1981年7月　天　気　Vol. 28, No. 7

ヒストグラムの誤差と描き方

第1報　標本誤差*

菊地原　英和**

目　次

1. 従来の研究及び第1報の概要

この問題に関する知識の現状は，ほぼ次の3つに要約される．

（1）入門的統計学書等の観測値整理又は度数分布の項に述べられている，描き方についての説明又は注意．

（2）標本の大きさ N から，階級別度数（等間隔）の最適階級数 n を求める関係式 $n=f(N)$ についての2，3の研究．

（3）正規母集団の標本分布を，N=10～200について例示した資料（気象庁観測技術資料 No. 17）．

このうち（1）の内容は，およそ次のような事項である．

1）階級の幅および境界位置は，きりのよい数値を採用する．

2）階級数は10から20くらいが良い．階級が少なすぎると分布がわからなくなり，逆に多すぎると余分な手数がかかる．

3）標本の大きさに応じて適当な階級数を選ぶ．標本が小さすぎたり，階級を細かく分けすぎると，度数分布が不規則になる（$n=f(N)$，又は N と n の対応表を引用）．

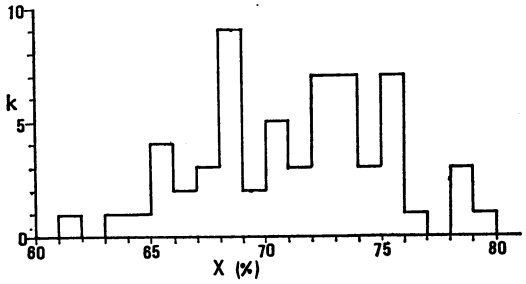
次に，（2）の階級数の式としては，次の2つの式がある．

* On the errors contained in a histogram and reasonable expression for histograms. Part 1 Sampling error.

** Hidekazu Kikuchihara, 気象大学校.

第1表　従来の式による最適階級数 n.

式 \ N	100	500	1,000	10,000
(1)	10.0	13.5	15.0	20.0
(2)	7.6	10.0	11.0	14.3
(3), 正規分布	6.8	13.9	18.7	47.7



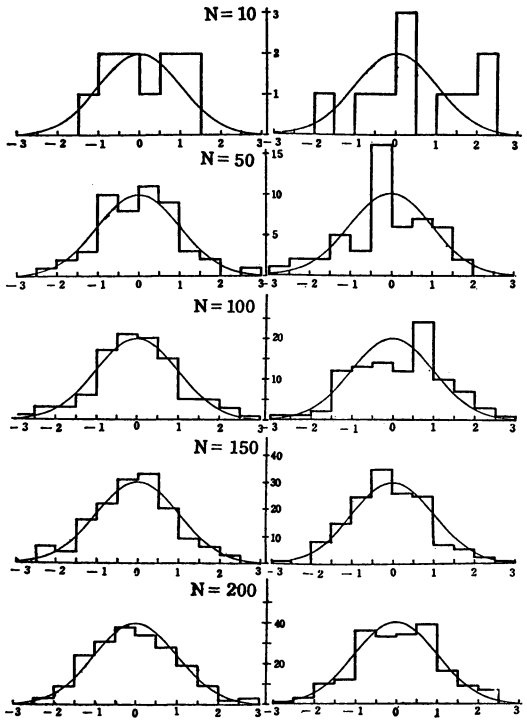
第1図　不規則なヒストグラムの例.
月平均相対湿度，東京，4月，1890～1949年，
$N=60$, $n=19$.

$$n=5\log_{10}N \qquad (1)$$

$$n=1+\log_{10}N/\log_{10}2\fallingdotseq 1+3.32\log_{10}N \qquad (2)$$

この（1）式は，Brooks and Carruthers (1953) の気象統計ハンドブックに，一応の目安であるとして述べられているもので，根拠ははっきりしない．（2）式は，Sturges (1926) によるものでよく引用されるが，その論拠は，2項係数 ${}_mC_r$, $r=0\sim m$ が度数分布としてきれいな形をしていることから，その r についての和 $N=2^m$ を大きさとする標本の階級区分数として，項数 $m+1$ を採用しようというのであって，例えば $m=4$ とすれば2項係数は1, 4, 6, 4, 1となりその和は $N=16$，項数は $n=5$ となる．このような N と n の関係が（2）式である．Sturges は，この式を平均値や分散，歪度などを計算するための階級幅を決めるのに提案したものであって，度数分布の表現法としての階級区分とは目的が異なる．現在は原著者の意図を離れて，一般的な階級数の式として引用されている傾向があるが，度数分布の階級数の式としては論拠に乏しく，その適否は（1）式と同様に不明確である．

森 (1974) は，2，3の数学的な条件のもとで，ヒストグラムの分布曲線に対する偏差の2乗の全領域についての積分の期待値を最小にするという条件で階級幅を求め，次の解を得た．



第2図　正規母集団の標本のヒストグラム，
10例中の最良（左側）と最悪（右側）.

$$\left.\begin{aligned} h^*(N)&=\left(\frac{6}{BN}\right)^{1/3} \\ B&=\int_{-\infty}^{\infty}f'(x)^2dx \end{aligned}\right\} \qquad (3)$$

ここで，$h^*(N)$ は階級幅，N は標本の大きさ，$f(x)$ は確率密度である．さらに．これを正規分布に適用し，標本範囲の式と組み合わせて数値計算を行ない，$N=100\sim1,000,000$ について正規分布の最適階級数を与えた．

標本の大きさ N から最適とされる階級数 n を求める上記3通りの関係式は，この報告で筆者が提案する式と共に後記第20図に描いてあるが，その一部の値を比較のために第1表に示す．このように（2）式より（1）式が大きく，（3）式は N が大きくなると，他より著しく大きくなる．

森の研究成果は論拠が明確であって，貴重なものであるが，実際にデータを扱う上では，これだけでは問題が解決しきれたとは言えない．第1図は気象データのヒストグラムの1例であるが，度数分布が非常に不規則に変動している．この例は $N=60$, $n=19$ であって，従来のどの式から見ても階級数が著しく過大である．それ故，この不規則の最大の原因が，階級を多くとり過ぎた

ための階級別度数の信頼度の低下，つまり標本誤差の増大にあることは容易に推測できる．しかし筆者の経験では，適当とされている階級数を採用した場合でも部分的には，程度の差はあれこのような不規則な変動を含むヒストグラムが得られることがしばしばある．このようなとき，その部分的な分布の不規則（例えば深い谷）が，階級別度数の標本誤差（偶然変動）による見掛けのものか，母集団分布に実在するものかを判断する必要が起こる．（3）項に挙げた資料は，規準正規母集団から，乱数表を使って，大きさ N の標本（N=10, 20, ……200）を10組ずつ作り，階級幅 d=0.5 のヒストグラムを描いた図集で，各Nについて，標本ごとのヒストグラムの変動がどの程度かを直感的に理解するのに役立っている．第2図にはこの資料から，10例中最も母集団分布に近いと思われるものと，最も不規則と思われるものを選んで掲げた．この資料は階級別度数の標本誤差の概念的な理解には役立つが，上に述べたような，目前のヒストグラムの不規則が母集団に実在のものかどうかを判定するには，直接役立たない．いま，変数 x の区間 $[a, b]$ で区切られる階級の実測度数を k，標本の大きさを N とすれば，変量がその階級に属する標本比率は，$p=k/N$ であって，これから母集団比率 p_0，つまり確率密度 $f(x)$ の区間 $[a, b]$ における積分値の，指定した信頼係数での信頼区間 $[p_L, p_U]$ は，周知の公式によって容易に計算できる．この p_L, p_U の N 倍を k_L, k_U とすれば，区間 $[k_L, k_U]$ は，$[a, b]$ の階級の母集団度数の，同じ信頼係数の信頼区間である．ヒストグラムの問題の箇所についてこの信頼区間を求めれば，それを基準として，不規則が実在のものか否かを客観的に判定できる．

これは理論的に目新しいことは何1つ無いが，計算が多少煩わしいためか，あまり実行された例を聞かない．それ故この報告では，いちいち計算しなくてもNとkから容易に信頼区間 $[k_L, k_U]$ が求められるようなノモグラムを先ず呈示し，これを利用して，モデル的なヒストグラムについて，Nとnを変えたとき標本誤差がどのように変化するかを量的に検討し，その結果から，新しい階級数の式を提案した．ヒストグラムモデルには正規母集団の標本を採用したが，階級数の式は，一般の一山型分布に適用可能である．

なお，N と k から，信頼区間 (p_L, p_U) を読み取るノモグラムには既存のものがあり（文献（6））, これをN倍すれば (k_L, k_U) が求まるが，線が混んでいて正確な読み取りはむずかしい．この報告のノモグラムでは，k_L, k_U を直接表示せず，誤差 $(k-k_L$ 及び $k_U-k)$ を採用することによって読み取りやすい図を作った．気象データの実際のヒストグラムへの適用例は，第2報の中で述べる．

2. 階級別度数の信頼区間の計算法

標本比率 $p=k/N$ が与えられたとき，信頼係数 α の母比率 p_0 の信頼限界 p_L, p_U が次の式で与えられることは周知の通りである．

$$\left.\begin{aligned} p_L &= \frac{m_2}{m_1 F^{m_1}_{m_2}(\beta)+m_2}, \quad m_1=2(N-k+1), \\ & \quad m_2=2k \\ p_U &= \frac{m_1' F^{m_1'}_{m_2'}(\beta)}{m_1' F^{m_1'}_{m_2'}(\beta)+m_2'}, \quad m_1'=2(k+1), \\ & \quad m_2'=2(N-k) \end{aligned}\right\} \quad (4)$$

