土木工学における確率論の二，三の応用

長　浜　友　治

Some Application of the Theory of Probability in Civil Engineering

Tomoharu NAGAHAMA

Recently in various fields of civil engineering the development of application of probability theory such as inductive statistics, information theory etc. is outstanding not only as a treating method of experimental values but also as an estimation means in the modern public work systems engineering.

In this paper the author attempted to apply the probability theory at first for a mathematical treatment of measured values in surveying. Then a way of thinking based on the probability theory about the occurrence of accidents in the traffic engineering, and an investigation of the adaptability of various discreat probability distribution functions for the occurrence frequency of deaths were carried out. As a result, it was made clear that the Polya-Eggenberger's distribution gives an exceedingly high adaptability.

1 まえがき

近年土木工学の分野においても，コンクリートの品質管理（control of quality of concrete），水文統計学（hydrological statistics），交通推計学（traffic inductive statistics），最近では，交通事故の統計的解析手法の研究（study on use of statistical method in the analysis of road accidents），土木計画学（public work systems engineering）における将来現象予測の解明手段としての多変量解析法（multivariate statistical analysis method），回帰分析（multiple regression analysis），実験計画法（experimental design），情報理論（information theory）などは，すべて確率論（Theory of Probability）を基礎とする応用面の基礎工学である。

今日の統計学は，大量観察を前提とした記述統計学（descriptive statics）と異なり，標本よりデータを解析し，その母集団（population）の情報を推定する推計学（inductive statistics）へと発展したものである。従来確率論は偶然および予測の純粋理論であり，近代統計学はその応用と一般に定義されているが，確率論は既知の母集団から標本を論じ，統計学は標本より母集団を推論する。つまり，確率論の基礎の上に立って，確率論の論法とその結果より標本から推論を行なうもので，統計的推測の基礎概念は確率論にほかならない。この意味において，本論文においては，あえ

て，確率，統計と区分せず確率論の応用と命題し，測量観測値の推計学的処理，交通工学の重要研究課題である交通事故発生の確率論的考え方，死者発生頻度の離散型確率分布関数（discrete probability distribution function）への近似など確率論の二，三の応用を試みた。

2　測量観測値の数学的処理への確率論の適用

測量においてある量を同一条件でくり返し測定した場合，すべての測定値は絶えず変動して完全に一致しない。$\bar{x} \longrightarrow \mu$ とする点推定（point estimation）により最確値（most probable value）を得ているが，これは真値（true value）の近似値（approximate value）でしかない。最確値の精度の決め方，すなわち，真値の偏差が与えられた probability をもつような限界を推定する，いわゆる推計学に確率論の方法を適用できるのは，測定結果が一定な確率分布（probability distribution）をもつ確率変数（stochastic variable）であるためである。たとえば，測定値および観測によって生ずる誤差は，正規分布（normal distribution）に従う。

著者は先に，水平角観測値の精度が，器械の adjustment の不良，器械構造上の限界，観測者の肉体的，精神的疲労因子のほか，気象状態，特に風力による影響の大きいことを検定するため，Wild T-2 Theodolite により「水平角観測精度におよぼす風力の影響」の実験を行なった。その精度計算においては，3対回観測を 1 set とした各組の最確値 $\theta=\sum_{i=1}^{n}/n$ は，観測値の算術平均値（arithmetic mean）であり，標本不偏分散（unbiased sample variance）s^2 は，$s^2=\sum_{i=1}^{n}(\theta_i-\bar{\theta})^2/n\ (n-1)$ で表わされるが，これは後述の平均二乗誤差（mean square error）の平方である。各組から，$\delta\theta_{ij}=\theta_i-\theta_j$ を求め，各母集団の不偏分散 $u^2=u_s{}^2+u_m{}^2$ の変化を，観測時刻 $\tau=t_j-t_i$ との相関で解析し，風力影響範囲と精度の関係を求めた。風速 0～3 m/s の階級に対する等分散性の検定には，F- 検定を行なったが，各組の s^2 は従来の誤差論に基づく計算であり，この値は，点推定による最確値と同様，確率変数の考え方よりすれば，小数観測回数のデータであり，誤差論としては不完全である。このことは，次の誤差拡張論からも説明できる。

