創立18周年記念講演会

昭和 37 年6月9日午後1時半より当所の講堂で，創立 18 周年を記念して，下記次第により公開講演会を行なった。

あいさつ

所長　末綱恕一

統計推論における信頼性について

第一研究部 第一研究室長　藤本熙

日本のホワイトカラー

所員　鈴木達三

水資源の統計数理的考察

付属統計技術員養成所長　菅原正巳

昭和36年度研究発表会アブストラクト

昭和37年3月29日，30日の両日に，当所の講堂で，昭和36年度の研究発表会を行った．

あいさつ　末綱恕一

昭和36年度第1研究部の研究概要

松下嘉米男

定常過程の積分値の有限和による近似について

藤井光昭

$x(t)$ を時間パラメター t が連続値をとる様な弱定常確率過程とし，実数値をとるものとする．更に $Ex(t)=0$ とし，covariance function $R(t)$ は $t=0$ で連続であると仮定する．そしてスペクトル分布函数を $F(\lambda)$ とする．

この時，積分値

$$\int_0^T x(t)dt$$

を

$$\sum_{j=1}^n x\left(\frac{j}{n}T+\xi_j\right)\frac{T}{n}$$

なる形で近似するわけであるが．ここで $\{\xi_j\}$ はすべて互に独立な平均0の確率変数で原点に関して対称な確率分布 $G(x)$ をもつものとする．そこでここでは誤差

$$V=\underset{T}{E}\underset{\Omega}{E}\left(\int_0^T x(t)dt-\sum_{j=1}^n x\left(\frac{j}{n}T+\xi_j\right)\frac{T}{n}\right)^2$$

を最小にする為に ξ_j いかにをとればよいか，云いかえれば $G(x)$ をいかに定めたらよいかという問題として考える．

$\varphi(\lambda)=\int_{-\infty}^{\infty}\cos x\lambda dG(x)$ とおけば

$$V=\int_0^T\int_0^T R(t-s)dtds$$
$$+\int_0^{\infty}\left[\frac{2T^2}{n}+A(\lambda)\varphi(\lambda)^2-B(\lambda)\varphi(\lambda)\right]dF(\lambda)$$

となる．但し

$$A(\lambda)=\frac{4T^2}{n^2}\sum_{m=1}^{n-1}(n-m)\cos\frac{mT\lambda}{n}$$

$$B(\lambda)=\frac{4T}{n\lambda}\left(2\sum_{j=1}^{n-1}\sin\frac{jT\lambda}{n}+\sin T\lambda\right)$$

である．

一般の $F(\lambda)$ に対して V を最小にする様な $\varphi(\lambda)$ を定めることは難かしい為，$F(\lambda)$ が step function の場合について考えてみる．$F(\lambda)$ が $\lambda=\lambda_0$ でのみ jump をもつ場合には

$$\frac{2T^2}{n}+A(\lambda_0)\varphi(\lambda_0)^2-B(\lambda_0)\varphi(\lambda_0)$$

を最小にする様な $\varphi(\lambda_0)$ を求めることになり，$A(\lambda_0)$ と $B(\lambda_0)$ に依存するが，例えば

（i）$A(\lambda_0)>0,\quad 0<\dfrac{B(\lambda_0)}{2A(\lambda_0)}<1$ の時

$\varphi(\lambda_0)=\dfrac{B(\lambda_0)}{2A(\lambda_0)}$ で定まる $G(x)$

（ii）$A(\lambda_0)>0,\ B(\lambda_0)=0$ の時

$\left[-\dfrac{\pi}{\lambda_0},\ \dfrac{\pi}{\lambda_0}\right]$ における一様分布

（iii）$A(\lambda_0)<0,\ B(\lambda_0)=0$ の時

原点で確率1をもつ分布

等といった様になる．

ある二標本の問題

藤　本　　熈

母集団 Π_1,Π_2 からの独立な標本変量 $X_1,\cdots,X_m$; $Y_1,\cdots,Y_n$ は連続な一次元の確率分布をもつとする．$\varphi(x_i,y_j)=1$ ($x_i>y_j$ ならば)，0 ($x_i<y_j$ ならば) とし，$U_{m,n}=\sum_i^m\sum_j^n\varphi(x_i,y_j)$, $i=1,\cdots,m$; $j=1,\cdots,n$ をつくると，$E\left(\dfrac{U_{m,n}}{mn}\right)=P(X>Y)$ はよく知られる．ここで更に一つの標本変量 Z が X か Y のいづれかであることが知られるとき，$m+n+1$ 番目の観測値 z から，$Z\epsilon\Pi_1$ か $Z\epsilon\Pi_2$ かを判断する方法を考える．

今 $x_i<z$ なる x_i の個数を m^*, $y_j>z$ なる y_j の個数を n^* とすると，$Z\epsilon\Pi_1$ のもとでは

$$\frac{U_{m+1,n}}{(m+1)n}=\frac{U_{m,n}}{mn}+\frac{1}{m+1}\left\{\frac{n-n^*}{n}-\frac{U_{m,n}}{mn}\right\},$$

$Z\epsilon\Pi_2$ のもとでは

$$\frac{U_{m,n+1}}{m(n+1)}=\frac{U_{m,n}}{mn}+\frac{1}{n+1}\left\{\frac{m-m^*}{m}-\frac{U_{m,n}}{mn}\right\}.$$

ここで若しこの分類が正しいとすると，

$$E\left\{\frac{U_{m+1,n}}{(m+1)n}\middle|Z\epsilon\Pi_1\right\}=E\left\{\frac{U_{m,n+1}}{m(n+1)}\middle|Z\epsilon\Pi_2\right\}=P(X>Y)$$

であるが，そうでないと $P(X>Y)\gtreqless\dfrac{1}{2}$ に従って，

$$E\left\{\frac{U_{m+1,n}}{(m+1)n}\middle|Z\epsilon\Pi_2\right\}\gtreqless P(X>Y),$$

$$E\left\{\frac{U_{m,n+1}}{m(n+1)}\middle|Z\epsilon\Pi_1\right\}\gtreqless P(X>Y)$$