ただし

$$\beta=\frac{1}{2}(1-\alpha) \quad (5)$$

また，$F^{m_1}_{m_2}(\beta)$ は自由度 m_1, m_2 の F 分布の超過確率 β の点で，$F^{m_1'}_{m_2'}(\beta)$ も同様である．

（5）式から，α=0.5 のとき β=25%，α=0.9 のとき β=5% となり，既存の数表が使える．

母集団度数の信頼限界は，（4）式と，

$$k_L=Np_L, \quad k_U=Np_U \quad (6)$$

から計算される．

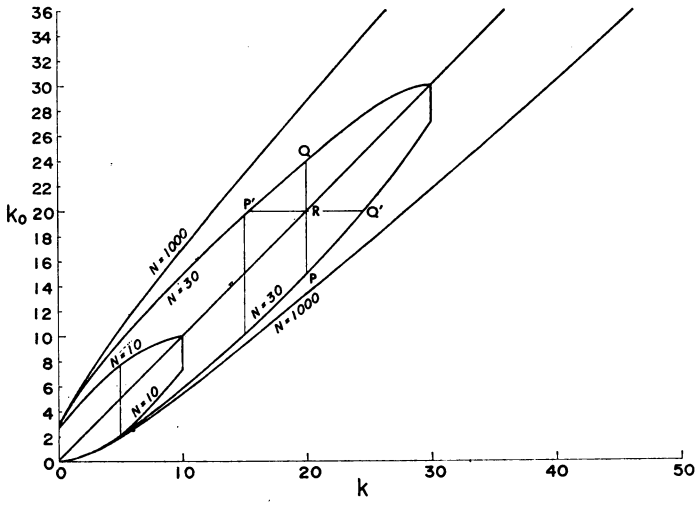
連続変量 x の確率密度 $f(x)$ を持つ母集団から抽出した大きさNの無作為標本の度数分布について，任意の階級の変域を I, その標本度数を k，母集団度数を k_0 とするとき，k_0 の信頼係数 α の信頼区間が（6）で与えられる．つまり，次の式が成り立つ．

$$\left.\begin{aligned} & p_0=\int_I f(x)dx, \quad k_0=Np_0, \\ & \mathrm{Prob}(k_L \leqq k_0 \leqq k_U)=\alpha \end{aligned}\right\} \quad (7)$$

3. 階級別度数の信頼区間の性質

前節の方法で，α=0.9 および 0.5 として，N=10～1000，k=0～N の種々の値について階級別度数の信頼区間を計算した．そのうち，α=0.9；N=10, 30, 1000 の計算結果を描いたのが第3図で，横軸に標本度数 k，縦軸に母集団度数 k_0 をとり，$k_0=k_L$ および $k_0=k_U$ の曲線と，$k_0=k$ を表わす勾配 45° の直線が描いてある．

k が与えられたとき，k_0 の信頼区間は，横座標が k



第3図　階級別度数の90%信頼区間
(k：標本度数　k_0：母集団度数).

のところに引いた縦線が両曲線で切り取られる線分で表わされる．例えば $N=30$, $k=20$ のときは図の線分 PQ の縦座標 [15.0, 24.2] が信頼係数90%の信頼区間である．この信頼区間の幅は，標本誤差の大きさの尺度と見ることができるので，以下これを信頼係数90%の「誤差範囲」と呼ぶ．同様に図の線分 PR および RQ の大きさ，つまり標本度数 k と信頼限界 k_L, k_U との差の絶対値を，「下方誤差」および「上方誤差」と呼び，次の記号で表わす．

$$\left.\begin{array}{ll} \text{誤差範囲} & D(N,\ k,\ \alpha)=k_U-k_L \\ \text{下方誤差} & D_L(N,\ k,\ \alpha)=k-k_L \\ \text{上方誤差} & D_U(N,\ k,\ \alpha)=k_U-k \end{array}\right\} \qquad (8)$$

ただし誤解のおそれがないときは括弧内の引数の一部又は全部を省略して，$D(N,\ k)$, D_L 等で表わす（他の記号についても同様）．また，度数である k_L, k_U など及び階級数 n を計算によって求めるときは，整数に限定せず実数として扱う．

3種の誤差 D, D_L, D_U の主な性質を次に挙げる．

（1）$D=D_L+D_U$

（2）$D_L(N,\ k)=D_U(N,\ N-k)$

（3）$D_L(N,\ 0)=D_U(N,\ N)=0$

（4）一定の N について，D は $k=N/2$ で最大となり，このとき $D_L=D_U=D/2$

（5）$k\lesseqgtr N/2$ のとき $D_U\gtreqless D_L$

以上は（4），（6）式および（8）式から容易に導かれる．性質（2）は図形が中心対称のことを意味し，また，階級別度数で $k>N/2$ ということはないから，（5）は実質上常に上方誤差が下方誤差より大きいことを意味する．

4. N, k の関数としての誤差 D_L, D_U

N と k を与えて D_L, D_U を計算した結果を見ると，一般に N による変化は緩やかである．それ故，計算結果は，横軸に N を対数目盛でとり，縦軸に誤差をとって，$D_L(N,\ k,\ 0.9)$, $D_U(N,\ k,\ 0.9)$, $D_L(N,\ k,\ 0.5)$ および $D_U(N,\ k,\ 0.5)$ を各1枚に描いたのが第4，5，6図および第7図である．これらの図には，$k=0$ から $k=500$ まで，15本の曲線が描いてあるから，その内挿によって，500以下の任意の k および $N=50$～$10,000$ についての上方誤差，下方誤差が容易に読み取れるであろう．

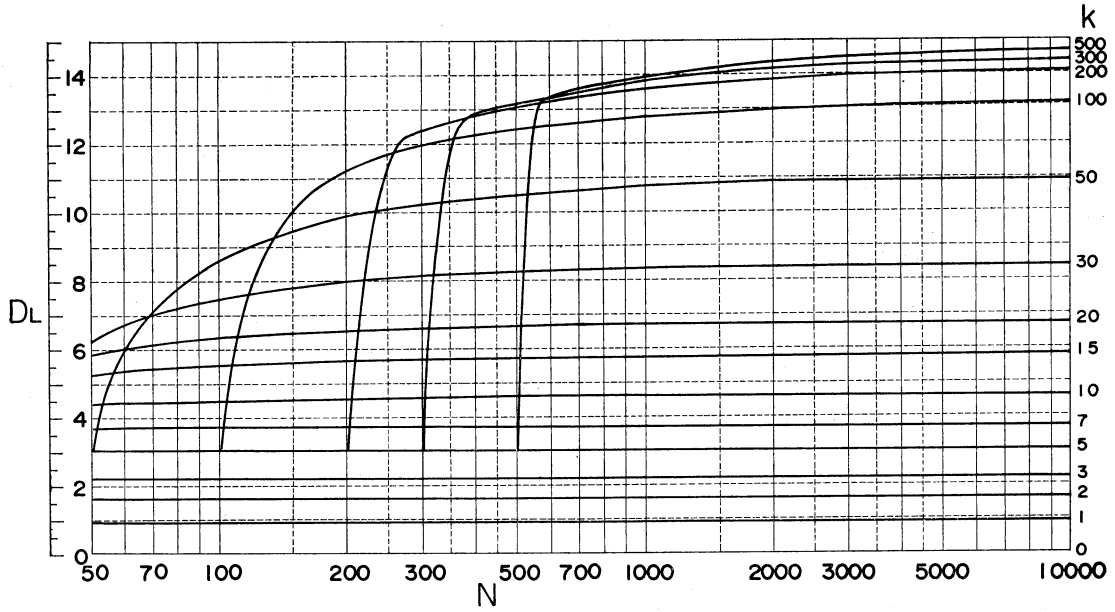
前節では誤差 D, D_L, D_U について（1）から（5）までの性質を挙げたが，第4～7図の考察で得られる傾向，特徴について追加すれば次の通りである．

（6）D_L, D_U は N と共に単調に増加する．

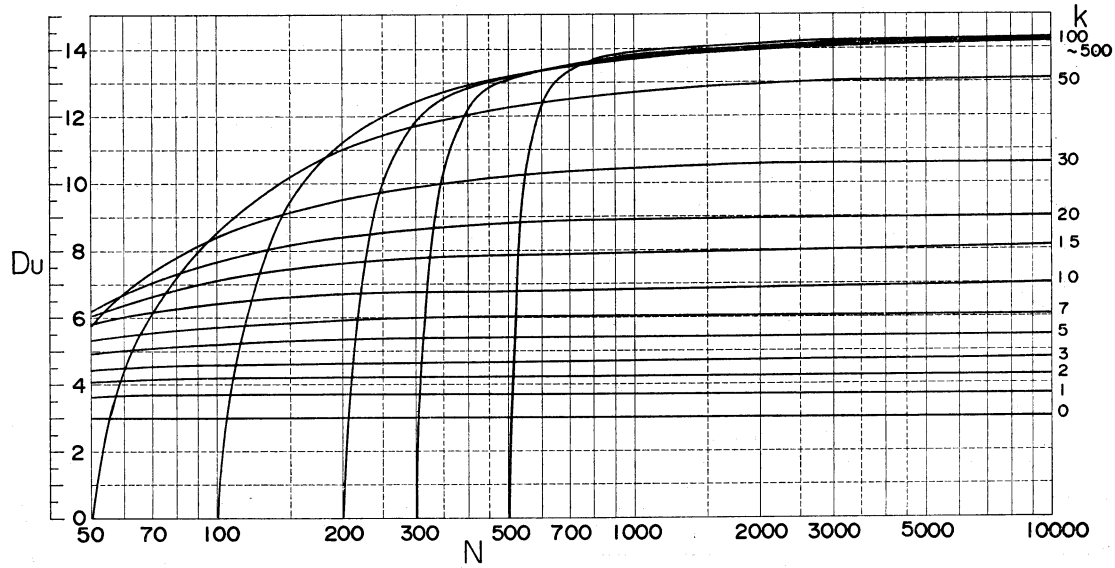
（7）k が小さいとき，D_L, D_U の N による変化は極めて緩慢で，近似的には k だけで決まる．

（8）k が N に近いときは，D_L, D_U の N による変化が著しい．

（9）D_L, D_U は k と共に増加するが，D_L は $k=2N/3$ 付近，D_U は $k=N/3$ 付近で極大に達し，そ



第4図　$\alpha=90\%$の下方誤差　$D_L(N,\ k,\ 0.9)$.