独立観測の諸量を x_1，x_2，……，x_n とし，平均二乗誤差を m_1，m_2，……，m_n とした場合，関数 $X=f\ (x_1,\ x_2, \cdots\cdots,\ x_n)$ の平均二乗誤差 M は

$$M^2=\left(\frac{\partial X}{\partial x_1}m_1\right)^2+\left(\frac{\partial X}{\partial x_2}m_2\right)^2+\cdots\cdots+\left(\frac{\partial X}{\partial x_n}\right)^2$$

$$X=x_1\pm x_2\pm\cdots\cdots\pm x_n \text{ のとき } M^2=m_1{}^2+m_2{}^2+\cdots\cdots+m_n{}^2$$

過去の測量士国家試験問題において，2区間にわけて，同一条件で行なった距離測量の平均二乗誤差を拡張理論より求めて，$M=\sqrt{m_1{}^2+m_2{}^2}$ が，全区間の平均二乗誤差 M' と一致せず，$M'>M$ となる理由を問うていることは，測定回数 n が小さいからであり，従来の誤差論がすべて測定回

数が無限回において適用されるものであり，その測定値の処理の不当性を裏付けるものである。ここに，小数のデータから無限集団または，これに近い有限数においてどのような性質であるか推定する確率論を基礎とした，推計学の測量観測値の処理が当然必要となるわけである。

Gauss の偶然誤差（accidental error）の確率曲線式において，測定誤差の大きさが ε と $\varepsilon+d\varepsilon$ との間にある確率は

$$p(\varepsilon)\,d\varepsilon=\frac{h}{\sqrt{\pi}}\,e^{-h^2\varepsilon^2}d\varepsilon$$

h は精度定数（precision constant）である。つまり，誤差が ε と $\varepsilon+d\varepsilon$ との間に生ずる頻度はこの式で表わされる確率で生ずる。この式は正規分布曲線であり，次のように表わされる。

$$p(x)d(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

x は測定値，μ は測定値の母平均で真値とみなされ，$x-\mu$ は測定値の誤差である。σ は測定値のバラツキを表わす母標準偏差（population standard deviation）であり，$p(x)$ は x の分布の確率密度関数（probability density function）である。

$$\mu=\int_{-\infty}^{\infty}xp(x)dx,\qquad \sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2p(x)dx$$

確率誤差（probable error）は，この誤差より大きい誤差の起こる確率と，小さい誤差の起こる確率とが等しいような誤差 γ で示され，平均二乗誤差の約67%に相当する。

$$p(x)dx=\frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{\mu-\gamma}^{\mu+\gamma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{2}$$

精度定数 h を求めることは，困難であるので，従来の誤差論では，測定値の精度を表わすのに平均二乗誤差，確率誤差が用いられている。

3　測定値の推計学的処理への事例

A，B 2人の観測者が，所定の2目標に対して，各4回ずつ水平角観測を行なって，表—1の結果を得た。この場合，従来の誤差論より，両者の最確値，平均二乗誤差および，平均二乗誤差を重みとして，この水平角の最確値を求めれば次のとおりである。

表—1　水平角観測結果

A 組		B 組	
番　号	水　平　角	番　号	水　平　角
1	54°　35′　18″	1	54°　35′　21″
2	20	2	17
3	25	3	22
4	15	4	24

観測値の最確値をAに対し，$X_{A,0}$　Bに対し $X_{B,0}$ とすれば，

$X_{A,0}=54°35'19.5''$, $X_{B,0}=54°35'21.0''$ となる。

Aの残差を δ_A, Bの残差を δ_B として，観側値をそれぞれ，l_A, l_B とすれば，

$\delta_A=l_A-X_{A,0}$, $\delta_B=l_B-X_{B,0}$ となり

表—2は，残差 δ と $[\delta^2]$ の計算表である。Aの平均二乗誤差を $m_{A,0}$ Bの平均二乗誤差を $m_{B,0}$ とすると

$$m_{A,0}=\pm\sqrt{\frac{[\delta_A{}^2]}{n(n-1)}}=\pm2.10''$$

$$m_{B,0}=\pm\sqrt{\frac{[\delta_B{}^2]}{n(n-1)}}=\pm1.47''$$

表—2　残差 δ, δ^2 の計算

A 組			B 組		
番号	δ_A	$\delta_A{}^2$	番号	δ_B	$\delta_B{}^2$
1	−1.5	2.25	1	±0.0	0.00
2	+0.5	0.25	2	−4.0	16.00
3	+5.5	30.25	3	+1.0	1.00
4	−4.5	20.25	4	+3.0	9.00
[ ]	0.″0	53.00	[ ]	0.″0	26.00