となるから，こうなることを防ぐ方向に Z を分類することにすると，

$$\frac{m^*}{m}>\frac{n^*}{n}\Rightarrow Z\epsilon\Pi_1,$$

$$\frac{m^*}{m}<\frac{n^*}{n}\Rightarrow Z\epsilon\Pi_2$$

と判断することになる．

又このような考えにもとずいて更に議論を展開することを考える．

周波数応答函数の統計的推定について

赤　池　弘　次

系への入出力の記録 $x(t)$, $y(t)$ が与えられた場合

$$C_{yx}(\tau)=\frac{1}{2T}\int_{-T}^{T-\tau}y(\tau+t)x(t)dt\qquad \tau\geqq0$$

$$=\frac{1}{2T}\int_{-T-\tau}^{T}y(\tau+t)x(t)dt\qquad \tau<0$$

$$C_{xx}(\tau)=\frac{1}{2T}\int_{-T}^{T-|\tau|}x(\tau+t)x(t)dt$$

とする．そこで $x(t)$, $y(t)$ ともに $\dfrac{1}{2\varDelta\tau}$ 以上の周波数成分のパワーが無視できるように $\varDelta\tau$ を定め次式により $\bar{p}_{yx}(f)$, $\bar{p}_{xx}(f)$ を算出する，

$$\bar{p}_{yx}(f)=\varDelta\tau\sum_{e=-h}^{h}e^{-2\pi ifl\varDelta\tau}C_{yx}^*(l\varDelta\tau)$$

$$\bar{p}_{xx}(f)=\varDelta\tau\sum_{e=-h}^{h}e^{-2\pi ifl\varDelta\tau}C_{xx}^*(l\varDelta\tau)$$

但し

$$C_{yx}^*(l\varDelta\tau)=C_{yx}(l\varDelta\tau)\qquad -h<l<h$$

$$=\frac{1}{2}C_{yx}(l\varDelta\tau)\qquad l=\pm h$$

$$C_{xx}^*(l\varDelta\tau)=C_{xx}(l\varDelta\tau)\qquad -h<l<h$$

$$=\frac{1}{2}C_{xx}(l\varDelta\tau)\qquad l=\pm h.$$

$\bar{p}_{yx}(f)$, $\bar{p}_{xx}(f)$ に次の smoothing を行なう．

$$\hat{p}_{yx}(f)=\sum_{n=-k}^{k}a_n\bar{p}_{yx}\left(f-\frac{n}{2h\varDelta\tau}\right)$$

$$\hat{p}_{xx}(f)=\sum_{n=-k}^{k}a_n\bar{p}_{xx}\left(f-\frac{n}{2h\Delta\tau}\right).$$

このとき系の周波数応答函数 $A(f)$ の推定値 $\hat{A}(f)$ は

$$\hat{A}(f)=\frac{\hat{p}_{yx}(f)}{\hat{p}_{xx}(f)}$$

として求められる.

実際の計算に際しては次式で与えられる T_f の大なる所では

$$T_f=-\frac{1}{2\pi}\frac{d(\arg A(f))}{df}$$

$$K_f\fallingdotseq\frac{1}{\Delta\tau}T_f$$

なる K_f を求め, $C_{yx}(l\Delta\tau)$ の代りに $C_{yx}((K_f+l)\Delta\tau)$ を用いて推定を行うことが必要となる. 更に上記計算式中の h, $\{a_n\}$ 等をどのように選定するかが問題となる. 筆者は $\{a_n\}$ の設計法を検討し

$k=2$, $a_0=0.64$, $a_1=a_{-1}=0.24$, $a_2=a_{-2}=-0.06$

なる型のものの使用をとくに提案している.

以上の研究の詳細は, Annals of the Institute of Statistical Mathematics, Vol. 14 に筆者の単独報告及び運輸技術研究所山内保文氏と筆者の協同報告として掲載される.

チェビシェフ型不等式の Sharpness について

石井恵一

ここで扱うチェビシェフ型不等式の問題とは, 次のようなものである.

Ω を空間, $\mathfrak{A}$ をその上の σ-集合体とし, $\boldsymbol{f}(\omega)=(f_1(\omega),\cdots,f_m(\omega))$ を Ω から R^m (m 次元実数空間) への可測写像とする. いま $(\Omega,\mathfrak{A})$ 上の未知の確率分布 P が $E_P(\boldsymbol{f})=\boldsymbol{M}=(M_1,\cdots,M_m)$ (与えられた値) によって規制されているとき, $\sup P(A)$ および $\inf P(A)$ の値如何という問題. (一方を考えれば充分である)

$P(A)\leqq\alpha$ 型の不等式を得るのに時々使われる方法は, $f(\omega)=a_0+a_1f_1(\omega)+\cdots+a_mf_m(\omega)$ の形の函数で, Ω 上 $f(\omega)\geqq\chi_A(\omega)$ なるものに対しては

$$P(A)=\int\chi_A(\omega)dP(\omega)\leqq\int f(\omega)dP(\omega)$$
$$=a_0+a_1M_1+\cdots+a_mM_m$$

であることより

$$P(A)\leqq\inf_{f\geqq\chi_A}(a_0+a_1M_1+\cdots+a_mM_m)$$
$$\equiv\lambda(A;\boldsymbol{M})$$

(ただし inf は上記の如き f のすべてにわたってとる).

しかし, この $\lambda(A;\boldsymbol{M})$ が sharp (これ以上小さくできない) な上界であるかどうかはこれだけでは分らない. 具体的な個々の場合には, 実際に $P(A)=\lambda(A;\boldsymbol{M})$ となる分布 P をみつけるなりして sharp であることを確かめることのできる場合もあるが, いつでも sharp であるとは限らないし, sharp であっても等号をとりうるとは限らない.

そこで, 一般に sharp な上界 (又は下界) はどのように定まるかを論じる. 問題は,

①$\lambda(A;\boldsymbol{M})$ はいかなる A, $\boldsymbol{M}$ に対して sharp か. 更に一般に, sharp な上界はどう与えられるか.

②上界の値を実際にとる分布 P が存在するか. また $\lambda(A;\boldsymbol{M})$ が sharp な上界のとき, $\lambda(A;\boldsymbol{M})$ の定義の式の inf を実際にとる f が存在するか.

③実際に sharp な上界を計算する便利な方法.

④更に, 考える分布の族を適当に狭くしたり, 広くした場合はどうか.

の4点である. これらについての結果は次の通り. 方法は凸集合の理論を用いる幾何学的方法が最も簡単である.

