し，その確率分布と密度関数とは容易に

$$\left.\begin{aligned} &P[n(1-U_n)\leqq y]=1-e^{-cy}\\ &f_1(y)=ce^{-cy}\end{aligned}\right\}\qquad(2.1)$$

となる．

しかしこれでは$U_n\to 1$だけで，$F(x_n)$が1に近づく“速さ”が考慮されていない．この速さについてはこの逆関数$x_n=F^{-1}(U_n)$の形状を用いた3つの定理と，分布$F(x)$の形状を用いた同様な3つの定理とがある．ここでは後述する Fisher-Tippett のタイプとの関連が分かり易い後者だけを要約した形であげる（鈴木栄一，1980～1983（I）を参照）．

一般に，極値統計量X_nはnが大きくなるほど大きくなる．そこでnに依存する実数a_n, b_nを用い$(X_n-b_n)/a_n$を安定した変数として考える．この極限分布については，Fisher, R.A. and L.H.C. Tippett (1928) 以来，多くの研究がなされ，B.V. Gnedenko (1943) によって極値極限分布が存在するための統計的な必要十分条件が明示されて，理論面はほとんど第2次大戦終期までに決定的な結論が導かれたと思われている．

それはもとの分布$F(x)$の具体的関数型に関係なく，$F(x)$について成り立つ極限的性質に依存して，3タイプの極限分布のどれかに収束し，これら以外の分布に収束しないという有名な結果である．

この3タイプを以下にあげる．

(1) すべての x に対し $F(x)<1$ で

$$\lim_{x\to\infty}\{1-F(kx)\}\{1-F(x)\}^{-1}=k^{-\alpha}$$

が $k(>0)$, $\alpha(>0)$ で成り立つなら n に依存する2つのパラメータ a_n, b_n を

$$F(a_n)=F(b_n)=1-n^{-1} \qquad (a_n>0)$$

とするとき確率変数 $(X_n-b_n)/a_n$ の極限分布は，

$$\left.\begin{aligned}\Phi(x)&=\exp(-x^{-\alpha}) \quad x\geqq 0\\ &=0 \qquad\qquad\quad x<0\end{aligned}\right\} \qquad (2.2)$$

(2) 有限な x_0 に対してのみ $F(x_0)=1$ で，どんな小さい $\varepsilon(>0)$ をとっても $F(x_0-\varepsilon)<1$ となっているとき

$$\lim_{x\to-0}\{1-F(kx+x_0)\}\{1-F(x+x_0)\}^{-1}=k^{\alpha}$$

が $k(>0)$, $\alpha(>0)$ で成り立つなら，パラメータ a_n, b_n を

$$F(b_n)=1-n^{-1},\ a_n=x_0-b_n$$

とするとき，確率変数 $(X_n-b_n)/a_n$ の極限分布は

$$\left.\begin{aligned}\Phi(x)&=\exp(-(-x)^{\alpha}) \quad x\leqq 0\\ &=0 \qquad\qquad\quad x>0\end{aligned}\right\} \qquad (2.3)$$

(3) $z\to x_0(x_0\leqq+\infty)$ のとき0に収束する z の連続関数 $A(z)$ が存在して

$$\lim_{z\to x_0}\{1-F(z(1+A(z)x))\}\{1-F(z)\}^{-1}=e^{-x}$$

ならば，パラメータ a_n, b_n を

$$F(a_n+b_n)=1-(ne)^{-1},\ F(b_n)=1-n^{-1}$$

とするとき，確率変数 $(X_n-b_n)/a_n$ の極限分布は，

$$\Phi(x)=\exp(-\exp(-x)) \qquad (2.4)$$

以上のように X_n を1次変換した $(X_n-b_n)/a_n$ の極限分布とは $F(x)$ の性質の具体的きめ方で導かれ，広義のワイブル型か2重指数型かのどちらかである((2.2), (2.3) は広義のワイブル型，(2.4) は2重指数型)．

数学的にはこれで良いのだが，実際はこの極限分布への収束が $n\to\infty$ のとき，決して速くない（場合によっては $\sqrt{n}^{-1}$ のオーダーである）点に注意しなくてはならない．

この点は，Chin, E.H. and J.F. Miller (1977) や筆者 (1979) らによって具体的に注意されている．（上記2人は limiting form of Fisher-Tippet Type（上記 (2.4)) is reached exceedingly slowly といったが他のタイプでもほぼ同様におそいである．）

最近，A.F. Jenkinson は融通性のある一般モデルを提示し，後で解説する四分位解析 (Quartile analysis) との比較を具体例で示した．

この理由は，“n が大きくないと，x_n に3種類の極限分布のどれが実際に適合するか分からないが，x_n の分布はこれらとあまりかけ離れた分布型にはならないであろう，n が大きくなれば，当然極限分布型のどれかを応用する分析方針をとるべきだ”．という点にある．その上，四分位解析とも併用可能な電子計算機向けの operational routine の開発も試みている．

A.F. Jenkinson は3パラメータ極限分布の一般型を

$$F(x)=\exp\{-(1-k(x-x_0)/\alpha)^{1/k}\} \qquad (2.5)$$

と書き，これが M. Frechet (1927) の「安定条件」（コーシー分布のように分布型の式，特性関数は存在しても，平均，分散が存在しないという不安定なものでないといった条件）を満たすことを指摘した．

ここで，α, x_0, k がパラメータであり，Jenkinson の換算変数 (reduced variate を仮にこう訳した) y を

$$y=-\frac{1}{k}\log\left\{1-\frac{k(x-x_0)}{\alpha}\right\} \qquad (2.6)$$

とおくと，明らかに x と y の換算関係は

$$x=x_0+\alpha\left\{\frac{1-\exp(-ky)}{k}\right\},$$

$$y=-\log\log F(x)^{-1} \qquad (2.7)$$

となる．この y への換算の意味は古く R. Von Mises (1936) により考察されているが，ここでは省略した．

問題は3つのパラメータ k, α, x_0 をどう解釈し，どう推定するかである．

まず k について考えると，容易に

(i) $k=0$ なら換算は線型式 $x=x_0+\alpha y$ であり，これを Fisher-Tippett Type I（つまり Gumbel の2重指数型）という．

(ii) $k<0$ なら換算は非線型式であり，これを Fisher-Tippett Type II という．

(iii) $k>0$ なら換算式はやはり非線型式であり，これを Fisher-Tippett Type III という．（第2図参照）

となることが分かり，(2.5) は3種類の極値極限分布のどれにでもなり得る一般性をもつことが確認される．

つぎに α は $k=0$ とした Jenkinson の2Pモデル (two-parameter model of extreme value distribtion, 7.2参照) における換算 $x\to y$ の比例係数であり，最後に x_0 はこの換算の定数であると理解できる．（この他の解釈，つまり確率分布としての型に対する解釈としてscale parameter, size parameter を持ち出す説明も一応

第1表　もとの分布と極値分布（V.T. Chow による）

もとの分布（確率密度）	極　値　分　布	極値 x の範囲
$x\to\infty$ のとき 0 に近づく exponential type 正規型：$f(x)=ce^{-ax^2}$ カイ2乗型：$f(x)=ce^{-bx}x^{n/2-1}$ 対数正規型：$f(x)=ce^{-a(\log x)^2}$ 指数型：$f(x)=ce^{-cx}$ ガンマ型：$f(x)=ce^{-bx}x^d$	$P(X\leqq x)=e^{-e^{-(a+x)/c}}$ $(\equiv F(x))$ $f(x)=e^{-e^{-(a+x)/c}}\times e^{-(a+x)/c}$ $a=\gamma\frac{\sqrt{b}}{\pi}\sigma-\mu$ $c=\frac{\sqrt{b}}{\pi}\sigma$ $\gamma=0.57721$	$-\infty<x<+\infty$ Type I
あるオーダー以上のモメントを持たないコーシー分布 $f(x)=\frac{1}{\pi}\frac{\lambda^2}{\lambda^2+(x-\mu)^2}$ $f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$	$P(X\leqq x)=e^{-(\theta/x)^k}$ $\theta=F(X_n)$ k=any one order of moments	$0\leqq x\leqq+\infty$ Type II
$f(x)=\begin{cases}f_1(x) & x\leqq\varepsilon_1\\ 0 & x<\varepsilon_1\end{cases}$ のように truncate された分布 $f(x)=cx^a(\varepsilon-x)^b$ $(c,a,b>0)$ といったベータ分布	$P(X\leqq x)=e^{-((x-\varepsilon)/(\theta-\varepsilon))^k}$ $k=f(x)$ の lowest derivative の order $\theta=F(X_n)$	$-\infty<x<\varepsilon$ Type III

可能である.）

つぎにこの3つのパラメータを順序標本換算値 y_1, $y_2,\cdots,y_n$ によって最尤法で推定する手法が示された．しかし，この一般化された3パラメータ極値分布を実際に適用するとして，3パラメータを推定する最尤法の結果に関し Jenkinson 自身，推定量の誤差分散行列に問題があることを次のごとく指摘している．

The variance-covariance matrix for the parameters shows some unstable characteristics, but experience indicates that we may use for a given x, the standard errors of estimates obtained for the same value x from the two parameter model.