Aの $X_{A,0}$ の重みを p_A, Bの $X_{B,0}$ の重みを p_B とすれば

$$p_A:p_B=\frac{1}{m_{A,0}{}^2}:\frac{1}{m_{B,0}{}^2}=1:2.04$$

重みづきの最確値 X_0 は

$$X_0=\frac{p_A\cdot X_{A,0}+p_B\cdot X_{B,0}}{p_A+p_B}=54°35'20.5''$$

この場合，Aの観測値は $54°35'19.5''\pm2.10''$，Bの観測値は $54°35'21.0''\pm1.47''$ と表わされ，Bの観測値はAの2倍の信頼度があるとしているのであるが，再三述べたように，このような解析は，測定値はすべて Gauss の誤差関数に従うとしたものであり，このような小数回観測のデータでは，最確値，平均二乗誤差とも不正確であり，おおよその見当を示すに過ぎない。

3．1　正規母集団の平均値 μ の区間推定

点推定によって，小数のデータより得た最確値は，推定値であって真の値ではない。つまり一定の分布にしたがって変動する確率変数であることは，前に述べたところである。そこで真の値と推定値との間に，どの程度の誤差があるかを，いいかえれば，ある確率をもって真の値が含まれる区間を推定することが必要となる。母数 θ に対して標本量 $(x_1, x_2\cdots\cdots, x_n)$ から2つの区間推定量 $\underline{\theta}(x_1,\cdots\cdots,x_n)$, $\overline{\theta}(x_1,\cdots\cdots,x_n)$ を作り

$$P_r\{\underline{\theta}(x_1,\cdots\cdots,x_n)\leqq\theta\leqq\overline{\theta}(x_1,\cdots\cdots,x_n)\}=1-\alpha$$

とした場合，区間 $[\underline{\theta},\overline{\theta}]$ は θ の信頼区間（confidence interval）であり，$1-\alpha$ は信頼係数（confidence coefficient）である。

分散が未知の正規母集団の平均値 μ の信頼区間は，標本平均値を X_0，標本分散を s^2 とすれば

$$t=\frac{X_0-\mu}{s/\sqrt{n}}$$

が自由度 $n-1$ の t-分布（t-distribution）；t（$n-1$）することを利用して求める。
t が $\pm t^{\alpha}/_2$（$n-1$）の間にある確率が $1-\alpha$ であることは

$$P_r\left\{-t_{\alpha/2}(n-1)\leqq\frac{X_0-\mu}{s/\sqrt{n}}\leqq t_{\alpha/2}(n-1)\right\}=1-\alpha$$

これより

$$P_r\left\{X_0-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\leqq\mu\leqq X_0+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right\}=1-\alpha$$

この場合の信頼係数 $1-\alpha$ に対する μ の信頼区間は

$$\left(X_0-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ X_0+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right)$$

前問題について，上記理論式より真値の区間推定を行なってみると

信頼係数 $1-\alpha=0.95$ として

$X_{A,0}=54°35'19.5''$，$X_{B,0}=54°35'21.0''$

$$s=\sqrt{\frac{\sum_{i=1}^{n}f_i(X_i-X_0)^2}{n-1}}$$ より $s_A=4.20''$，$s_B=2.94''$

t-分布表から，$\phi=n-1=3$，信頼係数 0.95 に対し $t_{\alpha/2}=3.182$ を得る。

よって信頼区間は

$$\underline{\mu}_A=X_{A,0}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}=54°35'12.8'',\quad \bar{\mu}_A=54°35'26.2''$$

$$\underline{\mu}_B X_{B,0}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}=54°35'16.3'',\quad \bar{\mu}_B=54°35'25.7''$$

A：〔$54°35'12.8''$，$54°35'26.2''$〕

B：〔$54°35'16.3''$，$54°35'25.7''$〕

である。

3．2　母分散 σ^2 の区間推定

A，B両者の母分散の信頼区間を求めてみると，n 個の確率変数 x_1，……，x_n が互に独立で，標準正規分布するとき

$$y=x_1^2+\cdots\cdots+x_n^2$$

この分布の確率客度関数 $p(x)$ は

$$p(x)=\frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1}e^{-\frac{1}{2}x}}{2\Gamma\left(\frac{n}{2}\right)}\ ,\quad 0\leqq x<\infty$$

で表わされ，これが自由度 n の χ^2-分布（chi-square distribution）；$\chi^2(n)$ である。

いまの場合，平均値 μ が未知で，標本分散 s^2 に対して，統計量 $(n-1)s^2/\sigma^2$ は自由度 $n-1$ の χ^2-分布にしたがい，危険率 α に対して

$$P_r\left\{\chi_{1-(\alpha/2)}{}^2(n-1)\leqq\frac{(n-1)s^2}{\sigma^2}\leqq\chi_{\alpha/2}{}^2(n-1)\right\}=1-\alpha$$

