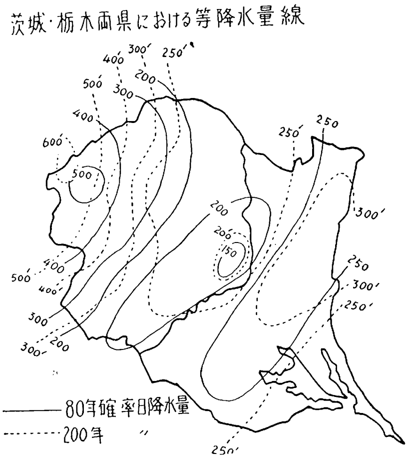


確率雨量

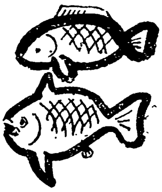
（気象用語）

最大日降水量の値は200mm，300mmという豪雨であつて，かつ地域によつて著るしい差異がある。このような豪雨はどこでも毎年起るものでなくて，10年間に１回とか，20年間に１回とかいうようなものである，それで各地についてこのような豪雨が何年間に起るものか，またその量が何mmかということを過去の雨量を基礎にして確率雨量を求めると次の表のようになる。

地名	統計年数	400年	200年	80年	40年	28年	8年	4年
		mm						
水戸	54	318	306	290	281	194	136	126
下妻	44	325	290	250	224	209	158	124
竜ケ崎	52	263	236	203	190	182	141	110
大子	44	268	243	217	207	182	140	121
大津	46	302	279	261	175	166	141	129
笠間	54	315	304	286	233	205	143	129
麻生	54	245	223	208	173	156	121	100
小瀬	48	271	241	216	197	185	131	108
神峯山	39	349	317	299	285	252	182	145
鉾田	54	224	215	205	157	149	121	105

この表から例えば，水戸で１日200mmの降水量を記録したとすれば，雨量確率からみて概ね30年に１回の豪雨であつたことが解るわけである。

また本県の大河川の上流域は栃木県に属するものが多いので，栃木県の値も併記して考察してみると，その共通した傾向は等雨量線が山脈の走向に一致していることである。確率最大日降水量の値は，阿武隈山脈から本県中部の山沿地方では大きく，栃木県の北西山岳地帯は極めて大きい。これら山岳の中間の平地や八溝山脈及び本県の南東部では少なく，殊に八溝山脈の西側には山脈に沿つて極めて小さい地域がある。また豪雨の起る確率は地形により著るしく異なり山岳の風上側では最も大きい，豪雨のもたらす最大の原因は台風であつて台風が接近するとき本県における一般気流は北東風であるので，阿武隈山脈と八溝山脈の南部（筑波山付近）の東側では確率雨量は大きくなる。



標本調査への手引 (4)

総理府統計局　高橋史朗

第1部　標本調査の理論（つづき）

10　期待値の推定

ひきつづいて，チェビシェフの不等式の，左辺の中括弧{ }のなかの不等式

$|t-E(T)|<\lambda.\sigma(T)$

についてかんがえてみます。

この不等式は，期待値E(T)を中心に，左右に，標準偏差σ(T)のλ倍の幅の区間をとつたとき，その区間のなかに，実現値tがはいる，ということをしめしております。

したがつて，もしも，この不等式が常に成り立つならば，実現値tは，許容区間の絶対誤差を標準偏差σ(T)のλ倍にとつた，期待値E(T)の推定値になつているわけです。あるいは標準偏差σ(T)を期待値E(T)で除せば変動係数CV(T)になりますから，実現値tは，許容区間の相対誤差を変動係数CV(T)のλ倍にとつた，期待値E(T)の推定値になつているわけです。

しかし，この不等式は常には成り立ちません。すなわち，チェビシェフの不等式によると，この不等式は確率（I－δ）以上で成り立つとしか言えないのです。いいかえると，成り立たない危険が，確率δ以下で，存在しているわけです。

したがつて，ここに，チェビシェフの不等式から，次のような重要な結果が導かれます。

> 高々，確率δの危険を認めるならば，確率変数Tの実現値tは，許容区間の絶対誤差を標準偏差σ(T)のλ倍にとつた，あるいは，相対誤差を変動係数CV(T)のλ倍にとつた，期待値E(T)の推定値になつている。

そこで，いま，ある統計を「正確」にもとめることにして，まず，その真値をA，許容区間の絶対誤差あるいは相対誤差を，それぞれ，ε, η(注)としてみましよう。

注　ε, ηーギリシヤ文字，エプシロン，イータの小文字.