（つまり理論的には誤差大きく不安定になり勝ちだが，実用性は一応あるという意味であろう.）

もとの分布の特性条件に応じて3種類の極限分布が導かれるといっても実際家にはピンとこないだろう．V.T. Chow (1964) が記述的に示したものをもとに，筆者が検討整理した結果を第1表に一括し，見やすくしておいた（多少疑問点もある）.

さらに WMO より寄贈されたレポートにより，$k=0$, $k<0$, $k>0$ に対応して $x\leftrightarrow y$ の換算関係 (2.7) をグラフに示すと第2図のようになる.

この第2図は，WMO のレポートによるもので原文そのままとしたが Type I の直線に比べ Type II と Type III の対照的な性格がよく反映されている.

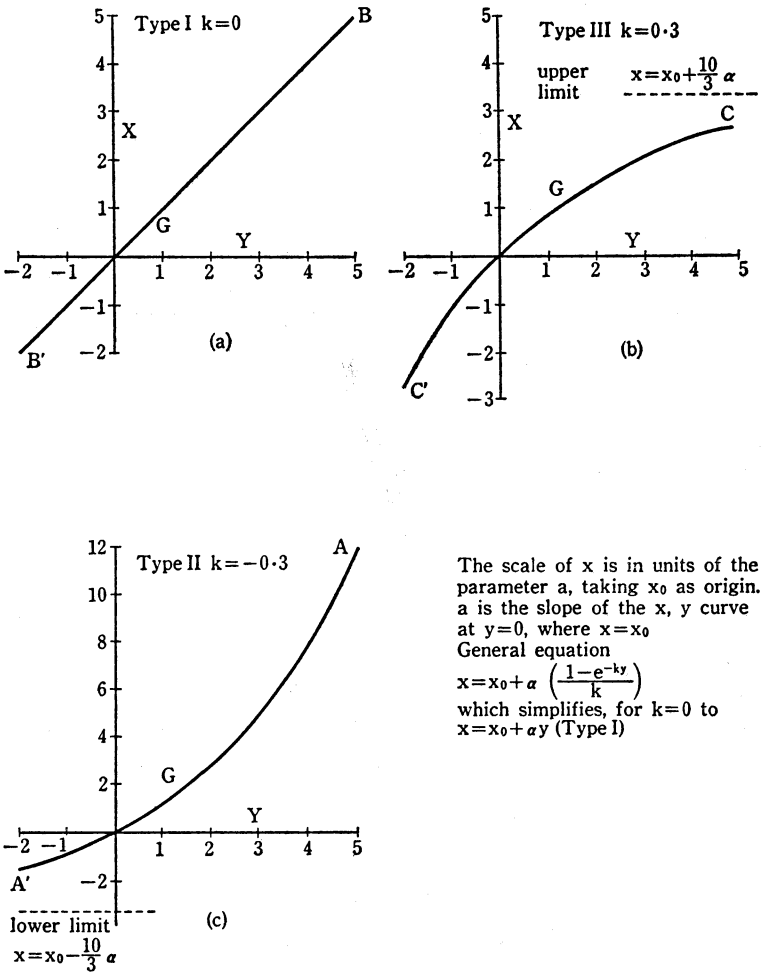
要するに，極限分布としては，第2図に示した3つの Type のどれかになるのだが，いずれもデータが十分多く（100以上），もとの分布状況を確実に見定め得る場合の結果であることを注意しておこう．（よってこれらは問題（b）への完全な解答ではない.）

3. 有限標本での極値統計量の分布

前章で説明したことは大きさ n の標本での極値統計量 X_n（確率変数）が $n\to+\infty$ のとき，どう分布するかの問題であり，現実には決して $n\to+\infty$ とはならないので，実用上はどこまでも有限標本とした上での極値を考察しなければならない．そこで有限標本に限定した取扱いが必要である.

（a）基本的出発点

一般に連続な確率変数 X の分布関数 $P(X\leqq x)\equiv F(x)$



第2図　Schematic representation of the Fisher-Tippett types of x, y diagrams (WMO レポート [13] による).

が与えられ，その密度関数 $f(x)=dF(x)/dx$ が $(-\infty, +\infty)$ で存在するとき，この母集団から得られた大きさ n の順序統計量（確率変数）

$X_1 \leqq X_2 \leqq \cdots \leqq X_n$

の同時密度関数（n 次元）は容易に

$$g_n(x_1, x_2, \cdots, x_n)=n!\ f(x_1)f(x_2)\cdots f(x_n) \quad (3.1)$$
$$-\infty<x_1<x_2<\cdots<x_n<+\infty$$

と書かれる．

これを基本的出発点として，多くの公式が解析的に導びかれている．

たとえば i 番目の順序統計量 X_i の密度関数は (3.1) の周辺分布密度として

$$g(x_i)=n_{n-1}C_{i-1}F(x_i)^{i-1}\{1-F(x_i)\}^{n-i} \quad (3.2)$$

だから，$X_i=X_n$ とおくとその最大（確率変数）X_n の分布の密度関数はよく知られているように，$i=n$ とおいて

$$g(x_n)=n\left(\int_{-\infty}^{x_n} f(x)dx\right)^{n-1} f(x_n) \quad (3.3)$$

と書かれ，有限な n を考える限り，明らかに $f(x)$, $f(x_n)$ の explicit form が決定的な役割を果たす．

さらに，順序統計量の間隔（spacing）の期待値は，

$$E(X_{i+1}-X_i)=\binom{n}{i}\int_{-\infty}^{+\infty}\{F(x)\}^i\{1-F(x)\}^{n-i}dx \quad (3.4)$$

$$i=1, 2, \cdots n-1$$

となるし，$X_1, X_2, \cdots, X_{n-1}$ を所与条件としたときの X_n の条件つき分布を示す密度関数は

$$f(x_n|x_1, x_2, \cdots x_{n-1}) = f(x_n)\{1-F(x_{n-1})\}^{-1} \quad (3.5)$$
$$x_{n-1} \leqq x_n < +\infty$$

となることも分かる．

この他の諸結果もいろいろあるが，ここでは省略する．

（b）極値予測に必要な理論的結果

予測に必要な条件つき期待値と分散の定式化をするため，$X_1, X_2, \cdots X_n$ を条件としたときの X_{n+1} の分布について考えると，その条件づき分布の密度関数は，もとの分布を $F(x)$，その密度関数を $f(x)$ として (3.5) の n を $n+1$ に置き換えて

$$f(x_{n+1}|x_1, x_2, \cdots, x_n) = f(x_{n+1})\{1-F(x_n)\}^{-1} \quad (3.6)$$
$$x_n \leqq x_{n+1} < +\infty$$

とし，条件づき期待値と分散は (3.6) からそれぞれ

$$E\{X_{n+1}|x_1, x_2 \cdots, x_n\} = \{1-F(x_n)\}^{-1}\int_{x_n}^{+\infty} x_{n+1} f(x_{n+1})\, dx_{n+1} \quad (3.7)$$

$$\mathrm{Var}\{X_{n+1}|x_1, x_2, \cdots, x_n\} = \{1-F(x_n)\}^{-1}\int_{x_n}^{+\infty} x^2_{n+1} f(x_{n+1})\, dx_{n+1} - E\{X_{n+1}|x_1, x_2, \cdots, x_n\}^2 \quad (3.8)$$

と表される．ここで実際の計算上困難なのは，これらの右辺の定積分項である．たとえば1例として，降水量のような Gamma 分布をするものを考えてみる．もとの分布がこの分布型の場合，密度関数は次の式で与えられる．

$$f(x) = \frac{\beta^\nu}{\Gamma(\nu)} e^{-\beta x} x^{\nu-1}, \quad 0 \leqq x < +\infty, \ \nu > 0, \ \beta > 0 \quad (3.9)$$

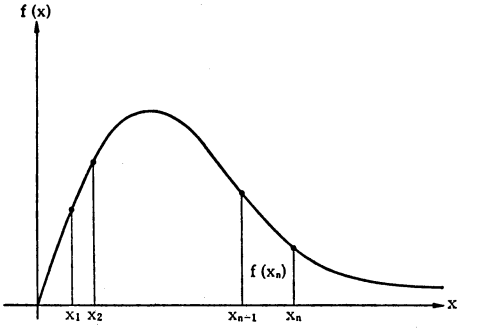
ここでパラメータ β と ν は，標本平均 $\bar{x}$ と標本分散 s^2 を用いることにより積率推定法を適用すれば

$$\nu = \frac{\bar{x}^2}{s^2}, \quad \beta = \frac{\bar{x}}{s^2}$$

として与えられる．従って条件づき期待値 (3.7) と分散 (3.8) を計算する場合，明らかに定積分値を explicit に数値計算可能な形で得ることはできない．

そこで我々は次のような近似的結果を導く手続きをとった．

分布関数 $F(x_n)$ の代わりにその期待値 $E\{F(x_n)\}$ で近似的に置き替えるために



第3図　ガンマ分布に対する近似モデル．

$$E\{F(x_n)\} = \frac{n}{n+1}$$

を用いることにする．

まず条件つき期待値を次の式で近似的に与える．

$$E\{X_{n+1}|x_1, x_2, \cdots, x_n\} \fallingdotseq (n+1)\int_{x_n}^{+\infty} x f(x)\, dx \quad (3.10)$$