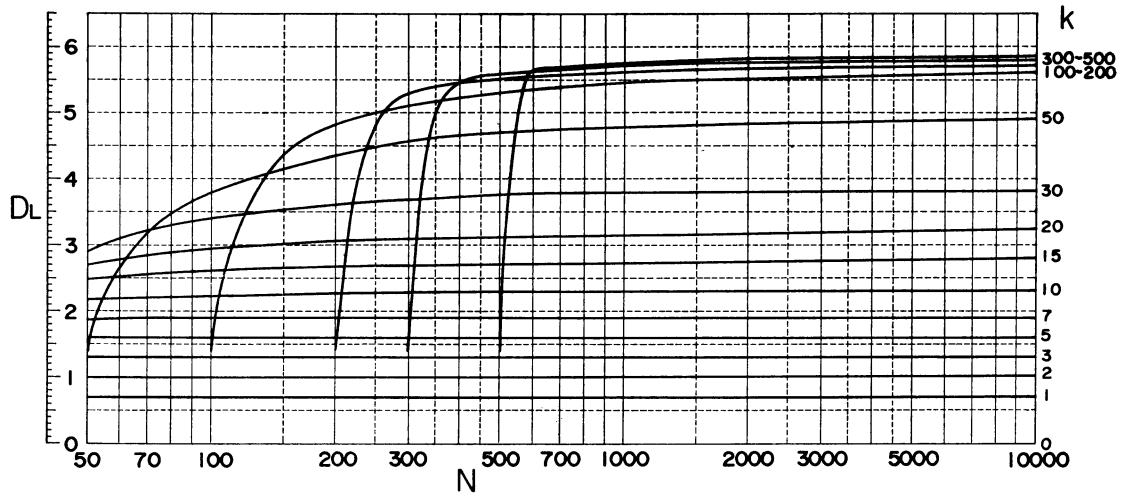
第5図　$\alpha=90\%$の上方誤差　$D_U(N,\ k,\ 0.9)$.

れ以後減少に転ずる．この増加領域で，D_L，D_U の増加率は k と共に小さくなる．

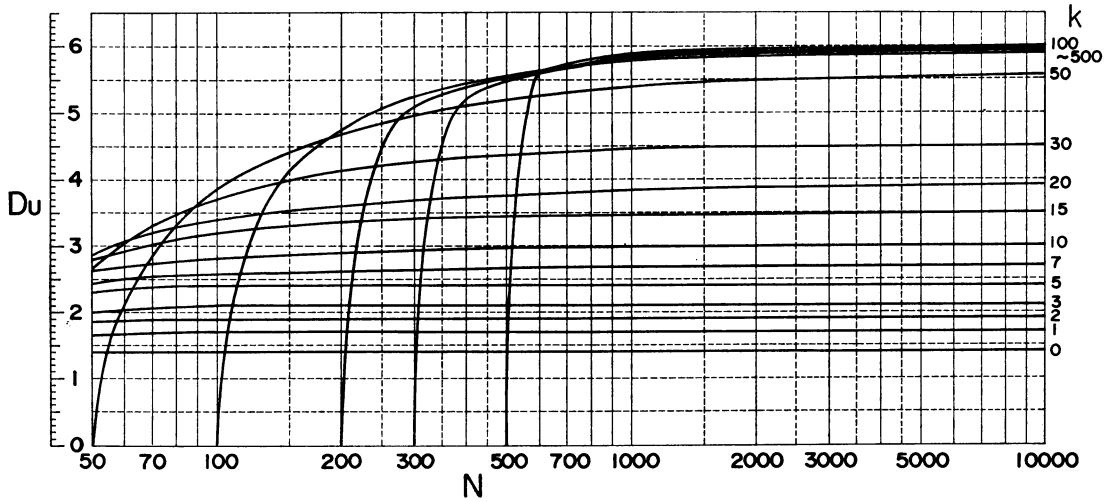
(10) 同じ N，k について，$\alpha=90\%$ の誤差 D_L，D_U は，$\alpha=50\%$ の D_L，D_U に比べて約2.5倍程度に大きい．

以上，特徴的な性質はアンダーラインで示したが，特に（7）は顕著な性質で，D_L，D_U は k が10以下では N が100から10000まで殆んど一定であり，k が30以下でも1ないし2増す程度である．

階級別度数の信頼度の尺度としては，D_L，D_U，D そ



第6図 $\alpha=50\%$の下方誤差 $D_L(N,\ k,\ 0.5)$.



第7図 $\alpha=50\%$の上方誤差 $D_U(N,\ k,\ 0.5)$.

第2表 誤差率 $Q=Q_L+Q_U$.

k \ N	$\alpha=0.9$			$\alpha=0.5$		
	30	100	1000	30	100	1000
1	4.4	4.6	4.7	2.3	2.4	2.4
2	2.7	2.9	2.9	1.4	1.5	1.5
3	2.1	2.3	2.3	1.1	1.1	1.1
5	1.5	1.6	1.7	0.7	0.8	0.8
10	0.9	1.1	1.2	0.4	0.5	0.5
20	0.5	0.7	0.8	0.2	0.3	0.4
50	—	0.3	0.5	—	0.1	0.2
100	—	0.03	0.1	—	0.01	0.1

のものよりも，次式の相対誤差（誤差率）のほうが適切な場合が多い．

$$\left.\begin{aligned} Q_L(N,\ k,\ \alpha)&=D_L(N,\ k,\ \alpha)/k \\ Q_U(N,\ k,\ \alpha)&=D_U(N,\ k,\ \alpha)/k \end{aligned}\right\} \qquad (9)$$

第2表には $Q=Q_L+Q_U$ の値を示す．Q_L, Q_U の傾向は Q と同様で，数値は表のほぼ2分の1程度である．この表から得られる性質を追加すれば

(11) 相対誤差は k の増加とともに急激に減少する．Nによる変化は緩慢である．

(12) 度数 k が10以下の階級では，$\alpha=90\%$ の相対誤差 Q が100%をこえる．$\alpha=50\%$の Q は50%をこえる．

結局のところ，大多数の階級で度数 k が10以上ないと，ヒストグラムはきわめて信頼度が低いと言えよう．

第4～7図は母集団の分布型によらないのはもちろん，k を任意の事象の出現度数としても利用できる．例えば，k_i を x の小さい方からの累積度数とし，標本分布関数を $F(x_i)=k_i/N$ とすれば，その信頼係数 α の信頼区間は次の式で容易に求められる．

$$\left.\begin{aligned} F_L(x_i,\ \alpha)&=\{k_i-D_L(N,\ k_i,\ \alpha)\}/N \\ F_U(x_i,\ \alpha)&=\{k_i+D_U(N,\ k_i,\ \alpha)\}/N \end{aligned}\right\} \quad (10)$$

5. 階級別度数の信頼区間と標本度数の変動幅の関係

階級別度数の信頼区間を表わす第3図で，45° 線から曲線までの鉛直距離，つまり PR，RQ が誤差 D_L，D_U であり，線分 PQ が標本度数 k が与えられたときの母集団度数 k_0 の信頼係数 α の信頼区間である．一方，水平の線分 $P'Q'$ は，母集団度数 k_0 が与えられたとき，標本度数 k の分布の中央部にとった，確率 α の範囲を表わす．つまり，この範囲を $[k_L',\ k_U']$，45° 線から曲線までの水平距離（線分 RP'，RQ'）を D_L'，D_U' とすれば，次の関係が成立する．

$$D_L'=k_0-k_U',\quad D_U'=k_U'-k_0 \quad (11)$$

$$\left.\begin{aligned} &\mathrm{Prob}(k\leqq k_L')=\mathrm{Prob}(k\geqq k_U')=1-\alpha/2 \\ &\therefore \mathrm{Prob}(k_L'\leqq k\leqq k_U')=\alpha \end{aligned}\right\} \quad (12)$$

この図形の縦の幅 D_L，D_U と横の幅 D_L'，D_U' とは当然一致しないが，数値的に比較すると，

$$\left.\begin{aligned} &N=50\sim1000,\ 5\leqq k_0\leqq N-5 \ \text{のとき} \\ &D_L'\fallingdotseq D_L,\ D_U'\fallingdotseq D_U \end{aligned}\right\} \quad (13)$$