① $\mathfrak{M}$ を, $(\Omega,\mathfrak{A})$ 上のあらゆる確率分布に対応する $E_P(\boldsymbol{f})$ の全体とすると, $\mathfrak{M}$ は $\{\boldsymbol{f}(\omega);\omega\epsilon\Omega\}$ によって張られる m-次元凸集合であることが示されて,

i) $\boldsymbol{M}$ が $\mathfrak{M}$ の内点のとき, 任意の $A\epsilon\mathfrak{A}$ に対して $\lambda(A;M)$ は sharp である.

ii) 一般の $M\epsilon\mathfrak{M}$ に対しては, $S(\boldsymbol{M})$ を, $\boldsymbol{M}$ を相対的内点として持つ最大の凸部分集合 ($\mathfrak{M}$ の) とし, $\boldsymbol{f}^{-1}(S(\boldsymbol{M}))=\Omega'(\boldsymbol{M})$, $A\cap\Omega'(\boldsymbol{M})=A'$ とおけば, Ω の代りに Ω', A の代りに A' を考えて $\lambda(A';\boldsymbol{M})$ が $P(A)$ の sharp な上界になる. ($\boldsymbol{M}$ が $\mathfrak{M}$ に属するかとか, 内点であるかなどの判定は多くの場合簡単な方式でできる).

④の問題に対しては, まず, $\mathfrak{A}$ によるかどうかの問題は, $\boldsymbol{f}$ や A が可測になるような $\mathfrak{A}$ でさえあれば上界の値は同じになる. また分布の族を, $P\{\boldsymbol{f}^{-1}(m+1\text{ 点集合})\}=1$ となるような P のみに制限しても上界の値は変らない.

②の問題. Ω がコンパクトな Hausdorff 空間, $\mathfrak{A}$ がその上の位相的 σ-集合体, $\boldsymbol{f}$ が Ω 上で連続, $\boldsymbol{M}$ が $\mathfrak{M}$ の内点ならば, 任意の $A\epsilon\mathfrak{A}$ に対し, $E_{P^*}(A)=\boldsymbol{M}$ となる P^* が存在して,

$$P^*(\bar{A})=\lambda(A;\boldsymbol{M})(=\lambda(\bar{A};\boldsymbol{M}))$$

したがって, A が閉集合ならば等号をとりうる.

Ω がコンパクトでなく, 局所コンパクトであっても (従って, $\boldsymbol{f}$ が有界でなくても), $f_j(\omega)/(1+|f_1(\omega)|+\cdots+|f_m(\omega)|)$, $j=1,\cdots,m$, が $\bar{\Omega}$ 上で連続になる

ような Ω のコンパクト化 $\overline{\Omega}$ をとれば，$\overline{\Omega}$ における上界と Ω における上界とは等しくなり，上の方法を用いることができる．たとえば，$\Omega=R^k$ で，$f_j(\omega)$ が k 変数多項式のときには，Ω に，原点を端とする各半直線の無限遠点を付加して球と同位相にしたものを $\overline{\Omega}$ とすればよい．

③の問題．完全にその方式を与えたのは，$\Omega=(-\infty,\infty)$ 又は $[0,\infty)$，又は $[0,1]$ で，$\boldsymbol{f}$ が Ω 上のチェビシェフ系のときには，$m+1$ 個の未知数をもつ $m+1$ 個の連立方程式をとけばよい．多変数で多項式のときも大体同様にできる．

ローレンツ・カーブの統計学方法論的諸特性について

田口時夫

昨年度の抽象的規定から，より具体的に，経済統計分析の場に於てその特質を明にすることに努めた．その結果は科学研究費により，「ローレンツ・カーブの基礎，発展及び国富調査結果えのその適用」1662 年 3 月，に一応集括した．

又，特に数学的側面を捉えて本誌上に発表を試みた．

昭和 36 年度第 3 研究部の研究概要

青山博次郎

本年度の研究概要についてその一部分は下述されているが，第 1 研究室では企業における decision making の問題を主として研究し，石油タンクの合理的容量決定についてモンテカルロ法による計算を実施し，また木材資源利用合理化推進本部に協力して，木材の需要予測を行った．

第 2 研究室では計算機に関連して，逐次寿命試験，optimal censoring などの問題，probit analysis，メディアン範囲を用いる $\bar{X}$-R 管理図などの研究を行い，モンテカルロ法に関して乱数発生の諸問題，数値計算における誤差の問題などについて研究を行った．電子計算機の購入がきまったので，自動プログラムの開発についても研究を進めている．外部に対しては東京都に協力し，青少年の生活についての実態と欲求についての調査を行い，また東京都立大学と協力して消費と貯蓄調査にも参加した．

指導普及室では日常普及業務の外に，前年度から引きつづき研究成果の評価に関する統計的研究（試験研究）を行っている．

また第 2, 1 研究部と共に総合研究「統計的模型解析についての総合的研究」に参加し，在庫管理の研究，計算機の誤差チェックの研究，確率過程モデルの研究などを行った．

私個人としては，市場調査の統計的研究を継続し，本年度は記念切手を利用して（西平重喜氏のアイデア）回収率 72% という郵便調査の成績をあげた．また選挙の報道の機械化に関しては，当確判定の問題に関し，random intervals の分布を利用する研究を行った（一部は講究会及び Annals に報告）．また前述の在庫管理については，本年度は文献の収集と共に，Karlin 等によるモデルのパラメータに基く変動を吟味し，需要分布が正規型，ポアッソン型などについて，その影響について分析を行った．

この他前年度に引きつづき因子分析法による能力転移の問題についての研究協力と．武蔵野市における罐詰調査に協力した．

原油タンクの最適容量に関する統計的研究

鈴木雪夫

石油工場において，原油処理能力，タンカーの運搬量，入港間隔，原油不足または船待ちによる費用を考慮して，モンテカルロ法を用い，最適タンク容量を決定した．

数値計算における誤差の問題

駒沢勉

数値計算による誤差の原因は，いろいろ考えられるが（たとえば解析的な誤差，計算に用いる数値それ自身に初めから含まれている誤差，数値計算途中における四捨五入による誤差等…が考えられる），ここでは Test Matrix としてよく知られている，性質の悪い行列（有限ヒルベルト行列 $\|1/i+j-1\|$ と有限ロットキン行列 $\begin{cases} L_{1j}=1 & i=2,\cdots,n \\ L_{ij}=1/i+j-1 & j=1,2,\cdots n \end{cases}$）の逆行列を求めたときの精度の悪さを，行列の性質の悪さを特徴づける種々の量，