しかしながら，上式右辺の定積分は x の関数で解析的に扱うのが困難であるので，x_n で切断された指数関数を用いて漸近近似をする．

第3図のように $x = x_n$ からスタートして減衰する曲線

$$f(x) = f(x_n)e^{-c(x-x_n)}, \ c > 0, \ x \in (x_n, +\infty)$$

で近似できるとすると

$$\int_{x_n}^{+\infty} f(x)dx = f(x_n)e^{cx_n}\int_{x_n}^{+\infty} e^{-cx}\, dx = \frac{f(x_n)}{c} \quad (3.11)$$

であり，さらに

$$\int_{x_n}^{+\infty} f(x)dx \fallingdotseq \frac{1}{n+1} \quad (3.12)$$

であるから，(3.11) と (3.12) より $c \fallingdotseq (n+1)f(x_n)$ という近似的関係式を得る．ゆえに，条件づき期待値は，

$$E\{X_{n+1}|x_1, x_2, \cdots, x_n\} \fallingdotseq (n+1)\int_{x_n}^{+\infty} x f(x)dx \fallingdotseq x_n + \frac{1}{\hat{c}} \quad (3.13)$$

ただし $\hat{c} \fallingdotseq (n+1)\ f(x_n)$

の形で表され，結局以下のようにして逐次に予測値 $\hat{X}_{n+1}, \hat{X}_{n+2}, \cdots, \hat{X}_{n+j}$ をもとめ得る．

$$\left.\begin{aligned}
&E\{X_{n+1}|x_1,x_2,\cdots x_n\}\\
&\quad\fallingdotseq x_n+\frac{1}{(n+1)f(x_n)} &&\to\hat{x}_{n+1}\\
&E\{X_{n+2}|x_1,x_2,\cdots x_n,\hat{x}_{n+1}\}\\
&\quad\fallingdotseq\hat{x}_{n+1}+\frac{1}{(n+2)f(\hat{x}_{n+1})} &&\to\hat{x}_{n+2}\\
&E\{X_{n+j}|x_1,x_2,\cdots,x_n,\hat{x}_{n+1},\cdots,\hat{x}_{n+j+1}\}\\
&\quad\fallingdotseq\hat{x}_{n+j-1}+\frac{1}{(n+j)f(\hat{x}_{n+j-1})} &&=\hat{x}_{n+j}
\end{aligned}\right\}\quad(3.14)$$

予測誤差を示す条件づき分散についても同様の近似が可能である（定式化省略）．ただし，この近似は2つの近似関係（3.10）および（3.12）が実際に適用されると考えられるときにのみ応用される．

もし，もとの X の分布が一様分布で，その密度関数が，

$$\left.\begin{aligned}
f(x)&=\frac{1}{\theta},\ 0<x<\theta,\ 0<\theta<+\infty\\
&=0\qquad\text{その他の } x,\theta
\end{aligned}\right\}\quad(3.15)$$

であるなら，容易に次のような結果が得られる．

$$\left.\begin{aligned}
g(x_n)&=n\theta^{-n}x_n^{n-1},\ 0<x_n<\theta\\
&=0\qquad\text{その他の } x,\theta
\end{aligned}\right\}\quad(3.16)$$

これに対応する分布関数 $F_n(x_n)$ は

$$\left.\begin{aligned}
F_n(x_n)&=\int_0^{x_n}n\theta^{-n}x^{n-1}dx=\left(\frac{x_n}{\theta}\right)^n,\ 0<x_n<\theta\\
&=1,\qquad\theta\leqq x_n
\end{aligned}\right\}\quad(3.17)$$

よって，$x_1,x_2,\cdots,x_n$ を与えて X_{n+1} の条件づきの諸確率（たとえば X_{n+1} が x_n と 0.9θ の間に入る確率など），条件づき期待値と分散 $E(X_{n+1}|x_1,x_2,\cdots,x_n)$，$\mathrm{Var}(X_{n+1}|x_1,x_2,\cdots,x_n)$ を解析的にもとめることは容易である．

さらに，もとのXの分布が指数分布（たとえば，特定の台風における最大風速の単純な度数分布のモデル）であるときも，同様に解析的表現が可能である．（もとの分布がこうした1母数を持つ単純な確率分布の場合は学生の演習問題水準なので，ここでは省略した．）

こうしたことから，最近の統計気候学国際会議での一傾向として，

「比較的単純な極値予測問題に対しては，もとの X の確率分布を一様分布か指数分布と仮定するか，そのいずれかに適当に変換して処理できる」

という風潮が見られるようになった．

しかし，いずれも1つだけの母数しかもたないこのような単純な分布におきかえ得るほど，現実の極値予測はやさしくないことも多い．

（ただ単純に処理できそうな場面に，わざわざ複雑な極値予測手法をもちこむのも賢明ではない．要するにデータをみて判断することになる．）

4. 極値確率紙のモデル

ここでは問題（c）を主として取り上げる．この確率紙は元来，面倒な数式を用いないで，稀現象や極値のデータをグラフ上で処理したいという要望に対して生まれたものである．

一般に，確率紙とは横軸に標本の値 $x_1,x_2,\cdots,x_n$ をとり，縦軸にこれらをこえない確率 $P(X\leqq x_1)$，$P(X\leqq x_2),\cdots,P(X\leqq x_n)$ をとって n 個の点 $\{x,P(X\leqq x_i)\}$，$(i=1,2,\cdots,n)$ をプロットするためのグラフ用紙である．

そこでまず，プロットの根拠から説明しよう．第4図にはプロットの基本的モデルが示されているが，ここでは縦軸に i 番目の順序統計量 $X_{(i)}$ の再現期間（return period）を併記することも多いことを示している．ただし極値確率紙（Extreme probability paper）に対しては通常の標本と区別するため，順序統計量をすべて X_i でなく，$X_{(i)}$ とした．

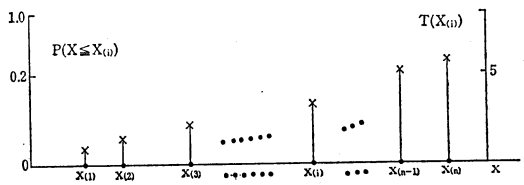
一般にある X が $X_{(i)}$ 以下である確率 $P(X\leqq X_{(i)})\equiv F(X_{(i)})$ を $X_{(i)}$ の非超過確率（non-exceedance probabilily）といい，$1-F(X_{(i)})$ を $X_{(i)}$ の超過確率という．このとき，この逆数

$$T(X_{(i)})=\frac{1}{1-F(X_{(i)})}\quad(4.1)$$

は $X_{(i)}$ をこえるものが存在する期間の長さを示すので，これは $X_{(i)}$ の再現期間（return period）である．

この基本モデル図では縦軸，横軸の目盛をとくに明示していないが，現実に用いられるものはともに等間隔でない非線型目盛で，プロットされた点が単調増加型直線状のとき，3タイプの極値極限分布のいずれかに合うように工夫されたものである．

一般に大きさの順序標本（確率変数）が連続な無限母集団から得られたとして，それを



第4図　プロットルールの基本モデル．

$X_{(1)}\leqq X_{(2)}\leqq X_{(3)}\leqq\cdots\leqq X_{(i)}\leqq\cdots\leqq X_{(n-1)}\leqq X_{(n)}$

と書き，確率変数 X が $X_{(i)}$ 以下である確率を

$P(X\leqq X_{(i)})$, $(i=1,2,\cdots,n)$

と書くことにする．（同じ $X_{(i)}$ が2度あらわれることもあるので，一応等号をすべて入れることにする．）

プロット手法は，大別して次の3つに分類されるので，その考え方とそれらの根拠を以下に示す．

（A）経験的方法

（i）California plot;

n 個中 $X_{(i)}$ 以下のものは i 個あるので

$$P(X\leqq X_{(i)})=i/n,\ i=1,2,\cdots,n \qquad (4.2)$$

とする．しかしこれは $i=n$ のとき，

$P(X\leqq X_{(n)})=n/n=1$, $P(X>X_{(n)})=0$

となり，$X_{(n)}$ をこえる標本は標本のとり方に関係なく，絶対にあらわれないこととなり，連続な無限母集団から有限個の標本抽出をしたという点で明らかに妥当でなく，今は用いられなくなった．（California 河の流量データ解析で，昔用いられた．）これを補正したのが Hazen のプロット方式である．

（ii）Hazen plot:

これは $P(X\leqq X_{(i)})=i/n$ の右辺がもつ前記のような欠点を補正するため，右辺分子 i をこの代わりに $i-1/2$ としたもので，その根拠は $X_{(i)}$ より小さい $X_{(i-1)}$ 以下のものは確かに $i-1$ 個あるが，$X_{(i)}$ については「折半した形」という意味で $1/2$ を加えることにし，$(i-1)+1/2=i-1/2$ を用いるもので，結局

$$P(X\leqq X_{(i)})=(i-1/2)/n=(2i-1)/2n \qquad (4.3)$$

とする方法によるプロット方式でかなり広く用いられてきた．

（B）期待値的方法

これは順序統計量 $X_{(i)}$ はもとの母集団から大きさ n の順序標本を抽出するとき，その取り方により違い得る確率変数だが，その「非超過確率 $P(X\leqq X_{(i)})$ という確率変数の期待値」がもとの母集団分布モデルの関数形に関係なく $i/(n+1)$ となることを利用する方法で Thomas plot とも呼ばれ，

$$P(X\leqq X_{(i)})=i/(n+1),\ i=1,2,\cdots,n \qquad (4.4)$$

となる．これはまた，第4図で示されたような標本による $n-1$ 個の横軸区間 $(x_{(1)}, x_{(2)})$, $(x_{(2)}, x_{(3)})$, $\cdots$, $(x_{(n-1)}, x_{(n)})$ の前後に $(-\infty, x_{(1)})$, $(x_{(n)}, +\infty)$ の2つを加えて $n+1$ 個の区間で全領域をカバーしたとき，どの区間に入る確率も等しく $1/(n+1)$ と考えることを根拠として上の Thomas plot を導いたと説明する向きもあるが，それはやや便宜的な根拠説明で正確ではない．