が成り立っていることがわかる．それ故，母集団度数 k_0 が与えられたとき，標本度数 k の確率 α(=0.9又は0.5) の変動範囲 $[k_L',\ k_U']$ は，第4～7図で $k=k_0$ として求めた D_L，D_U を使って，次の式で与えられる．

$$k_L'\fallingdotseq k_0-D_L,\quad k_U'\fallingdotseq k_0+D_U \quad (14)$$

誤差の大きさは (13) 式の N，k_0 について最大2以下で，Nが1000以上でも (14) 式は十分実用できると見られる．

6. モデルヒストグラムによる標本誤差の評価と階級数の検討

標本誤差を評価するためのノモグラムの準備ができたので，この節以降では，従来の（1），（2），（3）式で階級数を決めたヒストグラム（又は度数折線．以下同様とする．）が，どの程度の標本誤差を持つかを調べ，この点から，各式の適否を考察する．しかし，標本誤差は直接的には N と k の関数であり，度数 k は N と n だけでは決まらず，母集団の分布型に依存し，さらに，標本ごとに異なる．これらをすべて検討することは不可能であるから，分布型や標本度数等について，最も一般的と考えられるようなモデルヒストグラムを想定し，これについて検討する．以下にこのモデルの条件を挙げ，続いてその理由を述べる．

1）母集団は規準正規分布 $N(0,\ 1)$．

2）ヒストグラム全体の幅は，大きさ N の標本の最大値のモード $\tilde{x}_N$ から最小値のモード $\tilde{x}_1$ までの距離 $W=\tilde{x}_N-\tilde{x}_1$ とする．

3）階級区分は等間隔．

4）標本の大きさは $N=100$ および $N=1000$ とし，階級数は従来の式（1），（2），（3）による．

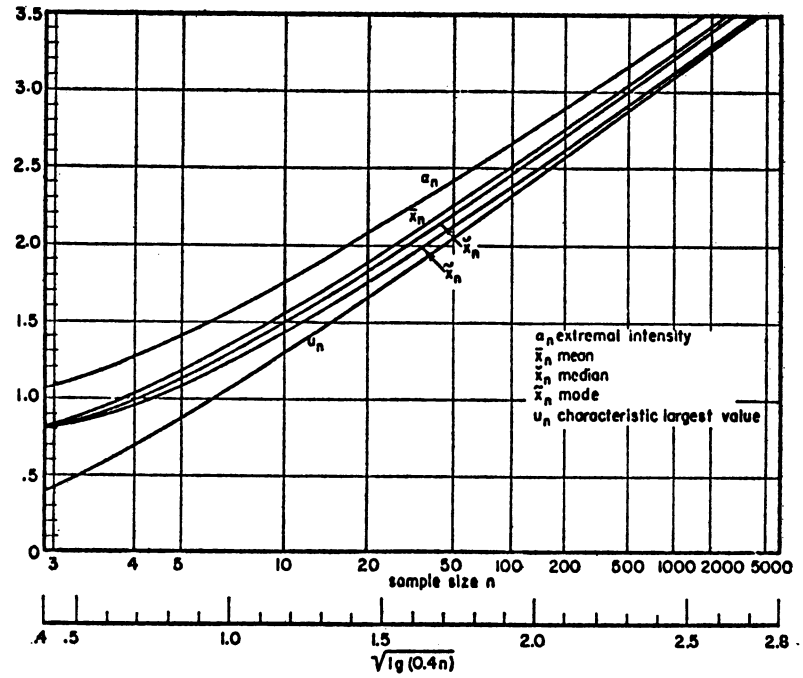
5）階級別度数 k は，期待度数 $k_0=Np_0$ に最も近い整数値．

従来の階級数の式のうち，（3）で具体的に階級数を与えているのは正規分布についてであり，（2）式は2項係数，つまり左右対称の一山型分布が根拠になっている．（1）式が適用できる分布型の範囲は明らかでないが，例えば雲量のような両側有界分布や，短時間の降水量（逆J字型分布）のように等間隔の階級区分を使うこと自体が適当でない特異な分布に適用できないのは明らかであり，恐らく著しい歪みは無い一山型分布を想定しているものと考えられるので，モデルでも同様な山型分布の典型として正規分布を採用した．$N(0,\ 1)$ でも $N(m,\ \sigma^2)$ でも，階級幅がちがうだけで，標本誤差の点では同等である．条件2）は次の節で説明する．条件3）は当然，4）は $n=f(N)$ の傾向が（1），（2），（3）式で異なるから，Nが小さい時と大きい時の2通りを調べるということである．

すべての階級で実測度数 k を期待度数 k_0 と一致させる条件5）は，ヒストグラム全体としては殆んど起こりえない特殊なものであるが，この調査で問題としているのは個々の階級の標本誤差であり，ひとつひとつの階級について見れば，k_0 は実測度数としても最も起こりやすい部類の値であるから，5）を採用した．

6.1. ヒストグラムの幅 W_N と N の関係

標本の大きさ N が大きくなるにつれて，標本の最小値 x_1 は小さくなり，最大値 x_N は大きくなって，必然的にヒストグラム全体の幅 W_N は拡がってゆく．x_1，x_N が標本ごとに異なるから，W_N も同様であるが，x_1，x_N の最も起こりやすい値として，それぞれのモードをこの



第8図　正規分布の各種極値統計量の標本の大きさによる変化（Gumbel による）.

モデルでは採用した．つまり次の式である．

$$W_N=\tilde{x}_N-\tilde{x}_1=2\tilde{x}_N \qquad (15)$$

なお，W_N は範囲のモード $\tilde{R}_N$ とは一致しないが，次の関係がある．

$$N\to\infty \text{ のとき } W_N\to\tilde{R}_N \qquad (16)$$

次に，W_N の具体的な求め方について概略説明する．正規変量の極値分布については従来から多くの研究があり，Gumbel の極値統計学（文献(2)）に要約紹介されているが，次の関係が得られている．

$$N-1=\tilde{x}_N F(\tilde{x}_N)/f(\tilde{x}_N), \qquad (17)$$

ここで，$f(x)$，$F(x)$ は規準正規分布の確率密度，分布関数である．Gumbel の前記著書には，K. Pearson による $F(x)/f(x)$ の数表を使ってこの式から求めた $\tilde{x}_N$ の数値が，Nが1000近くの値までについて掲載されているので，これを利用した．

さらに大きなNについては，正規分布が属する「指数型分布」に一般に成り立つ次の式

$$N\to\infty \text{ のとき } \tilde{x}_N\to u_N \qquad (18)$$

$$F(u_N)=1-1/N \qquad (19)$$

を近似的に利用した．u_N は特性最大値と呼ばれ，超過確率が $1/N$ に相当する変量値であるから，$F(x)$ の表（確率積分表）から容易に求められるが，$N=10^{10}$ までの u_N は極値統計学に表があるので，これを利用した．結局のところ，次の近似式を採用したわけである．

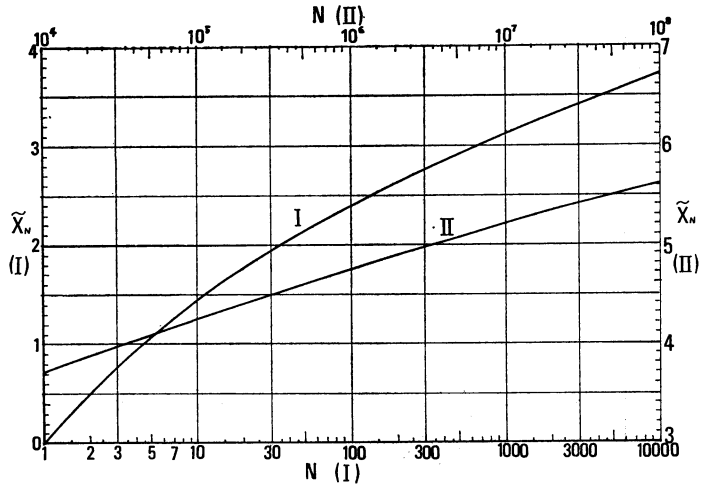
$$\tilde{x}_N\fallingdotseq u_N \ (N>1000) \qquad (20)$$

第8図は，正規極値統計量のNによる変化を示す．極値 x_N は歪度正の分布をしており，従って $\tilde{x}_N<\check{x}_N$（メジアン）$<\bar{x}$ であるが，その差は小さいので，かりに W_N に $\tilde{x}$ でなく $\bar{x}$ の2倍を採用したとしても，調査結果にそれほどのちがいは無いと考えられる．この図には特性最大値 u_N も描かれているが，近似式 (20) にも問題はないと見られる．

以上のようにして求めた $\tilde{x}_N$ と N の関係は第9図の通りで，Nの増加によるヒストグラムの幅$2\tilde{x}_N$ の増大は緩慢である．

6.2. 既存式の階級数によるモデルヒストグラムの標本誤差

標本の大きさ N から第9図でヒストグラム幅 W_N を求め，これを階級数 n で割って得た階級幅 d で正規変量を分割し，確率積分表から各階級の確率 p_0 を求め，これをN倍し，四捨五入して整数位にした階級別度数を k とし，ヒストグラムを描いた．標本誤差は，Nと k から第4～7図で求めた信頼区間を，図が見にくくならないために，度数折線（ヒストグラムの各柱の頂上中央の点



第9図　規準正規分布の最大値モード $\tilde{x}_N$ と N の関係.