$\det(A)$ （行列式）

$N\text{-number}=n^{-1}\operatorname{norm}A\cdot\operatorname{norm}A^{-1}$

$(\operatorname{norm}A=\sum_{ij} A_{ij}^2)$

$M\text{-number}=n\max_{ij}|A_{ij}|\cdot\max_{ij}|A_{ij}^{-1}|$

$P\text{-number}=\max\lambda_i/\min\lambda_i$ （λ_i は固有根）

等，近年考慮されている condition number をつかっ

て，行列の悪さをチェックすると共に，求められた逆行列の精度を知る情報として，求められた逆行列を再に逆転したものと最初のデータの行列を比較することによって，求められた解の精度の情報を知る検算方法を考えて見た．この方法の利点は内蔵型の計算機で逆行列を求める場合，同一プログラムによって，解を求めることと解の検算ができること，更に記憶部は逆行列を求めるプログラムと行列の要素（$n\times n$）と少しの仕事番地でよいことである．

従来この種の検算法は行列 $\|A\|$ に求めた逆行列 $\|A^{-1}\|$ を乗じて単位行列にどれだけ近いかで，解の精度の情報を得ていた，この方法で検算まで計算機で行う場合逆転検算法の2倍の記憶部が必要である．実際有限ヒルベルト行列と有限ロットキン行列（2元～6元）を使って FACOM-128A での実験では逆転検算法は，行列 $\|A\|$ に逆行列 $\|A^{-1}\|$ を乗ずる方法より精度を強くおさえすぎる心配はなかった．

乱数の生成と変換について

渋　谷　政　昭

累積分布関数 $F(x)$ に従う乱数を電子計算機で生成するには，一様乱数 U に，$F^{-1}(U)$ という変換を行えばよいが，種々の確率モデルを工夫することにより，より効率のよい生成法が得られる．ここで，効率がよいとは，1個の $F(x)$ 乱数を生成するのに要する機械時間が短いことを意味する．つまりなるべく少数の乗算加減算で，除算や開平算はなるべく行わずに生成しようというのである．ここでは G. Marsaglia (Ann. Math. Stat., **32** ('61), 899-900) の方法を拡張して指数，カイ二乗，正規乱数の新しい生成法を提案する．

指数乱数

1.　0打切りポアソン分布に従う整数値確率変数Nをつくる．

$$P(N=n)=p^{-1}e^{-\mu}\mu^n/n!,\ p=1-e^{-\mu},\ n=1,2,\cdots.$$

2.　N個の $(0,1)$ 一様乱数をつくり，その最小値を選ぶ．

3.　成功率 $p=1-e^{-\mu}$ の幾何分布に分布に従う整数値確率変数 M をつくる．

$$P(M=m)=pq^m,\ q=1-p,\ m=0,1,2,\cdots.$$

4.　$$Y=\mu\{M+\min(U_1,\cdots,U_n)\}$$

は標準指数分布に従う．

ただし μ としては 1/2 か $\log 2$ が望ましい．

カイ二乗乱数

自由度 2ν のカイ二乗乱数を生成するには

1.　平均値 μ のポアソン分布に従う整数値確率変数の系列 $\{M_i\}$ をつくる．

2.　$$\sum_{i=1}^{L-1}M_i<\nu\leqq\sum_{i=1}^{L}M_i$$

なる事象が生じたら系列を止める．

$$N=\nu-\sum_{i=1}^{L}M_i \qquad (1\leqq N\leqq M_L)$$

とおく．

3.　M_L 個の一様乱数をつくり

$Y=2\mu\cdot\{L-1+\text{the Nth smallest of}(U,\cdots,U_{M_L})\}$

をつくれば Y が求める乱数．

$\nu=2,5,10,15$ に対し，$\mu=1.5,2.5,3,4$ ととる．

さらに一般のカイ二乗乱数，正規乱数も同様の原理で作成できる．

渋谷：各種の分布をもつ乱数の生成，数理科学総合研究第4班，第3回プログラミング・シンポジウム報告集，1962 年1月．

渋谷：乱数と生成と変換，日本数学会 1962 年春季総合分科会，応用数学会科会特別講演．

Censored sample による推定法について

多　賀　保　志

ここでは次の2つの問題を扱うことにする．

a)　censored sample にもとづいて，パラメータの推定量をどうきめたらよいか？

b)　その推定量を使って推定を行う場合に，optimal censoring をどのようにして定めるか？

以上2つの問題を扱う方針についての概略を述べる（くわしくは近々のうちに Ann. Inst. Stat. Math. に発表の予定である）．

まづaについては，有効推定量又は最尤推定量を使いたいところであるが，一般にそれらを求めるのがむつかしかったり，偏差を含むため小標本の場合に適用できなかったりするので，ここでは最良線型推定量を用いることにする．そのような推定量については，A.K. Gupta (Biometrika, 1952) が正規分布の平均と標準偏差の最良線型推定量を，S. S. Gupta (Technometrics, 1960) がガンマ分布の尺度パラメターの最良線型推定量を求めている（life test ではさらに Weibull 分布や対数正規分布の位置パラメターおよび尺度パラメターの推定量を求めたい）．一般的にいって，censored sample による最良線型推定量を求めるには順序統計量の1次および2次のモーメントがわかればよいから，この問題は基準分布（位置パラメターを 0,

尺度パラメターを1とした分布）からとられた順序統計量の1次および2次のモーメントを求めることに帰着する．そこで，それらのモーメントを正確に求めれば一番よいが，それが困難なときには，それらの近似値を使うことも考えられる．そのような近似値を求める方法については，F. N. David と N. L. Johnson (Biometrika, 1954) が述べているが，われわれが $n=20$, $k=3$ の場合について計算したところ，正規分布の場合は近似があまりよくなかった（指数分布ではかなりよかった）．

一般に上の近似式を用いると，n が大きいとき，最小値（およびそれに近いもの）のモーメントの近似がわるくなるから，それを避けるには censored sample の中から適当にえらばれたいくつかの観測値を用いた方が，近似の点からみても，計算の簡易さの点からみてもよいと思われる．すなわち，censored sample における Optimal spacing の問題も考えておく必要が生じる．