従って，むしろ順序統計量に関する定積分表現による期待値的考慮から導かれたとする方が，統計的には説得力がある．（証明略）

第2表　パラメータ a, b に関する若干の数値例
（WMO による）

プロット名	a	b
Chegodayev	0.3	0.4
Blom	3/8	1/4
Tukey	1/3	1/3
Gringorten	0.44	0.12
Jenkinson	0.31	0.38

（C）数値実験的方法

これは正規分布とか他の適当な連続変量の確率分布から発生された適当な擬似乱数（現実に得られるのは擬似乱数で，真の乱数はあり得ない！）$n=50, 100, \cdots,$ 10000個をもとに，実験的関係式

$$P(X\leqq X_{(i)})=(i-a)/(n+b) \qquad (4.5)$$

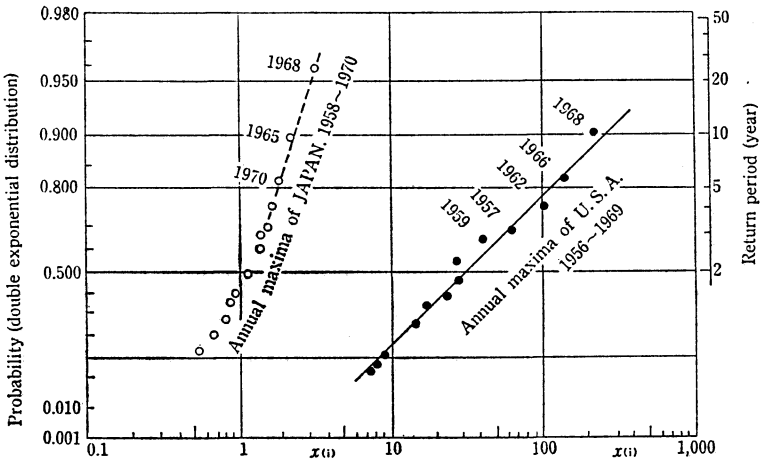
を想定し，正の小数 a, b を極値確率紙（主として2重指数型）上でこの乱数順序標本データがもっともよく直線上にのるようにもとめたもので，この具体的な実験結果のいくつかを第2表に示しておく．

この場合，極値確率紙の変数表示目盛（横軸），確率表示目盛（縦軸）をどう選択するかで小数 a, b はちがってくるが，ここでは横軸を線型目盛，縦軸を2重指数確率分布型が適合するとき，直線上にプロットされた点が並ぶような2重指数確率目盛が採用されている．

こうした数値実験的方法が提案された理由は，標本数 n が大きくなってもその中での極値 $X_{(n)}$ の分布が極値極限分布へ収束するのがかなりおそいため，2つの補正 a, b が必要となるからで，各プロットの優劣はなかなか決めにくいが，一応のスタンダードとして WMO（水文気象のワーキンググループ）では水文学的データに対して第2表の最下欄にある Jenkinson のプロットを実際家にすすめている．

上記の3つのプロット方式に関するこれまでの検討結果を要約すると Gumbel 型（Fisher-Tippett I型）なら Jenkinson のプロット方式がよく，Weibull 型なら Thomas のプロット方式（4.4）がよいとなっている．

いずれにしてもこの根拠については有限な n 個のデータ（つまり n 年間の年最大値を収集したデータ）を用いる限り，完全に問題が解明された訳ではなく，数理統計学



第5図（a）　2重指数確率紙へのある種の災害記銀をプロットした例（角川正義氏の御教示による）.

の理論面での専門家によっても決定的な解答案がないのが現状である．よって筆者は期待値的方法（Thomas のプロット方式）をどの分布型にも利用し，もっとも直線が適合しそうな極値極限分布型（3つの型のうち1つ）を導入することを現段階ではすすめている．

5. 極値確率紙の実際利用と極値分布

WMO では前記ワーキンググループが Maximum floods の評価分析手法確立を目指して発足し，1972年に Technical Note として総合的レポートを提示した．約300ページにのぼる厖大なもので，ここでは一部分だけを抽出し，紹介しよう．これは問題（A），（C）への具体的解答である．

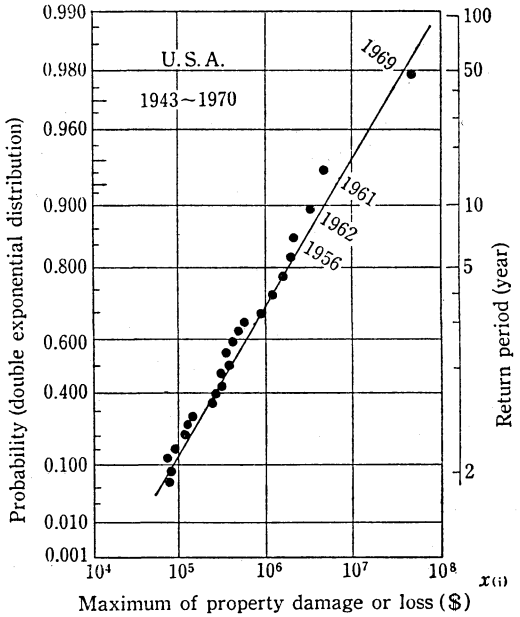
横軸に順序標本をとり，縦軸に非超過確率をとって，確率紙により分析することを主眼とする．

一般に大きさ n の順序標本

$X_1 \leqq X_2 \leqq \cdots \leqq X_i \leqq \cdots \leqq X_{n-1} \leqq X_n$

に対して，分布 $F(x)$ の値を考えるとき，$x=X_i$ に対する $F(x)$ の値は $(i-1)/n$ と i/n の間であるとする．このとき $F(X_i)$ の '中央値' としては $(i-0.31)/(n+0.38)$ が十分近いと考えられた．他方 N.N. Chegodayev (1953) は $(i-0.3)/(n+0.4)$ を提示し，A. Hazen (1930) は $(i-0.5)/n$，I. Gringorten (1963) は $(i-0.44)/(n+0.12)$ をそれぞれ提示したことを前章で示した(いずれも中央値的でなく平均値的の意味をもつ由)．

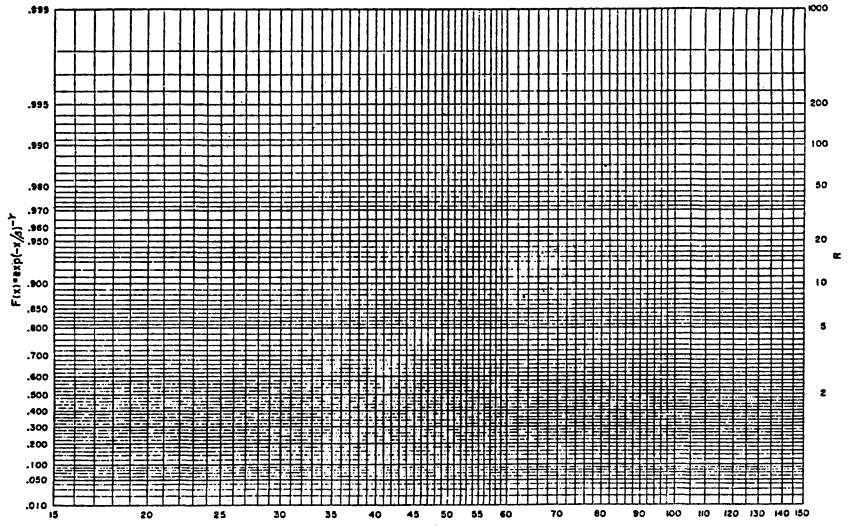
前記ワーキンググループは $(i-0.31)/(n+0.38)$ を種々の理由をあげてすすめている．そこで，$X_1, X_2, \cdots,$



第5図（b）　アメリカでのある災害記録の2重指数確率紙へのプロット例（日本原子力研究所角川正義氏の御教示による）.

$X_i, \cdots, X_n$ に対し，それぞれ，$F(X_1)=0.69/(n+0.38)$，$F(X_2)=1.69/(n+0.38)$，$\cdots$，$F(X_i)=(i-0.31)/(n+0.38)$，$\cdots$，$F(X_n)=(n-0.31)/(n+0.38)$ を対応させると，再現期間 $T(x_i)$ と $F(x_i)$，y_i との関係は，

$$T(X_i)=\{1-F(X_i)\}^{-1},\ i=1,2,\cdots,n \qquad (5.1)$$



第6図　Fisher-Tippett Ⅱ型の極値確率紙の1例（Thom, H.C.S. による）.

$$y_i=-\log\log\{T(X_i)/(T(X_i)-1)\},\quad i=1,2,\cdots,n \tag{5.2}$$

となり，縦軸の等間隔目盛に X_i，横軸の等間隔目盛に y_i をそれぞれ記入し，この図上で $(X_i, y_i)(i=1,2,\cdots,n)$ を記入すれば結果として n 個の点がプロットされる.