を結んだ多角形）で表示した．なお，ヒストグラムの外側の確率が0でないことと四捨五入のために，k の合計はNと一致しない場合もあるが，その差は－2～＋1で無視できる．

（1）N=100，n=10 の場合（第10図）

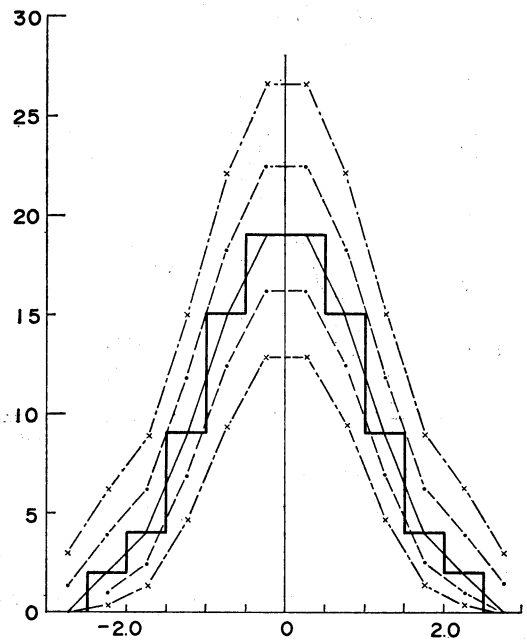
（1）式を採用した場合で，階級数 n=10は3つの式中最大である．外側の鎖線が α=90%，内側の破線が α=50%の，母集団度数の信頼区間を表わす．また，第5節の結果から，ヒストグラムを母集団度数の表現と見れば，破線，鎖線は，標本度数の確率50%，90%の変動範囲を近似的に表わしている．

第10図を見れば，このヒストグラムの標本誤差が著しく大きいことは明白である．誤差率が最も小さい中心階級でも，α=90% の信頼区間は13から27まで広がり，誤差率は70%を越える．このように大きな誤差の原因が度数 k の過小，つまり階級数の過大にあることも明らかで，誤差率 100% 以下の目安である「k が10以上」を満たしているのは中心部の4階級だけである．

（2）N=100，n=5 の場合（第11図）

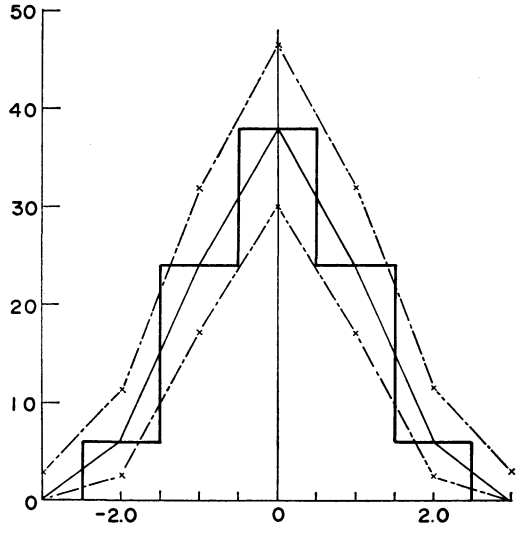
（2）式，（3）式で得られる階級数はそれぞれ n=7.6 及び 6.8 で，8又は7階級となるが，n=10 であまりにも標本誤差が大きいので，ここでは更に n を小さくし，n=5 として描いたのが第11図である．この図以降では α=50% は省略し，α=90% の誤差範囲（信頼区間）を鎖線で示した．

第11図では，中央階級で，k=38，90% 誤差範囲は約30～47，その誤差率は約43%と前よりかなり小さくなる

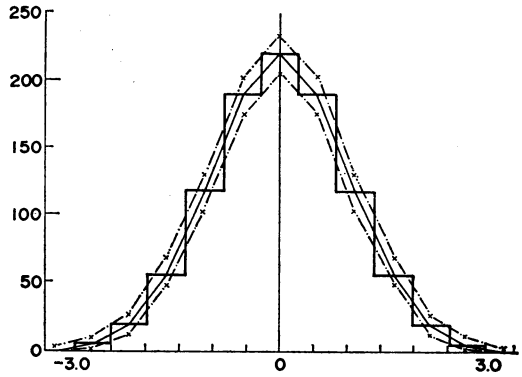


第10図　階級別度数の50%，90%信頼区間.
N=100，n=10，d=0.50
破線は α=50%，鎖線は α=90%

が，やはり誤差はかなり大きい．度数に関する望ましい目安条件，k が10以上の階級数は3で，かろうじて過半数に達している．階級数 n=5 は後に述べるヒストグラムの系統的誤差の点からは，かなり過少で，ましてこれ以上階級数を減らしては，ヒストグラムの用をなさな



第11図　階級別度数の90%信頼区間.
N=100, n=5, d=1.0

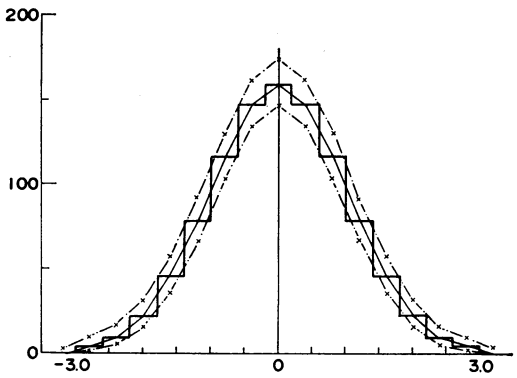


第12図　階級別度数の90%信頼区間,
N=1000, n=11, d=0.56

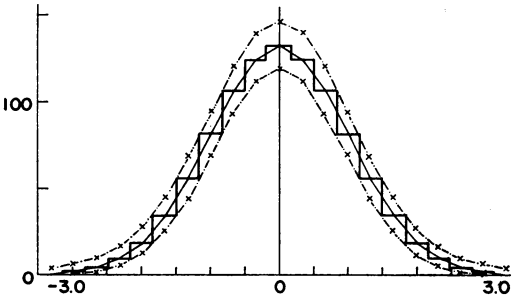
い．結局のところ，N=100では信頼が置ける ヒストグラムを描くことは，一般的にはかなり無理であって，それなりの大きな標本誤差を覚悟して描くほかはない．

（3）N=1000, n=11, 15, 19の場合（第12～14図）

第1表に挙げた（2），（1），（3）式による値を四捨五入して，n=11, 15, 19 を採用し，前図と同様に描いたのが第12, 13, 14図である．中心階級の度数が縦軸のスケールを変えるほどには違わないので3図共同じスケールであるが，第10, 11図と比べると度数が大きく違うので，縦軸のスケールは中心階級の度数にほぼ反比例して縮小してある．



第13図　階級別度数の90%信頼区間.
N=1000, n=15, d=0.40



第14図　階級別度数の90%信頼区間.
N=1000, n=19, d=0.33

第12～14図を見ると，標本誤差の幅は，N=100 の場合よりはるかに狭くなっている．90%誤差範囲の幅をヒストグラムの階段のステップと比べてみると．n=11 では誤差範囲のほうが狭く，n=15 で同程度，n=19 では斜面部で同程度，中心と両裾では誤差範囲のほうが広くなっている．ヒストグラムを母集団度数と見たとき，標本度数がこの誤差範囲の外に出る確率は10回に1回しかないから，n=11 及び n=15 では，ヒストグラムに不規則な凹凸を生ずる危険性は小さく，ヒストグラムの形が安定していると言える．n=19 でも斜面部で標本誤差による凹凸が起こる可能性は小さい．しかしこのようなヒストグラムの形の安定性は利用上重要なので，次節でさらに検討する．

中心階級の誤差率 $(D_L+D_U)/k$ の値は，n=11, 15, 19でそれぞれ約13%，17%，20%で，誤差率100% 以下の目安である「kが10以上」の階級の割合は，それぞれ 9/11, 11/15, 13/19であって，いずれも過半数に達している．

以上を要約して一応の結論をまとめれば次のようになる．

$N=1000$のとき，（1）式，（2）式の階級数は，標本誤差から見て過小である．

6.3. 度数逆転の可能性から見た検討

ヒストグラムの相隣る2つの階級を添字AとBで区別する．実測度数（標本度数）k_A，k_B の大小関係が，期待度数（母集団度数）k_{0A}，k_{0B}の大小関係と逆になる現象を，「度数逆転」と呼ぶことにすれば，その確率は次式で表わされる（k_0 が大きい方の階級をA）．

$$\mathrm{Prev}=\mathrm{Prob}(k_A<k_B|k_{0A}>k_{0B}) \tag{21}$$

母集団が一山型の単純な分布でも，ヒストグラムの全領域でこの条件付確率が大きければ，ヒストグラムに多数の凹凸ができる．また，母集団分布の斜面部に顕著な度数逆転が起こると，ヒストグラムは二山型のように見え，母集団分布の推測を誤まる危険が生ずる．このように度数逆転の確率の大小はヒストグラムの利用上重要なので，この節では前節のモデルヒストグラムを，この観点から検討する．

(21)式の確率についてあらかじめわかることを述べれば，k_{0A} と k_{0B} の差が小さい所，つまり一山型分布ならば中央部と両裾で起こりやすく，差が大きい所，つまり斜面部では起こりにくい．ヒストグラム全体で見れば，Nが小さかったり，nが過大であったりして標本誤差が大きいほど，度数逆転も起こりやすくなる．