さて，b の optimal censoring については，問題に応じて適当な loss function を定めれば，あとは容易に解決される．一般的にいって，loss function L は k, n, t（推定量），θ（パラメター）の函数と考えられるが，それは推定の誤差にもとづく loss L_1 とサンプリングに要する費用 L_2 の和で表わされるとする．したがってそれに対応する risk についても，$R=R_1+R_2$ となるが，一応 n と θ を固定しておくと，R_1 は k の単調減少函数，R_2 は k の単調増加函数となり，$1\leqq k\leqq n$ なる範囲の中で，R を最小ならしめる k が一意的に定まる可能性がある（とくに R が k について下に凸ならばよい）．そのようにして定まる $k=k_0$ の値によって，optimal censoring が定まるのである．問題は，上に述べたやや異質的な2つの loss L_1 と L_2 の和で全体の loss を定義するところにあり，その両者を如何に調和させるかという点が大切であろう．

研究費と研究成果に関する統計的研究

内　田　良　男

科学振興の諸策に関連して，研究費がどのような効果をもたらすかは重要な問題である．

このような研究の基礎として，文部省科学研究費による総合研究と，科学試験研究費補助金による試験研究とについて資料分析を行なうと共に，理論的研究も行った．

資料分析の主なるものは次の通りである．

1．研究費の要求・査定・実績に関するある実態
（1） 要求額と査定額の関係［昭 34．総合，試験］
（2） 科目別経費の査定と実績の間における変化［同上］
（3） 研究計画における研究費と補助金の関係［昭 34．試験］
2．研究費の内訳に関するある実態
（1） 研究費の内訳（科目）による分布（分野別）［昭 34．総合，試験］
（2） 研究費とその内訳（科目）別経費との関係（分野別）［同上］
3．研究成果の発表に関するある実態
（1） 分野別の発表数（年度による追跡）［昭 32—35．総合，昭 32—34．試験］
（2） 論文数（発表数）と研究費との関係（分野別・年度別）［同上］
（3） 論文数と消耗品費との関係［昭 32—34．総合，試験］

理論的研究としては，研究費，研究成果および研究価値に関し模型解析を行った．主なる点は次の通りである．

1．研究費と研究成果の基本的関係
2．研究成果の経済価値と研究費との関係
3．研究費の配分法と研究費総額の決定法
（1） 研究条件が同等である多数の研究課題がある場合
（2） 研究条件が異る多数の研究課題がある場合
（3） 研究条件によって多数の研究課題を分類した場合
4．研究費の分割投入法
（1） 単純型模型へ繰込み投入する場合の準備
（2） 単純型模型へ繰込み投入する場合の分割投入法

第2研究部の研究概要・その他

林　　知己夫

各研究部との協同研究として行われている特別研究課題があるが，本年度においては，国民性に関する統計的研究，統計的モデル解析についての研究（36 年度文部省科学研究費，総合研究）が行われた．

1．国民性に関する統計的研究

1958 年に全国調査を実施し，これをまとめると共に 1959 年 12 月に岐阜市において吟味調査を実施した．従来の諸調査をまとめあげ，その報告書として「日本人の国民性」（至誠堂 1961）を発刊した．この

研究は，調査委員会のみならず，旧第3部及び現在の第2，第3部の研究員の十分な協同によってなしとげられたものである．

2．統計的モデル解析についての研究

この研究は，統計的モデルを構成することによって，複雑な現象を有効に解析する方法を確立することを目的としている．各々異った方面での研究成果を示すことによって，新しい方法論が各分野に発展し適切な手段を提示することを期待している．この様な主旨で本来の意味での総合の実をあげ，新しい研究の芽を見出すことを目的としている．この研究のねらいは，(イ)考えている方法論が有用であらねばならない事（よく数理科学と言われる研究でみられる様な形式的な理論の展開は考えない），(ロ)適用分野は異るが，統計的モデル解析という方法論を作ることにおいて筋の通っていること（具体的方法は勿論異る），(ハ)各分野での知識を交換することによって刺激を受け，新しい方法論を発展できること，である．この内容は

I　過程事象の確率モデルの研究

(イ)　雑踏に関する確率モデルの研究(植松，林，石田)

(ロ)　生物現象における確率モデルの研究（崎野，高橋；黒川一予研一）

(ハ)　在庫管理におけるモデル解析（青山，鈴木）

(ニ)　経済モデルの統計的解析(石田；江口一日銀一)

(ホ)　確率過程モデルの研究（赤池）

II　質的構造の数量化モデルに関する研究

(イ)　味覚構造の数量的研究（塩谷；野田，武藤，布川一醸造研一）

(ロ)　音楽における統計的モデル解析（林，石田；門馬一ラジオ東京一）

(ハ)　美的判断に関する数量化の研究（石田，林；野村一武蔵美大一）

III　不連続構造の統計的モデルに関する研究

(イ)　粒体系の統計的模型解析法の研究（樋口）

(ロ)　粉体系の量子統計力学的模型(横田一電機大一)

(ハ)　蛋白質分子の会合状態分布の研究（野田一東大一）

IV　計算機の誤差チェックにおける統計的模型解析（多賀）

となっている．

このほか，第2研究部第4研究室としては，従来の交通現象の研究をつづけ，日比谷交叉点において，調査を行い，都電の停留所の位置による交通量の変化状況を分析した．

第1研究室では，従来のEF（マスコミ効果に関する研究），SSM（社会的成層と移動の研究），ホワイトカラーに関する研究，中間階級に関する研究，企業体における社会的態度調査研究，社会調査法の研究，が行われ，第2研究室では粒子統計に関する研究，第3研究室では，多次元解析に関する研究，味覚に関する統計的分析，検定論に関する研究が行われた．第4研究室の研究はのちにのべる通りである．これらの細目については，夫々の発表を参照されたい．

我々のところでは，数量化に関する理論的実証的研究，数量化による分類の問題が取扱われた．またサムプリングの問題としては，継続パネルによる非妥当性の問題，諸リスト（同一のものが多くのリストに記されている）のつきあわせによる総数推定の問題が考えられ，新しい解析方法としては，確率回答・回答誤差の問題が研究された．またモデル解析としては，音楽の問題が処理された．このほか，事故現象の解析（運転事故の分析，国鉄新幹線事故の推定）一国鉄との協同一，政治意識の分析一池内氏，内川氏，（新聞研）田中氏(東大文)，京極氏(東大教養)，牧田，加留部，斎藤のろ氏（輿論議会協会)一，数量化の応用としてのマスコミ現象の分析一朝日新聞広告部或はNHK放文研との協同一，農地解放に関する調査一内閣審議室一，国富調査に関する考察一企画庁経済研究所一，盲聾唖調査一文部省特殊教育課一，木材の需要予測一木材資源利用推進本部一，が行われた．