このプロットされた点群に対して，3つの曲線型

$$\left.\begin{aligned} &X=a+by \quad \text{(Fisher-Tippett Ⅰ型)}\\ &X=a+be^{cy} \quad \text{(Fisher-Tippett Ⅱ型)}\\ &X=a+be^{-cy} \quad \text{(Fisher-Tippett Ⅲ型)}\end{aligned}\right\} \tag{5.3}$$

のいずれかを適合させ，どれかもっともよく適合するものを選んで，それをもとに外挿を試みるのが極値予測の図式分析法の基本である（第2図参照）.

（ワーキンググループは多くの具体例とモデル，既往の研究集約とそれぞれに対する見解を述べているが，ここでは省略する.）

要するに(5.3)のような曲線あてはめ(Curve fitting)とその延長（extrapolation）の問題として，極値予測を図上で行うことをこのグループがすすめている由.

これは実際家向けのマニュアル提供に重点をおいたからであり，理論面は E.J. Gumbel (1958) および A.F. Jenkinson の書物（1969年に刊行，その後改訂されたもの）と彼の一連の研究（1955～1969）に負っている.

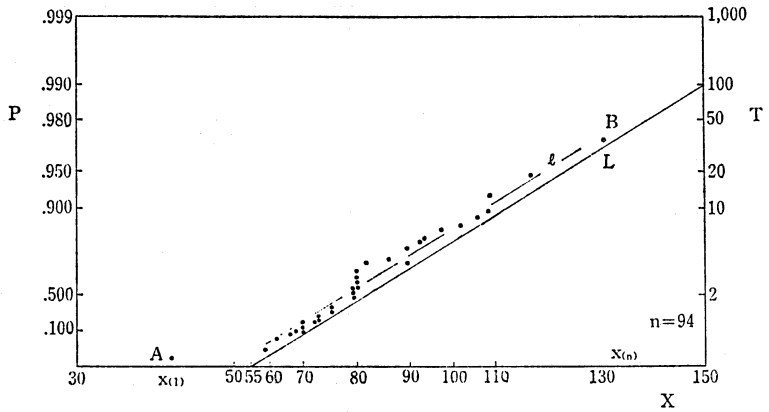
一方実際利用を簡単に説明するため，まず日本原子力研究所角川正義から頂いた2重指数確率紙への見やすい記入例をあげよう（第5図）.

つぎに Fisher-Tippett Ⅱ型の極値確率紙モデルとその記入例を第6図，第7図に示す.

実際的使用に便利な2重指数確率紙は，日本規格協会その他で販売されている.

第7図はアメリカで計測された，ある自然環境要因 X のデータ $n=94$ 個からつくられた順序標本データ $x_{(1)}, x_{(2)},\cdots, x_{(93)}, x_{(94)}$ をもとに，Fisher-Tippett Ⅱ型が比較的よく適合することを示す一種のモデル的具体例の紹介である.（つまり $n=94$ 個の全データをプロットすると重複があるので全データが記入されている訳ではない.）横軸には X としてこの環境要因順序標本を不等間隔目盛で示し，縦軸左側は Fisher-Tippett Ⅱ型が適合すれば直線的データ配列となるような非超過確率 $P\equiv P(X\leqq x_{(i)})$ の目盛，縦軸右側は再現期間（return period）$T=1/\{1-P(X\leqq x_{(i)})\}$ の P に対応する不等間隔目盛をとってある．明らかに最小値 $x_{(1)}$ を示すA点と最大値 $x_{(n)}$ を示すB点を除き，ほとんど左下から右上に向う単調で増大する傾向直線 l が適合するように見られる．さらにその下側直線 L は，この下には全くプロットされた点がないような一種の下限的境界直線を示す.

こうした「直線の式」は縦軸目盛特性が与えられれば容易に定式化されるであろうが，問題点としては，この第7図のように直線からみてとび離れた点 A，B（極大か極小であることが多い）が存在するとき，および，直線 l でなく，共通の中間的1点を共有する2つの折線が考えられるときに，合理的かつ適切な処理手法は何かということであろう.



第7図　Fisher-Tippett Ⅱ型極値確率紙への記入例（Wantz, J.W. によるが原論文中の図を少し変形拡大したのでやや不正確である）.

後者については1つの解決策として「複合型極値分布モデル」の適用がある.（たとえば，複合型 Weibull 分布適用例が石原健二により示された.）

前者のように $x_{(1)}$ または $x_{(n)}$ がとびはなれた位置にある場合はどうすればよいか.

この点の合理的処理法は今のところ未開発であるが，次のことが手法開発のためのヒントを与える.

（a）極値極限分布のパラメータ推定量（確率変数）の推定誤差を考慮して，第7図のような直線 l の信頼限界つまり順序標本という特殊な変量に見合う特殊な信頼双曲線を設定しその限界内にプロットされた点が入っていればよいとする.

（b）いずれかの確率紙上にプロットされた記録値の集団に対して，「データ自身の信頼限界」（あてはめた通常の直線回帰式に対する一種の許容限界（tolerance boundary））内に入っていればよいと考える.

（c）新たに別の有限極値分布を考察する.

このように，プロットされた順序標本データに対し，その信頼性とか誤差を評価するような限界をつける研究が I.I. Gringorten（1963～）以来，少しずつすすめられてきた. そこでこの点をなるべく実際に統計処理する立場からみて分かりやすい形で要約しよう.

6. 極値確率紙と包絡線（Envelope）

極値確率紙上に実際の順序標本データをプロットした場合，前章で述べたような意味での信頼限界や許容限界を設定することは極値予測の立場からも要請されていた. それに対する1つの考え方が I.I. Gringorten によって示されているので，その基本的考え方を紹介する.

いま，大きさ n の順序標本の値を小さい方から

$$x_{(1)}, x_{(2)}, \cdots, x_{(i)}, \cdots x_{(n-1)}, x_{(n)}$$

とし，これが母集団確率分布 $P(X \leqq x_{(i)}) \equiv F(x_{(i)})$ から得られたとするとき，記号的に

$$F_{i|n} \equiv F(x_{(i)}) \equiv P(X \leqq x_{(i)}) (\equiv P)$$

とかく（Gringorten の記号）.

たとえば標準正規分布 $N(0, 1^2)$ を母集団とすると，最大標本変量 $X_{(n)}$ が $n=15$ の標本から抽出されたとき，確率としては，

$$P(0.77 \leqq X_{(n)} \leqq 2.93) = 0.95 \qquad (6.1)$$

とかかれ，同様に2番目に大きい標本変量 $X_{(n-1)}$（$n=15$ とすると $X_{(14)}$）に対して

$$P(0.47 \leqq X_{(n-1)} \leqq 2.13) = 0.95 \qquad (6.2)$$

となるであろう. これはもとの連続変量確率分布が正規分布であるような母集団から大きさの n 標本をとったときの最大 $X_{(n)}$，2番目に大きい $X_{(n-1)}$ は，どんな区間内に，どれだけの確率で存在し得るものであるかを数値例として示したものである. このことは I.I. Gringorten が示唆しているように，「もとの母集団分布が与えられ，ここから大きさ n の標本を抽出したとき，このうちの小さい方から i 番目の標本変量 $X_{(i)}$（$i=1, 2, \cdots, n-1, n$）が存在し得る区間をある確率で示すことが，望ましい順序標本変量の統計処理法である.」という主張の具体例で，彼はこの区間を連続曲線的に結んで包絡線（Envelope）とよんだ.

この包絡線づくりはプロットされた点群の許容限界づくりに1つの有力な示唆を与える.（計算は基本的に2

項分布に基づく確率計算で以下のようになろう.)

よく知られているように，もとの連続な母集団確率分布と密度関数をそれぞれ $P(X\leqq x)\equiv F(x), dF(x)/dx\equiv f(x)$ とかくとき，大きさ n の標本において小さい方から i 番目の順序統計量をあらためて $X_{(i|n)}$ とかく.(このような順序統計量とは i と n とに依存する確率変数なので，こうかく方が正確であろう.)

この非超過確率 $F_{i|n}$，確率密度 $f(x_{(i|n)})$ が容易にそれぞれ

$$F_{i|n}=\int_{-\infty}^{X_{(i|n)}} f(x)dx,$$
$$dF_{i|n}=f(x_{(i|n)})dx_{(i|n)} \qquad (6.3)$$

となるので，これを用いた $P_{(i|n)}\equiv P(X\leqq X_{i|n})$ の確率微分と $F_{i|n}$ の期待値はそれぞれ

$$\left.\begin{aligned} dP_{i|n}&=\frac{n!}{(i-1)!\,(n-1)!}F_{i|n}{}^{i-1}\cdot \\ &\quad (1-F_{i|n})^{n-i}dF_{i|n} \\ E(F_{i|n})&=\int_0^1 F_{i|n}\,dP_{i|n}=i/(n+1) \end{aligned}\right\} \qquad (6.4)$$

で，後者がいわゆる Thomas plot の根拠であり，$F_{i|n}$ や $f(x_{(i|n)})$ の具体的関数型に関係なく (6.4) 第2式の右辺の値が求められるが，$F_{i|n}$ そのものの計算にはもとの分布をどうしても考慮せざるを得ない.