次に，高々確率δの危険は，これを，適当に小さくするという条件で認め，さらに，次の２つの条件を満たす確率変数Tを探がして，その実現値tをもとめてみましよう。

すると，この実現値tは，あらかじめ，絶対誤差εあるいは相対誤差ηによつて，決めておいた許容区間のなかにはいる，この統計の「正確」な推定値になります。

条件 1　期待値E(T)は，真値Aに等しい。

$E(T)=A$

条件 2　標準偏差σ(T)あるいは変動係数CV(T)のλ倍は，それぞれ，絶対誤差ε，相対誤差η以下である。

$\lambda.\sigma(T) \leqq \varepsilon$

$\lambda.CV(T) \leqq \eta$

なお，このように，ある統計を，その実現値からもとめるために採る確率変数Tを，その統計の**推定子**といいます，

これが，**抽出理論の基礎**です。結局，抽出理論は，この基礎の上に立ち，もとめようとする統計にたいして，この２つの条件を満たし，しかも，もつとも迅速・安価に，その実現値がもとめられる確率変数T，すなわち，推定子は，どのようにすれば得られるかを明らかにしているわけです。

ところで，さきの確率δの危険は避けられないかということですが，それは，確率の理論を利用しているかぎり，不可能です。なお，この危険は，標本調査だけにかぎらず，測定を必要とする自然あるいは社会科学のあらゆる分野にわたつて存在しております。

これで，抽出理論の基礎が明らかになつたわけですがしかし，そこには，なお，次のような２つの問題点が残されております。

問題点 1　まず，確率δはどのくらいにとればよいかという問題があります。普通，$\delta=5\%$ぐらいにとればよいと言われておりますが，それだけでは片付かないようです。この問題は，また，あとで，詳しく述べたいとおもいます。

問題点 2　次に，δからλをもとめる関係式が問題になります。チェビシェフの不等式では，これを$\lambda=\frac{1}{\sqrt{\delta}}$としておりますが，これは，確率変数Tに，まつたく条件がつかない場合です。

したがつて，推定子に採るというような条件がつく場合に，この関係式によつて，δからλをもとめれば，λは一般に，非常な過大評価となつて，好ましくありません。たとえば，推定子には，**中心極限定理**の成り立つものがかなりありますが，その場合には，次の関係式によつて，δからλをもとめて十分です。

$$\frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{\lambda} e^{\frac{-u^2}{2}}\,du = \mathrm{I} - \delta$$

かなり式の形が変わるので，比較しにくいのですが，たとえば，$\delta = 5\%$としますと，第5表にしめすように確率変数Tに，まつたく条件がつかない場合は，λ=4.48としなければなりませんが，中心極限定理が成り立つ場合には，λ=1.96として十分です。この問題も，また，あとで，詳しく述べたいとおもいます。

第 5 表

δ	λ		I－δ
	まつたく条件がつかない場合	中心極限定理が成り立つ場合	
%			%
1	10.00	2.58	99
2	7.08	2.33	98
3	5.78	2.18	97
5	**4.48**	**1.96**	**95**
10	3.17	1.65	90
20	2.24	1.29	80
30	**1.83**	**1.04**	**70**
40	1.59	0.85	60
50	1.42	0.68	50