当初は（21）式の確率を具体的に計算したいと考えたが，数値計算に相当の手数がかかることがわかり，計算は断念して，図的な便法で度数逆転の起こりやすさの程度を大まかに評価したが，後々のためもあるので，それを述べる前に，一応計算式を導いておく．

相隣る2つの階級を添字A，Bで区別し，次の記号を使用する．

$$\left.\begin{array}{ll} \text{母集団確率} & p_A>p_B,\ p=p_A+p_B,\\ & q=1-p\\ \text{母集団度数} & k_{0A}=Np_A,\ k_{0B}=Np_B\\ \text{標本度数} & k_A,\ k_B,\ k=k_A+k_B, \end{array}\right\} \tag{22}$$

この N, p_A, p_B を与えれば，（22）式の初めの2行の定数が決まる．2項分布を $Bin\,(N,\ p,\ x)={}_NC_xp^xq^{N-x}$ で表わせば（21）式は次の式で表わされることが容易にわかる．

$$\left.\begin{array}{l} \mathrm{Prev}=\sum_{k=1}^{N}\left\{Bin(N,\ p,\ k)\sum_{k_A=0}^{m}Bin\left(k,\ \frac{p_A}{p},\ k_A\right)\right\}\\ k\text{が奇数のとき}\quad m=\frac{1}{2}(k-1),\\ k\text{が偶数のとき}\quad m=\frac{k}{2}-1 \end{array}\right\} \tag{23}$$

さらに，自由度 ϕ_1，ϕ_2 のF分布の超過確率 β の点を $F_{\phi_2}^{\phi_1}(\beta)$ で表わし，（23）式右辺の2番目の2項分布の部分和をF分布で表わせば次のようにも書ける．

$$\left.\begin{array}{l} \mathrm{Prev}=\sum_{k=1}^{N}B_{in}(N,\ p,\ k)F_{\phi_2}^{\phi_1}(\beta)\\ k\text{が奇数のとき}\quad \phi_1=\phi_2=k+1,\\ \qquad \beta=\frac{p_A}{p_B}\\ k\text{が偶数のとき}\quad \phi_1=k,\ \phi_2=k+2,\\ \qquad \beta=\frac{k+2}{k}\cdot\frac{p_A}{p_B} \end{array}\right\} \tag{24}$$

あるいは，裾のごく少数の階級を除けば，Npq が十分大きいから，2項分布を正規分布で近似し，その部分和を正規確率から求める近似計算も可能である．どの方法にせよ，多数の隣接階級間について計算するのは，あまり簡単でない．計算の代りに行なった方法を次に述べる．

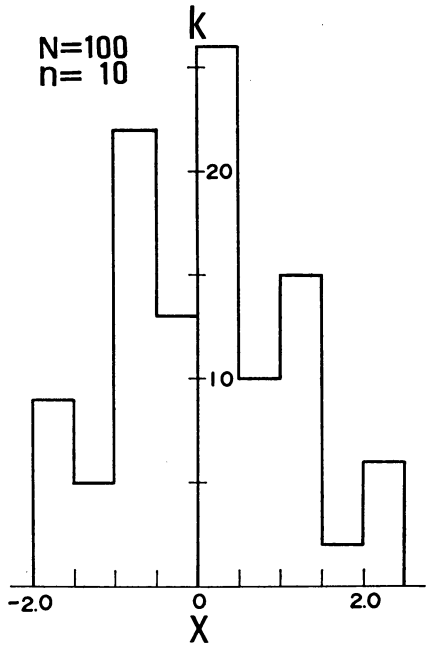
第10～14図のモデルヒストグラムを母集団度数を表わしていると見たとき，上下の鎖線は，標本度数の分布の中央部90％の変動範囲を近似的に表わしている（第5節）．それ故，この範囲内の度数は一応「起こりやすい」とし，この範囲外の度数は「起こりにくい」と見なして，「起こりやすい範囲で，できるだけ不規則なヒストグラム」を次のようにして描く．

1）階級数が偶数のときは，中央の2つの階級の一方を90％変動範囲内で最大の度数（整数）にとる．

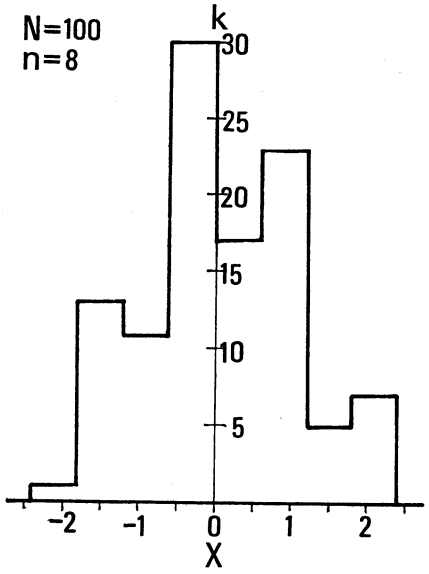
2）最大の度数の隣りの階級は90％変動範囲内で最小の度数（整数）にとり，その隣は再び最大の度数というように，最大の度数と最小の度数を交互にとる．

3）階級数が奇数のときは中央の階級が1つなので，その度数を最大とした場合を中央から一方の側に，中央の度数を最小にした場合を中央から他方の側に，2）と同じ方法で描く．

このようにして描いたモデル的な不規則なヒストグラムが第15～19図である．これらの図で度数逆転が起こっている所は，両方の階級の標本度数の90％変動範囲が一部重なった所であり，逆転がない所は，90％変動範囲に重なりが無い所である．この図で逆転の確率を量的に推定することはできないが，図に顕著な逆転が見られる所は実際にも逆転が起こりやすい所であり，図に逆転がない所では，実際に逆転が起こる確率は，大きくても5％以下である．以下各図について述べる．

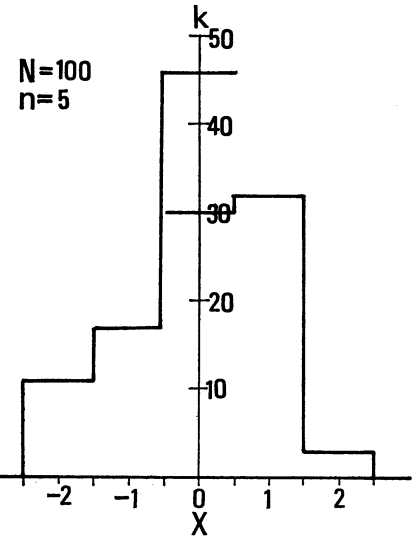


第15図 90%変動範囲内の不規則なヒストグラム.
$N=100$, $n=10$

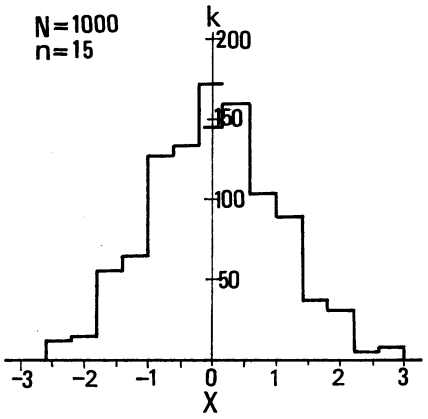


第16図 90%変動範囲内の不規則なヒストグラム.
$N=100$, $n=8$

第15図は $N=100$, $n=10$ の場合で，ヒストグラムの全領域で，極めて顕著な度数逆転が起こっている．もちろんヒストグラム全体としてはこのモデルのようなヒストグラムは殆んど起こり得ないが，局所的に見て，ヒストグラムのどの部分でも，顕著な度数逆転が容易に起こりうることを，この図は示している．



第17図 90%変動範囲内の不規則なヒストグラム.
$N=100$, $n=5$

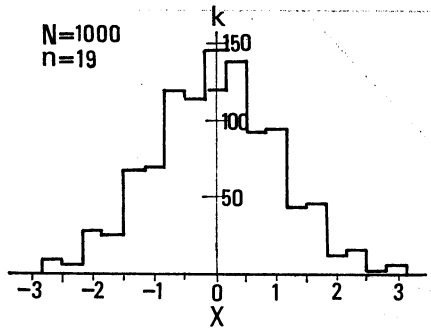


第18図 90%変動範囲内の不規則なヒストグラム.
$N=1000$, $n=15$

第16図は $N=100$, $n=8$ のときの図である．この場合も，至る所で度数逆転が起こっているが，その程度は $n=10$ のときより小さい．

第17図は $N=100$, $n=5$ の図で，中央部に弱い逆転が見られるほかは，逆転が無い．

第18図は $N=1000$, $n=15$ の図で，中央部と右裾に弱い逆転が見られるが，他に逆転はなく，$N=100$ の例に



第19図　90%変動範囲内の不規則なヒストグラム．
$N=1000$, $n=19$

くらべてはるかに規則的な，母集団に近い分布を示している．

第18図より階級数が少ない $N=1000$, $n=11$ の図が第18図より更に規則的になるのは明らかであるから，その図は省略し，第19図には，$N=1000$, $n=19$ の場合を挙げる．図に見られる通り，両裾の部分に浅い逆転があり，中央部にやや深い逆転がある．斜面部は2階級ずつほぼ平坦になっていて，逆転が起こらないぎりぎりの状態である．