そこで包絡線を作るには，まず上記のような $F_{i|n}$ の期待値 $i/(n+1)$ をプロットして，つぎに $F_{i|n}$ の可能な上限 $U(F_{i|n})$ と可能な下限 $L(F_{i|n})$ 内に確率変数 $F_{i|n}$ が入る確率

$$P(U(F_{i|n})\geqq F_{i|n}\geqq L(F_{i|n}))\equiv P^* \qquad (6.5)$$

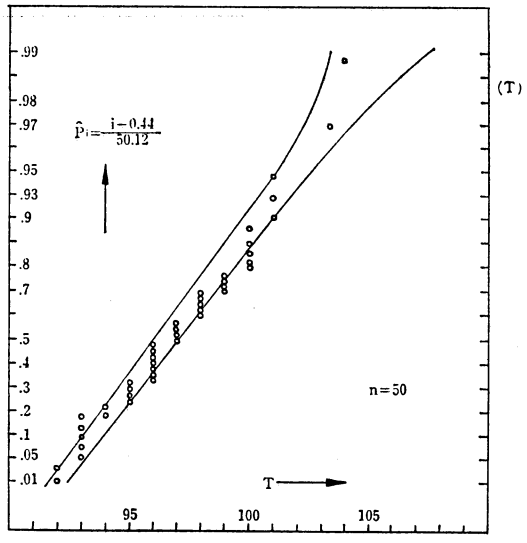
が97.5%とか75%とかになるように $U(F_{i|n})$ と $L(F_{i|n})$ を求める作業の図表を作成しておくと便利である. Gringorten らによると，USA の NBS (National Bureau of Standards) 1953年版 Table 2 には標準正規分布の場合の非超過確率 $P(X\leqq x)\equiv P$ に対し，

$$P=0.001(0.0001)0.005(0.001)0.988(0.0001)$$
$$0.9994(0.00001)0.99999$$

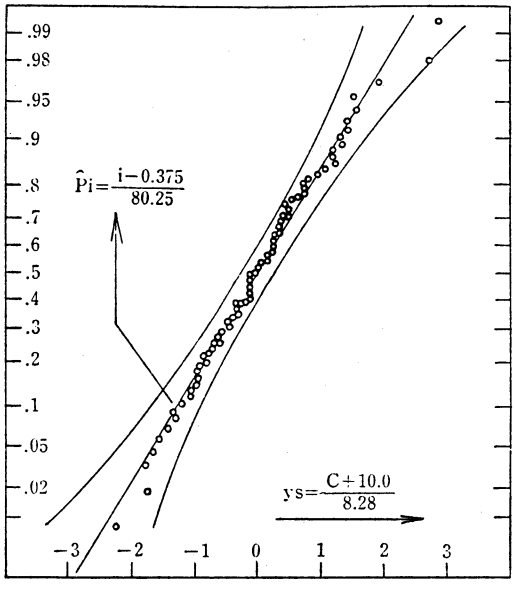
のときの x が詳しく示されているので彼は，(6.1)～(6.5) を根拠に工夫をこらして極値確率紙上に包絡線を作った．ここにそのうちの2例だけを紹介する.

その第1例は第8図で50個の順序標本データを横軸に等間隔目盛でとり，縦軸に Gringorten 方式の確率 P をとり75%包絡線をつけた．この包絡線外に出るデータも13個あることを示す.

その第2例は第9図で $n=80$ 個のある別要因 C の順序標本データを標準化し，y_s の等間隔目盛で横軸にとり，



第8図　ある要因Tの極値確率紙プロット例と75%包絡線 (Gringorten, I.I. による).



第9図　標準化された変量 y_s の確率紙プロット例と95%包絡線 (Gringorten, I.I. による).

縦軸に Blom による目盛で $\hat{P}_i$ をとり，95% 包絡線をつくったもので，信頼双曲線の形状を示し，全データがこの限界内にある.

7. その他の再現期間値予測法

以上の他に数多くの予測技法が開発工夫されている．それらを列挙すると，

四分位解析（Quartile Analysis 略して QA という）

2P 法（Jenkinson の3パラメーター一般モデルで $k=0$ とした2パラメーターによるもの）

曲線あてはめによる外挿法

カルマン・フィルター理論の応用手法

物理要因重合確率による手法

などである．これらすべてを詳述する必要もないし，紙面もないので，ここでは QA 法と 2P 法を具体例中心に説明し，他の技法は「考え方」だけにとどめる．これらは一応問題（a）に対する直接的解答である．

(1) QA 法

これは A.F. Jenkinson が前記の WMO 極値解析総合研究報告書を作成した際，有効な技法としてすすめたものの1つである．（とくにこの手法だけを取り上げた論文は見当たらない．）

この技法を手順の形で要約し，具体例を2つあげよう．

QA 法とは次の手順による極値予測技法である．

〔手順1〕 大きさ n の順序標本を大体4等分して各4分位平均 QM 1, QM 2, QM 3, QM 4 を計算する．

〔手順2〕 再現期間 (return period) T_i を 5, 10, 20, 30, …, 200, 300, …, 1000 としこれに対応する換算変数 y_i を

$$y_i=-\log\log\{T_i/(T_i-1)\}\quad i=1,2,\cdots \tag{7.1}$$

とする．(T_i は (10, 1000) の間でどうとってもよく，たとえば設計基準から予めきめられる．)

〔手順3〕 QM 4, QM 3 より，換算係数

$$\alpha_1=(\mathrm{QM}\,4-\mathrm{QM}\,3)/1.55 \tag{7.2}$$

をもとめる．

〔手順4〕 再現期間 T_i に対応する外挿推定値（予測値）x_i は次式でもとめられる．

$$x_i=\mathrm{QM}\,4+\alpha_1(y_i-2.32)\quad i=1,2,\cdots \tag{7.3}$$

〔手順5〕 上記の4分位平均を単なる算術平均でなく，幾何平均としてもよく，そのときの推定値 x_i は $\alpha_2=\log(\mathrm{QM}\,4/\mathrm{QM}\,3)/1.55$ を用い，

$$x_i=\mathrm{QM}\,4\exp\{\alpha_2(y_i-2.32)\}\quad i=1,2,\cdots \tag{7.4}$$

とすればよい．

上記手順は明らかに一応極値予測のための経験的技法であるが，手順3～5の根拠を一応示しておく．換算変数 (reduced variate) y をもとに例えば2重指数型極値分布モデルでの4分位平均 (quartile mean) を示すと，第3表のようになっているからである．

第3表　換算された4分位平均値（2重指数型の場合）

n	QM 1	QM 2	QM 3	QM 4
16	−0.74	0.04	0.79	2.22
32	−0.77	0.03	0.78	2.27
>40	−0.80	0.02	0.77	2.32

しかし，予測ではなく同定 (identification) の場合はこの手順3～5は多少変更されなければならない．

一般に「同定」とは片山徹 (1983) の3頁にあるように，「時系列や順序標本のモデルで，係数を最尤法か最小2乗法で決めること，および2変数時系列で相互相関をもとめ，モデル確定をすること」である．「予測」は未知な最大値や将来値を推定することで，「予測」は時間ズレの関係を重視するが「同定」では全く無視するか，軽視するというちがいがある．同定のコンピュータプログラムも片山の書物にある．

（手順3，5の 1.55 は明らかに QM 4—QM 3の第3表における値であり，手順4，5の 2.32 は QM 4 の第3番目の値である．）

(2) 2P 法

つぎに2P モデル（2パラメータ極値分布モデル）における計算手順を示そう．

〔手順1〕 上記 QA 法で得られた α_1 または α_2 を α の初期値とし，x_0 の初期値を QM 2 とする．

〔手順2〕 α, x_0 のこの初期値を2P モデルが所与データによく合うように修正する．(2P モデルをもっとよく適合させる最小2乗法は不可能なので，Newton-Raphson による反復法の利用をする．)

〔手順3〕 上記修正後の α および x_0 を用いて所要の再現期間 T_i に対応する $y_i=-\log\log\{T_i/(T_i-1)\}$ を計算し，T_i に対する極値予測値 x_i を

$$\hat{x}_i=x_0+\alpha y_i\quad i=1,2,\cdots \tag{7.5}$$

とし，x_i の予測誤差を

$$s(x_i)=(\alpha/\sqrt{n})\{1+0.6079(y_i+0.4228)^2\} \tag{7.6}$$

とする．（この予測誤差は標準偏差の意味をもつ）

これが要約された2P モデルの計算手順であるが，この方法には不明な点が2カ所ある．その1つは手順2の修正の妥当性が十分確認されていないこと，および，第2の点として手順3の極値推定量（予測値）x_i の分布