注　不等式の性質から，λは切り上げにしてあります。

なお，中心極限定理については，説明を省略します。

11　サイコロと乱数表

サイコロを振ると，1から6までの6個の目が，それぞれ，確率$\frac{1}{6}$で現われてきますが，いま，このうち，たとえば，6の目が出ても，これをすべて捨てて，振り直すとしたらどうなるか，かんがえてみます。

すぐ，お分かりのように，このようにすれば，1から6までの6個の目ではなく，1から5までの5個の目がそれぞれ，確率$\frac{1}{5}$で現われてきます。すなわち，丁度1から5までの5個の目のあるサイコロが作られたことになるわけです。

おなじようにして，4個，3個，2個の目のあるサイコロが作られます。

ところで，われわれは，さらに，7個以上の目のあるサイコロを作りだすこともできます。もちろん，そのようなサイコロを実際に作ることはできませんが，そのようなサイコロを振つたのと，まつたく等しい結果を作りだすことができるのです。それには，乱数表がもちいられます。

乱数表は，その一例を，第6表にしめしたような表ですが，これから，この乱数表をもちいて，7個以上の目のあるサイコロを作りだしてみましょう。

第 6 表
乱 数 表

6	2	6	8	6	1	2	3	9	8	4	7	6	9	2	5	5	4	0	0
2	2	0	0	3	4	1	6	9	6	4	0	9	8	3	9	7	8	7	9
3	6	6	2	2	7	8	0	4	3	2	6	5	8	2	0	1	3	6	9
1	9	2	2	1	0	1	0	7	3	4	5	7	1	9	6	9	2	8	3
8	9	3	0	4	7	4	4	3	2	2	7	2	9	0	0	9	6	2	0
4	6	8	2	3	0	3	6	6	9	4	9	8	2	0	9	1	0	1	5
9	6	2	0	7	7	5	4	5	9	9	2	9	6	6	3	2	2	2	3
3	9	2	7	0	5	4	6	0	2	4	0	7	5	4	8	4	1	6	9
8	3	4	3	6	4	9	1	1	2	5	4	1	3	7	1	7	6	8	0
7	5	7	0	9	4	5	2	7	4	7	0	4	5	9	9	8	9	7	3
0	1	7	9	4	7	2	1	8	1	7	6	0	2	3	3	0	5	5	1
6	0	3	1	8	1	1	8	4	8	0	5	4	6	8	4	1	2	4	1
6	3	1	3	9	8	9	4	0	4	0	3	5	8	0	6	2	4	5	3
6	4	7	1	2	6	6	1	9	2	7	1	8	2	2	1	8	1	2	7
7	8	1	6	4	9	1	1	6	9	3	4	8	3	5	7	1	3	4	4

例(1)　まず，76個の目のあるサイコロを作つてみます。いま，たとえば，乱数表の左端の2列をとり，上から下にみてゆくと，62，22，36，19，89，46，96，39，83，75，……というようになつていますが，これは，丁度，00から99までの100個の目のあるサイコロを，くり返して振つた結果とみなしてよいのです。したがつて，1から76までの76個の目のあるサイコロを，くり返して振つた結果を作るには，たとえば，01は1の目，02は2の目，……，76は76の目とみなして，あとの00や77から99まではすべて捨ててしまえばよく，その結果は，62の目，22の目，36の目，19の目，46の目，39の目，75の目，……という風になります。

例(2)　次に，23個の目のあるサイコロを作つてみます。これは，例(1)とまつたくおなじように作つても，差支えないのですが，ただ，そうしますと，01は1の目，02は2の目，……，23は23の目とみなし，あとの00や24から99までをすべて捨ててしまうことになるので，これではあまり捨てる方が多すぎて，不経済です。そこで，このような場合には，01，31，61は1の目，02，32，62は2の目，……，23，53，83は23の目というようにみなし，あとの00や24から30まで，54から60まで，84から99までをすべて捨てるというようにすればよく，その結果は，2の目，22の目，6の目，19の目，16の目，9の目，23の目，15の目，……という風になります。