層別について

鈴　木　達　三

東京23区内の有権者を対象にする標本調査は良く行われているが，サンプリングとして通常は投票区を第1次抽出単位，個人を第2次抽出単位にする層別2段サンプリングを用いている．

この際のサンプリングの誤差や層別の効果の推定あるいは第1次抽出単位と第2次抽出単位との抽出比の問題につづいて調査の結果，得られたデータから数値的に検討してみた．

通常は地点に割当てられるサンプル数は10～20くらいでかなり少いので得られた推定値は不安定なおそれがあるが，今回行ったものは，各地点当りの割当数が75であるので，割合安定した数値がえられていると思われる．

計算に用いた資料は：

研究所のマスコミの効果の第15次調査

教育大学社会学科の社会階層調査

文部省科学研究費による社会階層移動調査

である．各調査とも同一の層別抽出をおこない，サンプルを割りふったものである．

項目（男のみ）	推定値				
	$\hat{P}$ 比率	内分散 σ_W^2	外分散 σ_b^2	σ_W^2/σ_b^2	層別の効果（%）
職　業（工員，職人労働者の比率）	0.294	0.18676	0.00986	18.9	20.2
学　歴（大学卒の比率）	0.334	0.19880	0.00700	28.4	28.9
趣　味（ありの比率）	0.335	0.21619	0.00351	61.5	9.1
家の構え（普通住宅の比率）	0.487	0.22274	0.01883	11.8	11.5
帰属階層（中の上，上の比率）	0.119	0.09949	0.00368	27.1	7.9
年　令（20～24 才の比率）	0.203	0.15564	0.00504	30.9	4.9

三者に共通な質問項目について計算した結果は上の表の通りである．

但し内分数の推定式は $S_W^2=\sum n_jS_j^2/n$，外分散の推定式は $S_{bi}^2=\sum n_{ij}(\bar{x}_{ij}-\bar{x}_i)^2/n_i$，$S_b^2=\sum n_iS_{bi}^2/n$，層別の効果の欄は，サンプル数 1000 を，50 層に比例割当した場合の例である．この場合母集団は 50 の層にほぼ等分されるものとした．

以上の計算例からみると，通常おこなわれているサンプリングの計画（50 層別，各層より 1 地点抽出，各地点に 20 サンプル割当）はいろいろな点から考えてほぼ妥当なものであることが分る．

世論の変化

西　平　重　喜

昭和 36 年度には，第 1 に昭和 28 年以来の国民性の研究を「日本人の国民性」（至誠堂）として出版することができた．これは所長監修のもとに 12 人が執筆したが，第 2 研究部第 1 研究室はその完成に全力をあげた．

第 2 に研究所有志からなる委員会は，自由社からの研究費の援助をうけて．中都市および農村における調査を実施した．これは，前年度の東京における調査とともに，1955 年以来の社会的成層の研究に関連している．問題の重点を，中間階級においたので，「中間階級調査」と称している（自由 36 年 12 月号）.

第 3 には，やはり社会的成層と社会的移動に関して文部省科学試験研究費を交付をうけた．これも前年度から引きつづいたが，東京・金沢でパネル調査を実施し，一部は上掲の雑誌に発表した．また，意見の変化が，30 才台に起るようであるので，とくに 30 才台だけのサンプルを東京でとり，集中的に調査を実施した．この結果は 37 年末か 38 年始め頃発表の予定である．なお，この研究費に関しては．研究所では鈴木達三・多賀保志両氏の援助をうけたほか，協同研究者は分担したテーマについて発表している．

第 4 には四国電力ＫＫおよび東京電力ＫＫにおいて従業員の世論調査を実施した．これは東大尾高邦雄教授と共同である．この結果は，組織労働者の意見を知る上で，必要であった．

第 5 には，経済企画庁の人的能力部会，ユネスコの東アジア文化研究センターにおいて，社会的成層や移動の研究を役立たせている．なお，足かけ 10 年におよぶマス・コミの効果の研究は，本年度から鈴木達三が主に担当しているが，これにも関係した．また，放送文化研究所（NHK）の国民生活時間調査に委員として参加した．

さて，世論の変化をみるためには，傾向分析（trend analysis），同年令集団分析（cohort analysis），パネル研究（panel study）があるが，支持政党，憲法改正，安保改正などについて，それらの方法を用いてみた．いずれの方法にしろ，サンプリング以外による誤差が相当に大きいことが予想され，世論調査の精密な分析の限界を感じさせられ，データによって結論がかわるおそれがある．こうしてみると，世論調査のデータは全体としての意味はあるが．それをもとにしての多重クロスや，計算をかさねても，その結果は信頼するにたりない．この世論の変化については，いずれとりまとめる．

なお，調査実施に当っては，いくつかの新らしい方法を採用した．青山博次郎氏が発表した，郵便調査の回収率を高める方法も，第 3 にのべたときに使ったものにもとずくものである．

嘘の効用について

沼　田　　久

或る種の粒度分布関数測定法の限界

樋 口 伊 佐 夫

細長い粒子（高分子）の粒度分分の測定法としては Peterlin-Stuart によって創始せられた流動による粒子配向の揃いを偏光解析によって測り，それから計算する方法がある．コラーゲン分子の解合状態の研究がこの方法によって東大野田研究室で行われたが，解析が思うようにゆかない点があった．そこで方法自体に粒度分布測定という点で困難があるのではないかを検討した．行ったことは次の通りである．