が正確に分からない（おそらく正規分布ではないだろう）のに，標準偏差相当の尺度で誤差評価してよいかである．

要するにこの方法の根拠は A.F. Jenkinson (1955) が示した前記3パラメーター一般モデル((2.5))

$$P(X(\leqq x)\equiv F(x) = \exp\{-(1-k(x-x_0)/\alpha)^{1/k}\} \qquad (7.7)$$

で，$x \leftrightarrow y$ の換算関係

$$\left.\begin{aligned} x &= x_0+\alpha\{(1-\exp(-ky))/k\} \\ y &= -k^{-1}\log\{1-k(x-x_0)/\alpha\}, \\ y &= \text{loglog}\, F(x)^{-1} \end{aligned}\right\} \qquad (7.8)$$

が $k=0$ のときパラメータは α, x_0 の2つとなり，そこで F の代わりに再現期間 T を用いることであり，Jenkinson はこれを2Pモデルと呼んでおり，その換算は

$$x = x_0+\alpha y$$

で示される直線になることを既に述べておいた．QA 法（四分位解析法），2P法（2-パラメーター法）の計算手順について数値的に具体例をあげよう．

例1　毎年計測されているある連続な環境変量のうち年間最大であるものを28年間に亘って記録した順序標本データが下のようであった．

2.20 2.60 2.69 2.84 3.14 3.22 3.33
3.48 3.49 3.50 3.59 3.62 3.75 3.80
3.84 4.05 4.28 4.75 5.34 5.35 5.57
5.64 6.00 6.51 6.98 7.09 9.50 10.04

そこで4分位平均は，それぞれ QM 1=2.85, QM 2=3.60, QM 3=4.69, QM 4=7.24
と計算される．)これらの4分位は Thomas plot などとくらべ少しちがうので原文のままとした)．

2Pモデルの計算手順に従い，数値計算結果を述べる．まず初期値として

$$\alpha = (\text{QM3}-\text{QM1})/1.57 = (4.69-2.85)/1.57 \fallingdotseq 1.20$$
$$x_0 = \text{QM2} = 3.60$$

をとる．α の計算に入ってくる除数1.57は2重指数型極値分布モデルの4分位平均より得られる．次に Newton-Raphson 法による反復修正計算を用いて，この例の場合，次の結果を得ている．

$\alpha=1.3205$, $x_0=3.8130$

予測を示す第4表によると，再現期間が10年第4表の M10 より短いものについては QA 法が実測に合う4分位点を用いるので信頼できそうであり，2P法では QA 法に比較してやや under estimate になっている．

例2．ある自然現象の57年間におけるデータがあり，

第4表　極値予測における各手法の比較例 (1)
(Jenkinson による)

	2P	S.E.	QA
2*M*			2.85
1*M*			3.60
*M*2			4.12
*M*5	5.79	0.45	5.83
*M*10	6.78	0.58	7.24
*M*20	7.74	0.71	8.68
*M*50	8.97	0.88	11.26
*M*100	9.89	1.01	13.68
*M*1000	12.93	1.45	26.07

（注）2*M*：1年に2回
*M*10：10年に1回
S.E.: 標準誤差 (Standard Error)

第5表　57年間実測データの統計値

QM1	QM2	QM3	QM4	H4	H3	H2	H1
55.9	62.0	65.8	73.2	75	77	83	90

ただし，H_1, H_2, H_3, H_4 はそれぞれ既存データの最大値，2番目，3番目，4番目に大きい値を示す．

第6表　極値予測における各手法の比較例 (2)
(Jenkinson による)

	2P	S.E.	QA
2*M*			55.91
1*M*			61.95
*M*2			63.88
*M*5	70.33	1.53	69.25
*M*10	75.14	1.96	73.19
*M*20	79.75	2.40	76.50
*M*50	85.72	2.98	81.25
*M*100	90.19	3.43	84.81
*M*1000	104.97	4.92	96.56

その4分位平均と4つの極値（最大のものから4位までの順序統計量）が第5表のようになっていた．

そこで例1と同様にして，2Pモデルの α, x_0 の反復計算による修正を行った結果，次のようになった．

$\alpha=6.403$, $x_0=60.72$

ここで前例と同様に，2P, QA 法による結果の比較を第6表に示す．

この結果と前例第4表の結果の大きな違いは，2P法はQA法に比較してでoverestimateとなった点である．

ということになり，2P法の基本である2重指数分布モデルの応用では少し不安定な結論となっている．

(3) カルマン・フィルター理論の導入

まず離散型の線型フィルター理論の基本である状態方程式と観測方程式を，順序標本に合う形で提示しておこう．

(状態方程式)

$$T(x_{i+1})=A(x_i)T(x_i)+\eta(x_i)\quad(i=1,2,\cdots,n)\quad(7.9)$$

(観測方程式)

$$y(x_i)=c(x_i)T(x_i)+\varepsilon(x_i)\quad(i=1,2,\cdots,n)\quad(7.10)$$

ここに

$x_1,x_2,\cdots x_n$: 所与の順序標本，$T(x_i),T(x_{i+1})$: x_i, x_{i+1} に対する真の再現期間（未知）．$A(x_i)$: 予測のための状態変換係数で x_i と共に変化する．(一般的なカルマン・フィルター理論では所与であるが，この場合は未知)．$\eta(x_i)$: 状態システムの誤差．$y(x_i)$: 観測された所与順序標本による再現期間の値．$c(x_i)$:観測システムの変換係数．$\varepsilon(x_i)$: 観測方程式の誤差．

である．

再現期間の同定は，まず $x_i(i=,2,\cdots,n)$ から $y(x_i)$ $(i=1,2,\cdots,n)$ を Thomas plot により求める．つぎに線型推定量 $T(x_i)(i=1,2,\cdots,n)$ は，つぎの線型方程式 (7.11) により得られるとする．

$$\hat{T}(x_i)=\alpha y(x_i)+\beta\quad(i=1,2,\cdots,n)\qquad(7.11)$$

ここでα，βは連立方程式（通常の最小2乗法における正規方程式）

$$\left.\begin{aligned}&\sum_{i=1}^{n}\{(c\alpha-1)y(x_i)+\beta c+\varepsilon(x_i)\}cy(x_i)=0\\&\sum_{i=1}^{n}\{(c\alpha-1)y(x_i)+\beta c+\varepsilon(x_i)\}c=0\end{aligned}\right\}\quad(7.12)$$

の解としてもとめられる．この連立方程式の中に出てくる c は，カルマン・フィルター理論の中では本来 x_i の関数で観測システムの変換係数と呼ばれているものであるが，ここでは関数型を決めにくいし，実用上，簡単のために定数と置いている．

結局，観測誤差について $E\{\varepsilon(x_i)\}=0$ を仮定すると，再現期間 $T(x_i)$ の同定は，次の式で与えられる．

$$\left.\begin{aligned}&\hat{T}(x_i)=\bar{T}+\frac{c\sigma_\varepsilon^2}{\sigma_T^2+c^2\sigma_\varepsilon^2}(y(x_i)-\bar{T})\\&\sigma^2(\hat{T})=(\sigma_{\bar{T}}^{-2}+c\sigma_\varepsilon^{-2})^{-1}\end{aligned}\right\}\quad(7.13)$$

(7.13) により，$\hat{T}(x_1)$, $\hat{T}(x_2),\cdots,\hat{T}(x_n)$ が得られたら，状態変換係数 $A(x_1),A(x_2),\cdots,A(x_n)$ を求めなければならない．しかし，ここに1つの問題がある．それは，カルマン・フィルター理論では，$A(x_i)$ は関数形自身が既知の関数であるが，極値再現期間を取り扱うこのケースでは，観測された順序統計量によって定まるため**既知ではない**ことである．そこで，上記方程式から $A(x_i)$ の推定を必要とする．ここでの一案は，(7.13) による線型推定 $\hat{T}(x_1),\hat{T}(x_2),\cdots,\hat{T}(x_n)$ から

$$A(x_1)=\frac{\hat{T}(x_2)}{\hat{T}(x_1)},\quad A(x_2)=\frac{\hat{T}(x_3)}{\hat{T}(x_2)},\cdots,$$

$$A(x_{n-1})=\frac{\hat{T}(x_n)}{\hat{T}(x_{n-1})}$$

と置き，$A(x_n)$ を外挿によって求める方法である．このため次の線型変換推移モデルを仮定する．

$$A(x_i)=aA(x_{i-1})+b(x_i-x_{i-1})+\varepsilon(x_i)\quad(i=1,2,\cdots,n)\qquad(7.14)$$

この係数 a，b は容易に次の正規方程式により求まる．

$$\left.\begin{aligned}&\{\sum_{i=2}^{n}A(x_{i-1})^2\}a+\{\sum_{i=2}^{n}A(x_{i-1})(x_i-x_{i-1})\}b\\&\quad=\sum_{i=2}^{n}A(x_i)A(x_{i-1})\\&\{\sum_{i=2}^{n}A(x_{i-1})(x_i-x_{i-1})\}a\\&\quad+\{\sum_{i=2}^{n}(x_i-x_{i-1})^2\}b=\sum_{i=2}^{n}A(x_i)(x_1-x_{i-1})\end{aligned}\right\}\quad(7.15)$$