なお，この技巧は，すこしぐらい不経済でも，あつさりとやるのがよく，あまり経済ばかりかんがえて，細々とやるのはよくありません。

〔練習問題〕　814個の目のあるサイコロを，くり返して振つた結果を作つてみてください。

12　期待値と算術平均

標本調査の例として，たとえば，次のような場合をかんがえてみます。

――東京都で営業している飲食店について，この９月の平均営業利益および１従業員当たりの平均売上高をもとめる――

ここで，平均営業利益および１従業員当たりの平均売上高は，それぞれ，**算術平均**および**2個の算術平均の比率**の一例となつております。

まず，飲食店の総数を，たとえば，26,204軒であるとして，その名称と所在地とをしめした**名簿**を用意します。そして，飲食店に，１番，２番……，26,204番と，通し番号をつけます。

次に，１から26,204までの26,204個の目のあるサイコロを用意して(これは，乱数表によつて，用意できます)たとえば，その１の目に１番の飲食店，２の目に２番の飲食店，……，26,204の目に26,204番の飲食店というように対応させます。

このようにすると，このサイコロを振るたびに，26,204軒の飲食店のうちのどれかが，それぞれ，確率$\frac{1}{26,204}$で現われてくることになります。

(1)　平均営業利益をもとめる場合

いま，飲食店の営業利益に着目しますと，このサイコロを振るたびに，どの飲食店かの営業利益が，それぞれ確率$\frac{1}{26,204}$で現われてくることになります。
そこで，これをXと表わせば，このXは，１次元の確率変数です。

この確率変数Xの期待値E(X)，標準偏差σ(X)，変動係数CV(X)は，26,204軒の飲食店の営業利益の算術平均M_x，標準偏差σx，変動係数CV_xに，それぞれ，等しいという関係があります。

期待値　E(X) ＝ **算術平均** M_x
標準偏差　σ(X) ＝ 標準偏差 σx
変動係数　CV(X) ＝ 変動係数 CV_x

ここで，とくに，期待値E(X)が，いま，もとめようとしている平均営業利益（すなわち，算術平均M_x）に等しいということに注意してください。

(2)　1従業員当たりの平均売上高をもとめる場合

いま，飲食店の従業員数と売上高に着目しますと，このサイコロを振るたびに，どの飲食店かの従業員数と売上高が，それぞれ，確率$\frac{1}{26,204}$で現われてくることになります。そこで，これを（X，Y）と表わせば，この（X，Y）は２次元の確率変数です。

この確率変数（X，Y)の相関係数ρ(X，Y)は26,204軒の飲食店の従業員数と売上高との相関係数ρxyに等しいという関係があります。

相関係数 ρ（X，Y）＝ 相関係数 ρxy

また，(1)とおなじように，確率変数XおよびYの期待値E(X)，E(Y)，標準偏差σ(X)，σ(Y)，変動係数CV(X)，CV(Y)は，26,204軒の飲食店の従業員数および売上高の算術平均M_x，M_y，標準偏差σx，σy，変動係数CV_x，CV_yに，それぞれ，等しいという関係があります。

E(X) ＝ M_x　　　E(Y) ＝ M_y
σ(X) ＝ σx　　　σ(Y) ＝ σy
CV(X) ＝ CV_x　　　CV(Y) ＝ CV_y

ここで，とくに，期待値E(X)，E(Y)が，いま，もとめようとしている１従業員当たりの平均売上高の分母，分子（すなわち，算術平均M_x，M_y）に，それぞれ，等しいということに注意してください。

計算機のむかしばなし

計算の機械は数千年ものあいだ玉をならべたソロバンにしか頼るしかなかつた。1642年仏ローエンのパスカルは19才のとき，父の税務事務の繁忙から免れようとして世界最初の自動計算機を発明した。歯車を廻転させ，つぎつぎに上位の歯車で積算してゆくものであつた。

彼はケタ数をふやそうと50台以上の試作品を作つたが，ついに技術的に歯車の抵抗力の集積を克服することができずに失敗に終つた。

―文芸春秋より―