廻転シリンダーを用いる流動複屈折装置で monodisperse の場合の配向の分布を求めるための P-S の式の解を Scheraag-Edsall-Gad に従って Legendre 多項式を用いる Fourier 級数であらわし，更にその係数を $\alpha=$(速度勾配)/(廻転拡散係数)で展開した．この展開を割合くわしく計算し，monodisperseの場合の消光角の近似式としては，S-E-G の論文にある近似式よりも二項くわしくとれる程度にした．（これは非常に労力を要する仕事であるが，あとで FACOM 128 による数値計算を行うためにはやむを得なかった）かくて Polydisperse の場合の式として，

$$\chi=\frac{\pi}{4}-\frac{1}{2}\cdot\frac{m_2}{m_1}G-\frac{1}{6}\left\{3a_1\frac{m_2}{m_1}\cdot\frac{m_3}{m_1}-3b_1\frac{m_4}{m_1}-\left(\frac{m_2}{m_1}\right)^3\right\}G^3-\frac{1}{10}\left[\left(\frac{m_2}{m_1}\right)^5+5\left\{b_2\frac{m_6}{m_1}-a_2\frac{m_2}{m_1}\cdot\frac{m_5}{m_1}+\left(a_1\frac{m_2}{m_1}\cdot\frac{m_3}{m_1}-b_1\frac{m_4}{m_1}\right)\times\left(a_1\frac{m_3}{m_1}-\left(\frac{m_2}{m_1}\right)^2\right)\right\}\right]G^5$$

を得た．ここに $a_1=\dfrac{5371}{4900}$, $a_2=\dfrac{7736}{4900}$, $b_1=\dfrac{64436}{49000}$, $b_2=\dfrac{79498}{49000}$ で m_i は重量粒度分布に於ける $\dfrac{\alpha}{6}$ の i 次の積率である．（廻転拡散係数は粒度に関係するから α は粒度に関係する．この関係は Perrin の式を用いる）．

mondisperse の場合はこの式で $\alpha=3$ の時に S-E-G の数値計算と 0.19° しかちがわない．従って，コラーゲンの大きさと実験の廻転数の範囲ではこの近似式で式の展開切捨てによる誤差はまずないものと考られる．

そこでコラーゲン解合に対応する仮想モデル分布を 32 種類つくり，この我々の近似式によって，それぞれの（廻転速度）―（消光角）の関係を計算した．その結果分布が離れていても曲線の区別がつきにくい場合もあり，この方法は分布函数に対して感度が甚だ悪いことがわかった．くわしい結果は，更に種々の分布について計算を行った後発表するつもりである．

もう一度方法論的にまとめると，この測定方法は数学的にいうと二つの既知の荷重函数による重量分布の二つの荷重平均の比を測定し，それから逆に分布を求めることになっている，逆変換は解析的にはとけない．数値的に積分方程式をとくことになるが，甚だ尨大な計算を要する．そこで逆に多くのモデル分布を想定し，その場合に得られるべき測定値関数を計算し，実験にデータにあう測定値関数から対応するモデル分布を探すという方法をとる方が誤差が少く得策である．

しかし結果的には逆が（実際上）一意でなく測定結果からは分布が弁別されにくいということが判明したわけである．

一次統計量による正規分布の特徴づけ

清 水 良 一

$x_1,\cdots,x_n$ は互いに独立に同一分布（分散 σ^2）に従うとする．

1. 統計量 T は，分散有限なすべての分布について

$$E(T)=\sigma^2 \quad \text{であるとする．}$$

もし，$E(T|\bar{x})=\sigma^2$ なら，x の分布は正規分布である．

2. ある $a_1,\cdots,a_n$ $(\neq 0)$ について

$a_1x_1+\cdots+a_nx_n$ が x_1 と同じ分布に従うなら，この分布は正規分布である．$(n\geq 2)$

多変数正規母集団に対する許容領域とこれに関連した二，三の統計量

塩 谷 実

分散・共分散行列が未知の場合に，多変数正規母集団に対する，楕円体の形をもつ，許容領域を設定する方法を述べ，この時，問題となる統計量について得た結果を発表した．一つは，自由度1の χ^2 変数の一次式の分布に関するもの，他の一つは，標本分散・共分散行列の最小特有根と最大特有根の比の分布に関するものである．

菌数測定について

高　橋　宏　一

総合モデル解析研究の一つ，細菌増殖モデルの研究の為に，基礎資料として大腸菌の培養に於ける全菌数（total count）と生菌数（Viable count）の時間的変動を追求する実験を行っている．全菌数は，Wright's method（但し，単位体積中の菌数既知の標準液と培養からとり出したものとを染めわけて混合）により，生菌数は通例の plating method によっている．

total count の sampling errors について考察する．ここでは，その formulation を幾つか見る．

（1）確率変数 N は整数値 $n=0,1,2,3,\cdots$ をとる一つの分布に従うとし，確率変数の組 $(X,\ Y)$ は $X+Y=n$ の条件下では，二項分布 $B(\cdot;n,p)$ に従うとするとき，$(X,\ Y)$ の moment generating ft. $\Psi(u,v)$ は

$$\Psi(u,v)=\sum_{n=0}^{\infty}P(n)(pu+qv)^n,$$

但し，$P(n)=P_r(N=n),\ q=1-p$

で与えられる．

$(X,\ Y)$ の相関係数 $\rho_{X,Y}$ は，$w=\dfrac{V(N)}{E(N)}$ とおくとき，

$$\rho_{X,Y}=\frac{w-1}{\sqrt{w+\dfrac{q}{p}}\sqrt{w+\dfrac{p}{q}}}$$

よって，例えば $P(n)$ が Poisson 分布のときは

$$\rho_{X,Y}=0$$

実際の測定資料により，顕微鏡下の各視野毎の標準の菌数（X に対応），テストの菌数（Y に対応）の値から相関係数を出し，一方標本からの $V(N),E(N),p$, の推定値を使い上式より $\rho_{X,Y}$ を計算したものと比較して前述の formulation の可否が判定できようが，それには $\rho_{X,Y}$ の標本分布が必要である．

（2）資料から各視野毎の $N,\ X,\ Y$ の分布を調べると，それぞれ負の二項分布に従うことがみられる．そこで，$(X,\ Y)$ の同時分布としては，

$$p(x,\ y)=\int_0^{\infty}\frac{1}{q\Gamma(p)}\cdot e^{-\frac{\lambda}{q}}\left(\frac{\lambda}{q}\right)^{p-1}\cdot\frac{e^{-\rho_1\lambda}(\rho_1\lambda)^x}{x!}\cdot\frac{e^{-\rho_2\lambda}(\rho_2\lambda)^y}{y!}d\lambda$$

$$=\frac{1}{q^p\Gamma(p)}\cdot\frac{\rho_1{}^x\rho_2{}^y}{x!y!}\cdot\frac{\Gamma(p+x+y)}{\left(\dfrac{1}{q}+\rho_1+\rho_2\right)^{p+x+y}}$$