この数値例は筆者らにより，第1回気候統計国際会議に報告されているが，問題（a）への解答として妥当な数値的結果を与えている(ここでは省略)．

(4) 曲線のあてはめと外挿手法

横軸に順序標本かまたはそれらの換算変量，縦軸に非超過確率または再現期間を併記入した，いわゆる確率紙(必ずしも極値確率紙目盛でなくてもよい)上にデータを記入し，これになるべくよく適合し，かつパラメータ数の少ない曲線をあてはめて，それを右上方向に延長し，外挿する全く実験的（経験的）な方法である．

筆者の試みはすべて依託研究であり直接引用できないので，曲線あてはめに限定したモデル例だけをあげることにとどめた．

一般に気象的物理量Xについては，物理的考察から，これ以上大きな値は存在し得ないはずの可能上限界Uと，これ以下に小さくなり得ないはずの可能下限界Lとがあるとするのが自然であろう．

このとき，$L\leqq X\leqq U$ なる X を $-\infty<\zeta<+\infty$ なる

確率変数モデル ζ に変換する1つの方法はたとえば，

$$\left.\begin{aligned} X&=(U-A)\zeta^2/\sqrt{(\zeta^2+a^2)}+A \quad \zeta\geqq 0 \\ X&=(A-L)\zeta^2/\sqrt{(\zeta^2+b^2)}+A \quad \zeta\leqq 0 \end{aligned}\right\} \qquad (7.16)$$

であろう（この他にも種々ある）．ここで A は順序標本中央値であるが，再現期間値として中央値 A 以上を考察する場合が多いから，実用上は (7.16) の前者だけで十分である．要するに右上りの単調増加関数の1モデルであり，$\zeta\geqq 0$ として，ζ の確率分布（たとえば半正規分布）を利用し，その超過確率から X の再現期間予測をする手順が作成される．ここでは確率紙上にプロットされた必然的に単調増加傾向をもつ点群に適合しそうな曲線式モデルを提示するのが目的なので，この他に横軸を順序標本 X_i，縦軸を再現期間値 $T(X_i)$ として

$$T(X_i)=\alpha X_i^{-b},\quad T(X_i)=Ue^{-a/X_i^2} \qquad (7.17)$$

とか，成長曲線などをあてはめる試みをした方々もある（詳細は省略）．

(5) 複数個の要因が重合しておこる再現期間値の予測

これまでの予測技法はいずれも順序標本だけによるものであった．しかし，気象学的には，2つ以上の要因が同時的誘因として作用し，結果として稀現象を発生させるメカニズムを考慮することが現実に説得的であることが多い．

たとえば，台風の接近か否かを示す定性要因 X_1 とある限度をこえる収束水蒸気量を示す定性要因 X_2 とによって，稀現象である豪雨 Y が発生する場合である．

このとき，要因が Y の発生に有効に働くとき $X_1=1$，$X_2=1$ とし，$P(X_1=1)=p_1$，$P(X_2=1)=p_2$ とすると，Y の発生確率 $P(Y)$ を

$$\begin{aligned} P(Y)&\equiv P(X_1=1,\ X_2=1) \\ &=p_1p_2+\rho\sqrt{p_1p_2(1-p_1)(1-p_2)} \end{aligned} \qquad (7.18)$$

とすることができる．ただし，ρ は X_1，X_2 の相関係数である．そこで，この右辺を標本からもとめるとき，p_1，p_2，ρ の最尤推定量 $\hat{p}_1$，$\hat{p}_2$，r（標本からの ρ の最尤推定はいわゆる関連係数）でおきかえ，それらの各推定誤差から最終的に $P(X_1=1, X_2=1)$ の推定量とその総合推定誤差 $VarP(Y)$ をもとめようとする方法がある (Suzuki, E. *et al.*, 1980)．

また，3個の要因 X_1，X_2，X_3 が同時に誘因となる場合の同様な $P(X_1=1,\ X_2=1,\ X_3=1)$ も3重相関論として定式化されている．しかし総合推定誤差のもとめ方は全く未開発である．（多変量2項分布が応用可能）

8. あとがき（残された諸問題）

稀現象発生確率分布モデルの研究は，R. Katz (1963) による2パラメータ確率母関数 pgf (probability generating function) モデル提案を発端として3パラメータは Bhalerao N.R. and J. Gurland (1980)，4パラメータは M. Ahmad (1980)，5パラメータは筆者 (1984) により一般化の方向に進められているが，いずれも予測技法化されてはいない．詳細は筆者 (1984) の109—111頁に述べられているが，稀現象や異常極値のこうしたモデルによる統計処理手法は多種多様である．

そして国際会議で毎回議論されるこの問題のセッションでも，理論家は極値極限分布を出発点とした理論モデルの開発ならびに上記のように一般化された多くの数式作成に強い関心をもつのに対し，実際家は従来からの3種の極限分布タイプのどれを選択し適合させるかの実例的考察に追われているために甲論乙駁している．

ともかく，この問題は「有限標本しか得られないのに何故 n が十分大きいときの極値極限分布（たとえば2重指数極値分布型）をベースにした考え方しかとらないのか？」という基本的論点を残しつつ，多面的に技法の工夫がなされているというのが現状である．

I.I. Gringorten も筆者もこの点では実際家と理論家との間に可成りのギャップがあると痛感している．

この解説では当初の研究から現状までを実際家向けに要約したが，詳細なワイブル分布やその確率紙利用などは石原健二氏の研究解説の併読をすすめる意図もあって割愛し，多くの定理の証明，公式の誘導プロセス，Gringorten の包絡線作りにある詳細な数式的または図的工夫プロセス，大学院生の数理統計学的演習とすべき諸事項をすべて省略した．

なお，紙面の都合もあって参考文献も最小限としたが，これから実際利用と研究をすすめる方々にとってはこれで一応十分と考えている．

最後に拙文に有益なコメントを与えられた菊地原英和氏に心から感謝し，コメントに従い，書き加えさせて頂いた．

文 献

Ahmad, M., 1980: On the determination of some probabilistic models in forecasting rare events, WMO Symposium on Probability and Statistical Methods in Weather Forecasting, 337-341.

有本 卓，1977：カルマンフィルター，産業図書．

Chin, E.H. and J.F. Miller, 1977: On the estima-

tion of daily precipitation extremes, 5 th Conf. on Prob. and Stat. in Atm. Sci., 217-220.

Chow, V.T., 1964: Handbook of applied hydrology, 8-13-8-42, McGraw-Hill Book Company.

Fisher, R.A. and L.H.C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc. Cambridge Phil. Soc., **24**, 180-190.

Gringorten, I.I., 1962: A simplified method of estimating extreme values from data samples, J. Appl. Met., **2**, 82-89.

————., 1963: Envelopes for ordered observations applied to meteorological extremes, Journ. Geophys. Res., **68**, 3, 815-826.

Gumbel, E.J., 1941: The return period of flood flows, Ann. Math. Stat., **12**, 163-190.

————., 1958: Statistics of extremes, Columbia Univ. Press, New York, (この訳書は河田，岩井，加瀬監訳で1962年に極値統計学，広川書店として出版された.)

石原健二，1981：気象極値の再現期間について，気象研究ノート，No. 143，125-142.

Jenkinson, A.F., 1969: Statistics of extremes, WMO Techn. Note, **98**, 183-244.

————., 1975: Extreme value analysis in Meteorology, Fourth Conf. on Prob. and Stat. in Atm. Sci., 83-89.

片山　徹，1983：応用カルマンフィルタ，朝倉書店. 3-8.

Kendall, M.G. and A. Stuart, 1961: The advanced theory of statistics, Vol. II, 3 rd ed., Griffin London, 532-550.

菊地原英和，1959：確率雨量について，気象研究ノート，No. 10，125-139.

Kimball, B.F., 1960: On the choice of plotting positions on probability paper, J.A.S.A., **55**, 545-560.

Megreditchion, G.D. 1980: The statistical forecast of extrmes or rare phenomena, WMO Symposium on Prob. and Stat. in Weath. Forecasting, 331-336.

Oliveira, J.T., 1983: Extreme values and meteorology, II Intern. Meet. on Stat. Clim. Lisbon. Portugal, **10**, 1. 1-8.

Suzuki, E., 1961: A new procedure of statistical inference on extreme values, Papers in Met. and G., **12**, 1-17.

鈴木栄一，1968：気象統計学，地人書館，238-248.

————，1979：環境統計学，地人書館，19-29.

Suzuki, E., M. Miyata and S. Hongo (1980): Statistical prediction of climatological extreme values and return period in the case of small sample, Statistical Climatology, Elsevier Sci. Publ. Comp., Amsterdam 1980, 207-216.

鈴木栄一，1980～1983：極値予測論と予測手法の諸問題（Ⅰ），（Ⅱ），まとめ，青山経済論集，**31**，97-110，**32**，108-125，**34**，1-29.

————，1984：離散型分布の一般化と最近の確率母関数理論，青山経済論集，**36**，1，97-114.

Takle, E.S. and J.M. Brown, 1978: Notes on use of Weibull statistics to characterize wind-speed data, J. Appl. Met., **17**, 556-559.

Tomas, H.A., 1948: Frequency of minor floods, J. Boston Soc. Civil. E., **108**, 1110-1160.

Thom, H.C.S., 1960: Distribution of extreme winds in the United States, J. Struct. Div. (Proc. ASCE.), 11-24.

WMO, 1972: Estimation of maximum floods, Report of a Working Group on the Commission for Hydrometeorology, Technical Notes, No. 98, 183-227.