以上を要約すれば，$N=100$ のときは，（1），（2），（3）式による階級数のヒストグラムは，その全領域で，かなり顕著な度数逆転が容易に起こりうる．つまり，すべて階級数が過大である．また，$N=1000$ のヒストグラムは，（1）式，（2）式では中心部と両裾で弱い逆転が起こりうるほかは，度数逆転は非常に起こりにくい．（3）式の場合も斜面部で逆転が起こることは少ない．前にも述べたように，隣接階級の母集団度数差が小さいモード付近と両裾の弱い逆転の可能性はどんな場合も避けられずそのことを知っていれば母集団分布の推定に支障はない．重要なのは斜面部の逆転である．このことを考えると，（3）式の階級数19くらいが許容できる限界に近いと考えられ，この意味で（1）式，（2）式の階級数は過少と言える．結果的には，この節の検討は，前節の結論を再確認したことになる．同じ90%変動範囲を使ったので，当然という見方もできよう．

6.4. 中心誤差率を基準とした階級数の式（正規分布）

第6.2. 節と第6.3. 節の結果から，従来の階級数の式（1），（2），（3）は，ヒストグラムを局所的に見た標本誤差に関しては，適切でないことがわかったので，この節では，局所的な（つまり階級別に見た度数の）標本誤差を基準として，階級数の式を求める．

第3表　モデルヒストグラムの中心階級の誤差率（$\alpha=90\%$）．

N	n	式	D_U/k	D_L/k	図の番号	
					ヒストグラム	度数逆転
100	10	(1)	40%	32%	10	15
100	8	(2)	34%	29%	—	16
100	5	—	22%	21%	11	17
1000	18	(3)	10.4%	10.0%	14	19
1000	15	(1)	8.7%	8.4%	13	18
1000	11	(2)	6.3%	6.2%	12	—

この場合，ヒストグラム全体の標本誤差の大きさの代表値としては，モデルヒストグラムの度数最大の中心階級の誤差を，その度数で割った誤差率を使うのがよい．次にその理由を述べる．なお，信頼係数は前と同様90%とした．

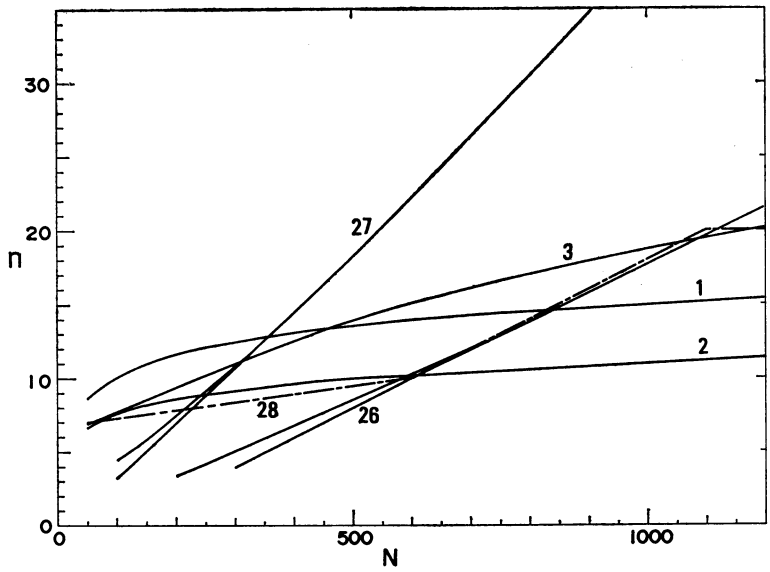
第10～14図を見ればわかるように，信頼区間の幅は中心階級で最も広く，端へゆくほど狭いから，中心階級での幅をおさえておけば，他の階級での幅は，それ以下に確保される．また，標本の大きさがちがうヒストグラムを描くとき，縦軸のスケールが同じならば N が大きいほど中心階級の度数も大きく，ヒストグラムの背が高くなるが，ふつうこのような描き方はせず，スケールを変えて，扱いやすいほぼ一定の高さにそろえる．つまり，縦軸のスケールは，ほぼ中心階級（最多度数階級）の度数に逆比例させる（第2図参照）．従って，視察した時のヒストグラムの標本誤差による不規則の程度は，誤差 D_L, D_U ではなく，これを中心階級の度数で割った，誤差率で表わされる．結局，N がちがっても中心階級の誤差率が同じならば，図の上では同程度の標本誤差があるから，この値を許容できる一定値に保つように，N と n の関係を決めるのが合理的である．

モデルヒストグラムについて，中心階級の下方誤差率 D_L/k 及び上方誤差率 D_U/k を求めると，第3表のようになる．表には，対応する既出の図の番号も示してある．

k は $N/2$ より小さいから上方誤差率のほうが下方誤差率より大きい．それ故ヒストグラム全体の誤差の程度を保証する基準値としてこの D_U/k を採用し，これをヒストグラムの「中心誤差率」と呼び，次の記号で表わす．

$$P_M=D_U/k \quad (\alpha=90\%,\ \text{中心階級}) \qquad (25)$$

第3表のこの値と，誤差を表現した第10～19図を対比してみると，P_M が10%ならばヒストグラムが十分信頼



第20図　標本の大きさNと階級数nの関係式（数字は式の番号）.
1：$n=5\log N$　2：$n=1+3.32\log N$　3：森（正規分布）
26：$\varphi_1(N)$，（$P_M=10\%$，正規分布）　27：$\varphi_2(N)$，（$P_n=20\%$，正規分布）
28：提案した実用式

第4表　$P_M\leqq10\%$，$P_M\leqq20\%$に最小限必要な標本の大きさN.

n	$P_M\leqq10\%$	$P_M\leqq20\%$
7	440	200
6	390	170
5	330	130

できることがわかる．しかしこの条件では，Nが小さいとき階級数が小さくなってヒストグラムが描けなくなるので，一応第11図，第17図の程度の誤差を許容するとすれば，$P_M=20\%$が基準となる．それ故，次の2通りの条件，それぞれについて，Nとnの関係式を求めてみた．

$$P_M(N,\ n)=10\%\rightarrow n=\varphi_1(N) \qquad (26)$$

$$P_M(N,\ n)=20\%\rightarrow n=\varphi_2(N) \qquad (27)$$

条件を満足する階級数は，Nを固定して種々のnについて求めたP_Mの値を内挿して決定した．計算の結果から得られた$\varphi_1(N)$，$\varphi_2(N)$のグラフは，従来の（1），（2），（3）式のグラフと共に，第20図に描いてある．この図には鎖線の折線も1本描いてあるが，これについては後の節で述べる．

今回求めた$\varphi_1(N)$，$\varphi_2(N)$のグラフは，従来の（1），（2），（3）式より勾配が急であって．Nが小さい所では従来の式より小さい階級数を与え，Nが大きい所では従来の式より大きな階級数を与えて，かなり傾向がちがう．また，グラフが左下の部分で枝分かれしているのは，モードが階級の中央にあるときと端にあるときで中心誤差率がちがうためで，モードの位置により2本の分枝の間を変動し，階級中央のときP_Mが最小（P_M一定ならばnが最大）になる．

第20図の$\varphi_1(N)$，$\varphi_2(N)$の2本の曲線を，階級別度数の標本誤差が小さい領域（$P_M<10\%$）大きい領域（$10\%<P_M<20\%$），非常に大きい領域（$P_M>20\%$）の境界と見れば，従来の式の標本誤差から見た性格が良くわかる．

次に，ヒストグラムが作れる階級数の下限値を7，6，5として，標本誤差を$P_M\leqq10\%$，$P_M\leqq20\%$におさえるのに必要な最小限のNの値を図から求めてみると，第4表のようになる．ここで$\varphi_1(N)$，$\varphi_2(N)$が2本に枝分かれした部分は，その中間の値を読んだ．

この表から，階級数6か7で，データが400個あれば標本誤差から見て十分信頼できるヒストグラムが描け，200個あれば一応よいが，100個程度では大きな標本誤差

第5表　(28) 式による階級数 n.

N	50	100	200	300	400	600	800	1000	1100以上
n	7.0	7.3	7.8	8.4	8.9	10.0	14.0	18.0	20.0

を覚悟して利用するほかはないことがわかる.

7. 他の誤差も総合した階級数の実用式

第6節以降の調査で得たヒストグラムの局所的な標本誤差の大きさや性質は，正規分布を仮定したモデルによって得たものではあるが，この誤差は直接的には N と k で決まり，分布型には依存しない．それ故，調査結果は必ずしも正規分布でなくても，これと似た形の一般の一山型分布についても，ほぼ同様と見てよい（詳しくは本節の後半で述べる）．それ故，前節までの結果から見ると，通常最も頻繁に現われる一山型のふつうの分布について，ヒストグラムの信頼性（つまりその形の安定度，あるいは標本ごとの形の不規則変動の大きさ）を決定する階級別度数の標本誤差の大きさや傾向を，従来のどの階級数の式も全く反影していない．このことと，従来の（1）式，（2）式がほとんど根拠のない式であり，（3）式の根拠であるヒストグラム全体としての適合性よりも，局所的な不規則性の程度のほうが利用上重大であるという判断から，標本誤差に根拠を置いた，新しい階級数の実用的な式を作ることにした.

この式を作る際の基本的な考え方は次の通りである.

（1）ヒストグラムの標本誤差が十分小さい，$P_M=10\%$ を基本とする.

（2）N が小さいときは，階級数が少ないための各種系統的誤差の増大を考慮し，標本誤差の増大を覚悟して階級数を増やす.

（3）上記に関連して，ヒストグラム作成の最低条件を，$N=50$，$n=7$ とする.