が考えられる．すなわち，その generating ft は $\Psi(u,v)=\left\{\dfrac{1}{1+q(\rho_1+\rho_2-\rho_1u-\rho_2u)}\right\}^p$ であり，bi-variate negative-binomial を想定したことになる．

従って $X+Y=N$ も negative-binomial に従う．$p(x,\ y|n)=\dfrac{n!}{x!y!}\left(\dfrac{\rho_1}{\rho_1+\rho_2}\right)^x\left(\dfrac{\rho_2}{\rho_1+\rho_2}\right)^y$ となり，(1) の場合に含まれる．

以上の考えをデータに当てはめると，なかなかよく合ってくれる．そこで (2) をもとにして，標本誤差を取り扱ってよいものと思われる．

系統式信号機による交通制御について

植　松　俊　夫

車の流れに混乱が起らぬように交叉点で交通制御を行う事に関して，その制御方法の統計数理的研究を前から行ってきているが，これ迄取上げたのは一交叉点だけを考えた場合であった．然し実際問題に於ては，路線上に並んだ各交叉点における制御が互に影響しあう事から考えて，系統式の交通制御こそ重要である．

実際に幹線道路においては，系統式の信号機が設置されている事が多い．所がこれ等の信号機系列によって，系統式制御の効果を最大限に発揮するやり方は考えられていない．車の進行のダイアグラムを考え，その作図を工夫するという勘に頼った方法が行われている現状である．それでこの問題について取上げてみた．

系統式制御の効果を判断する一つの目やすとして，通常は through band の幅の広さが用いられる．これは各交叉点での信号の規制の下で，信号の1サイクル当りについて，車が一度も赤信号にかからないで通過できる時間幅の事で，車の流れの対立する二つの流れの各々に対して考えられる．この through band の幅が広い程制御の効率が高いと考えるのが通常の考え方で，ここでもそれを効率判断の基準にとった．流れの二つの方向の各々に対するこの through band の幅ができるだけ大きくなる様に，而も二つの方向の混み具合のちがいも考慮に入れて，各交叉点の間での信号時間のずれ，車に対し指定される時速をきめるという事をここで問題とした．

その為に，各交叉点毎にそこでの規制を最良にきめるという考え方を入れ，この考え方で上述の諸量をきめるやり方を与えた．

この結果の実際の応用に当っては，計算機によって数値計算で，信号時間のずれと指定速度をきめる事になるが，これによって，制御効率の出来るだけ高いも

のを勘に頼った作図的方法によらずに確実に求める事が可能となる.

伝染病の伝播モデルについて

崎 野 滋 樹

これまでに発展した伝染病の伝播モデルに於ては, age distribution 即ち, 発病からの経過時間の分布が考慮されていない. 時間区間 $(t, t+dt)$ で新患者が発生するか或は病人が回復するかどうかは t 時点に於ける age distribution に関係する. このような観点から, age distribution を考慮した伝染病の伝播モデルを作ることは, 伝染病の伝播の問題に於ける重要な課題である. が併し, これを解決することは容易なことではない.

この問題を解決する糸口としての詳しい議論は, 近く彙報に発表する.

第4研究室の概要

石 田 正 次
(林 代読)

交通現象における統計的解析及び人出・混雑現象の分析(もちまきと人出)—植松—, エピデミックモデルの研究, 公衆衛生における統計的分析, 心臓の状況検査に関する統計的研究—崎野—, 細菌の増殖に関する統計数理的研究—高橋・崎野—, があり, また土地価格指数に関する研究, 木材需要予測に関する研究, 森林資源調査に関する研究, 抜取検査方式の研究, 検定論の基礎に関する研究, 音楽・美術に関する統計的モデル解析の研究—石田—が行われた.

昭和36年度養成所事業報告および研究成果

菅 原 正 巳

昭和36年度も, 例年の通り, 基本科, 研究科(前期, 後期), および専攻科(教育統計および工業統計)を開講した. 参加人員は下記の通りである.

基本科	研究科前期	研究科後期		教育統計	工業統計	合計
		A	B			
53	99	149	144	80	152	677

今年度で例年と異った点は研究科後期をA, Bの2コースに分離したことで, 前者では主として社会科学方面, 後者は理工系方面に目標を置いた. また, 工業統計は「不規則振動の統計的解析」を主題として行ない, 多くの第一線の専門家の参加を得た.

つぎに菅原の研究について述べる. 今年度も流出の問題が中心であった. 今年度は, 北米サンフランシスコの近くの小河川の資料を得る機会に恵まれ, 半乾燥地帯の河川の流出を解析した. このような乾燥地帯については, 地表の土の含水量に関して, 新たな機構を考える必要があるように思われる. また, 積雪地帯の水の流出の問題を考え, 解決の緒に着いたと思われる. これは来年度の研究課題であるが, 高速の計算機の利用が望ましい. 最後に一つの簡単な注意を述べる.

時系列にある種の操作を加えた場合, また観測が定時に行なわれなかった場合に, 不等間隔の時系列が現れる. これを適当に補間して等間隔の時系列にする一つの簡便法である. いま, 不等間隔の時系列が次のように与えられたとする.

時刻 $\cdots, t_{-1}, t_0, t_1, t_2, \cdots$

函数 $\cdots, x_{-1}, x_0, x_1, x_2, \cdots$

函数の変化はゆるやかで, t_i は単位時間に近い間隔で並んでいるとする. x_i は時点 $[t_i]$ と $[t_i]+1$ の間に落ちる. ここに $[t_i]$ はガウスの記号である. $\tau_i=t_i-[t_i]$ と置き, x_i を時点 $[t_i]$, $[t_i]+1$ に $(1-\tau_i)$, τ_i の比で比例配分する. 各時点に配分された函数値の和を, 重さの和で割ったものを, その時点における函数の推定値とする. すなわち

$$\tilde{x}_i=\left(\sum x_\nu(1-\tau_\nu)+\sum' x_\nu\tau_\nu\right)\Big/\left(\sum(1-\tau_\nu)+\sum'\tau_\nu\right)$$

ここに, $\sum$ は $[\tau_\nu]=i$ である ν についての和, $\sum'$ は $[\tau_\nu]=i-1$ である ν についての和を表わす.

この方式は自動式計算機で簡単に計算できる所に長所がある. また, 変化のゆるやかな例について, $\{\tilde{x}_i\}$ の点列は, $\{x_i\}$ のグラフの上によくのって十分に実用になる.