（4）計算，作図の手数とその効果を考慮し，階級数に上限値 $n=20$ を設ける.

（5）以上により階級数の式は3領域に分かれるが，各部は1次式で表現する.

このような考え方で，結局，次の式を得た.

$$\left.\begin{array}{ll} n=0.0055N+6.73 & (50\leqq N\leqq 600) \\ n=0.020N-2.00 & (600\leqq N\leqq 1100) \\ n=20 & (1100\leqq N) \end{array}\right\} \quad (28)$$

第5表は，この式による階級数である.

(28) 式は，既出の第20図に鎖線で描いてある.

第6表　(28) 式の中心誤差率 P_M（正規分布）.

N	P_M	P_M の傾向
50～ 230	49～20%	N の増加により減少
230～ 600	20～10%	N の増加により減少
600～1100	10%	一　定
1100以上	10%以下	N の増加により減少

また，この式の標本誤差を中心誤差率で示したのが第6表である.

(28) 式と従来の式との数値的比較は第20図の通りであるが，N が600以下では（2）式と同じ又は1小さい程度でかなり一致し，（3）式とは $N=50$ と $N=1000$～1100の2か所でほぼ一致し，それ以外では（3）式より小さい.

ここで，(28) 式を求めた（1）～（5）の条件及び適用条件等について，説明を補足する.

1）ヒストグラムは母集団分布曲線に対して数種類の系統的誤差があるが（第2報で扱う），共通して，誤差は階級幅が広いほど大きい．つまり標本誤差とは逆に，階級数 n を大きくするほど誤差が小さくなる．通常の一山型分布であれば，n が10以上あればほぼ十分であるが，それ以下は望ましくない．それ故 $P_M=10\%$ を採用する N の下限は n がちょうど10になる $N=600$ とし，N が600以下の所は，第20図でこの点（600, 10）と上記条件（3）の点（50, 7）を結ぶ直線を採用した.

2）N が600以下では，N の減少につれて標本誤差が増すが，誤差のおよその程度は第6表から推定できる.

3）ヒストグラム作成の最低条件を $N=50$，$n=7$ としたのは，このくらいのデータ数でヒストグラムをどうしても描きたい場合が起こるであろうという常識的な感覚を条件化したもので，この点に限れば，従来の（2）式を採用したと言ってもよい．この場合 P_M は約50%と非常に大きいから，著しく不規則なヒストグラムになる可能性はかなり大きいが，必ずそうなるとは限らないし，不規則な凹凸の実在性を，第4～7図でチェックすることもできる.

4）系統的誤差は n が大きいほど小さいから，標本誤差を $P_M=10\%$ でおさえておけば，N が大きいほど n が大きくとれて，系統的誤差からもヒストグラムの信頼度が増す．しかし元々特殊な分布は N だけから n を決めるのは不可能で，この式の対象外とする

第7表　範囲の変動率　$2\sigma_{RN}/W_N$（正規分布）.

N	50	100	400	1000	5000
$2\sigma_{RN}/W_N$	31%	25%	21%	17%	14%

ので，nは20もあれば十分で，それ以上は計算，作図の労が増すだけで，効果が少ない．

5）この式の適用の対象は，変域の中央部にモードを持ち，両側で確率密度が0に漸近する一般の一山型分布であって，例えば次のような特殊な分布は対象外である．

①雲量，日降水量などの変域が有界な分布，②二山型分布，③著しく歪んだ分布，④著しく裾が伸びた分布．

これらは分布の型に応じた階級分けが必要であり，等間隔区分が不適当なものもある．これらについても．分布型がわかり，階級区分が決まれば階級別の期待度数が求まるから，標本誤差を第4～7図で評価できる．標本誤差を大きくしないためには，期待度数が小さい階級をできるだけ減らすような区分を考えるべきであることが，今回の調査から言える（分布型に適した階級区分は第2報で扱う）．

6）(28) 式が適用されるとした一山型分布でも，歪んでいたり，尖度が正規分布とかなりちがっていたりすれば，第5表の誤差の量的な表現は適用できない．しかし一応の基準として階級数の式を利用するのは支障ないと考えられる．

8. 範囲の標本変動の大きさと影響

モデルヒストグラムでは，その全体の幅として，$W_N=2\tilde{x}_N$を採用したが，実際の標本でヒストグラムを描くときは，標本の範囲，$R_N=x_N-x_1$を階級数nで割って階級幅を求めて階級区分するから，階級幅は範囲の標本変動と同じ割合で変動する．ここでは，この変動の大きさと影響を考察する．

規準正規変量の標本範囲 R_N の平均値と標準偏差，$\bar{R}_N$，σ_{RN} は，$N\leqq100$については，統計数値表（参考文献参照）に値があり，$N=50$～5000 については，前記森の論文に，R_N の変動係数 $\sigma_{RN}/\bar{R}_N$ を求めた数値がある．また，筆者が採用した W_N と，$\tilde{R}_N$ を結びつける次の関係が知られている．

$N\to\infty$ のとき $\tilde{R}_N\to W_N$（既出(16)式）および $\bar{R}_N\to\tilde{R}_N+2\gamma/\alpha_N$．つまり，次のように書ける．

$$N\to\infty\text{ のとき }\bar{R}_N=W_N+2\gamma/\alpha_N,\ \text{ただし}\ \gamma=0.5772157,\ \alpha_N=Nf(u_N) \quad (29)$$

ここで γ は Euler の定数，u_N は特性極値，f は確率密度で，α_N は特性極値強度と呼ばれる量である．

森による R_N の変動係数の値を（29）式で変換して，R_N の W_N を基準とした変動率として，$2\sigma_{RN}/W_N$ の値を求めてみると，第7表のようになる．

表のように，範囲の標本変動を $\pm2\sigma_{RN}$ で見積れば，その値は，通常使われる標本の大きさ50～5000では，モデルヒストグラムの幅 W_N の ±15～±30% 程度で，N が小さいほど比率は大きい．したがって，ヒストグラムの階級幅も，これと同じ比率で変動する．

なお，正規変量の範囲 R_N の分布は，正に歪み，正規分布よりやや尖った分布であるから，表の $\pm2\sigma_{RN}$ が確率95%の範囲とは言えないが，変動の大部分をカバーする範囲と考えることはできるであろう．

範囲の変動に伴う階級幅の変動が誤差に及ぼす影響は，ほぼ次のように考えられる．

範囲が大きく変動するのは，分布の裾の部分の標本変動が原因であって，元々母集団度数が小さい階級で，標本変動で度数が0になれば範囲が狭くなり，母集団度数が殆んど0の階級にたまたまデータが現われれば範囲が広がる．範囲を変動させるデータはごく少数であって，度数が多い階級の標本分布には殆んど差違が無い．それ故分布の主要部については，階級幅が狭まっただけ（広がっただけ），同じ標本を細かく（粗く）階級区分したことになる．

階級幅が狭まれば階級別度数が減って，偶然誤差（標本誤差）は大きくなるが，各種の系統的誤差は小さくなる．広がった場合はその逆である．例えば，モデルヒストグラムの階級数が10のとき，範囲が W_N の±20%変動すれば，誤差は階級数を8階級から12階級まで変えたことに相当する．

特に，ごく少数のデータが他と著しく離れた値のときは，これを除いた範囲を階級数で割って階級幅を決める必要がある．

あとがき

研究のまとめは第2報で行なうが，主な結果はアンダーラインをつけてある．ヒストグラムの実例での標本誤差の評価の例も第2報で述べる．標本誤差に関しては，ヒストグラムと度数折線は同じである．

文　献

Brooks, C.E.P. and N. Carrutheres, 1953: Handbook of Statistical Methods in Meteorology, Meteorological office 538, Her Majesty's Stationary Office, London.

Gumbel, E.J., 1958: Statistics of Extremes, Columbia Univ. Press.

気象庁統計課，1960：正規母集団の標本分布，気象庁観測技術資料，No. 17.

森　俊夫，1974：ヒストグラムの最適級間隔について，応用統計学，4，No. 1，17-24.

日本規格協会，1972：統計数値表.

小河原正巳ほか，1957：統計公式および図表とその使い方，地人書館，気象学講座，19.

Sturges, H.A., 1926: The Choice of a Class Interval, Journ. Amer. Stat. Assoc., 21, 65-66.

気象学会および関連学会行事予定

行　事　名	開催年月日	主催団体等	場　　所
第3回夏季大学「新しい気象学」教室	昭和56年8月4日～6日	日本気象学会関西支部	大阪市立労働会館
第15回夏季大学「新しい気象学」教室	昭和56年8月10日～13日	日本気象学会	日本教育会館
グローバル水収支の変動に関するシンポジウム	1981年8月9日～15日		英国オックスフォード
IAMAP Third Scientific Assembly	1981年8月17日～28日		西独ハンブルグ市
月例会「大気数値シミュレーション」	昭和56年9月4日	日本気象学会	気象庁第1会議室
第19回粉体に関する討論会	昭和56年10月28日～30日	日本薬学会ほか	岐阜信用金庫本店大ホール
第7回「リモートセンシングシンポジウム」	昭和56年11月17日～18日	(社)計測自動制御学会	機械振興会館
第28回風に関するシンポジウム	昭和56年11月27日	日本建築学会ほか	東京大学生産技術研究所第1会議室
昭和56年日本気象学会秋季大会	昭和56年12月1日～3日	日本気象学会	愛知県中小企業センター