を満足する χ^2 の値が得られる。

$$P_r\left\{\frac{(n-1)s^2}{\chi_{\alpha/2}{}^2(n-1)}\leqq\sigma^2\leqq\frac{(n-1)s^2}{\chi_{1-(\alpha/2)}{}^2(n-1)}\right\}=1-\alpha$$

信頼係数 $1-\alpha$ に対する母分散 σ^2 の信頼区間は次のように表わされる。

$$\left[\frac{(n-1)s^2}{\chi_{\alpha/2}{}^2(n-1)},\ \frac{(n-1)s^2}{\chi_{1-(\alpha/2)}{}^2(n-1)}\right]$$

いま信頼係数 $1-\alpha=0.95$ とすると，自由度3に対する $\chi_{\alpha/2}{}^2$，$\chi_{1-(\alpha/2)}{}^2$ は，χ^2-分布表より

$$\chi_{0.025}{}^2=9.35,\ \chi_{0.975}{}^2=0.216$$

標本分散 $s_A{}^2=17.97$，$s_B{}^2=8.67$
故に求める信頼区間は

$$A:\left[\frac{(n-1)s_A{}^2}{\chi_{0.025}{}^2(3)},\ \frac{(n-1)s_A{}^2}{\chi_{0.975}{}^2(3)}\right]=\left[5.67,\ 245.42\right]$$

$$B:\left[\frac{(n-1)s_B{}^2}{\chi_{0.025}{}^2(3)},\ \frac{(n-1)s_B{}^2}{\chi_{0.975}{}^2(3)}\right]=\left[2.78,\ 120.42\right]$$

3.3　母分散の比の検定

A，B両者の平均二乗誤差，$m_{A,0}=\pm2.10''$，$m_{B,0}=\pm1.47''$ で $m_{A,0}>m_{B,0}$ の結果，Bの技術が勝れていると判断したのであるが，果して技術的に相違があるだろうか。

いま互いに独立な確率変数 x，y が，自由度 m と n の χ^2-分布，$\chi^2(m)$，$\chi^2(n)$ に従うとき，

x/m と y/n の比で与えられる確率変数

$$F=\frac{x/m}{y/n}$$

を考えると，この分布の確率密度関数 $p(x)$ は

$$p(x)=\frac{\Gamma\left(\frac{m+n}{2}\right)m^{\frac{m}{2}}\ n^{\frac{n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\ \frac{x^{\frac{m}{2}-1}}{\left(mx+n\right)^{\frac{m+n}{2}}}\quad (0<x<\infty)$$

$\sigma_A/\sigma_B=1$ に対して統計量

$$F=\frac{s_A{}^2/\sigma_A{}^2}{s_B{}^2/\sigma_B{}^2}=\frac{s_A{}^2}{s_B{}^2}$$

が自由度 n_1-1，n_2-1 の F-分布；F（n_1-1，n_2-1）することから検定を行なうと，この場合は片側検定となり

$F=\frac{s_A{}^2}{s_B{}^2}$ ただし $s_A{}^2>s_B{}^2$，$F>1$ として検定する。

F の値を計算すると

$$F=\frac{s_A{}^2}{s_B{}^2}=2.04$$

危険率 $\alpha=0.01$ のとき，$F_0=29.5$；$\alpha=0.05$ のとき，$F_0=9.28$

$$F<F_0\ (3,\ 3\ ;\ 0.05)$$

母分散が等しいという仮説は，5％の危険率をもっても棄却できない。すなわちA，B両者の間に技術的優劣の差はないと判定される。一般に k 組の試料の母分散 σ^2 に，差がないと判定された場合の母分散の信頼区間は，次のように計算される。

$$s^2=\sum_{i=1}^{k} s_i{}^2,\qquad \text{自由度}\ \phi=\sum_{i=1}^{k}(n_i-1)$$

$$\left[\ \frac{(n-1)\sum_{i=1}^{k} s_i{}^2}{\chi_{\alpha/2}{}^2\left\{\sum_{i=1}^{k}(n_i-1)\right\}},\ \frac{(n-1)\sum_{i=1}^{k} s_i{}^2}{\chi_{1-(\alpha/2)}{}^2\left\{\sum_{i=1}^{k}(n_i-1)\right\}}\ \right]$$

A，B二組の母分散が等しいと検定されたので，この場合 $k=2$ とおいて母分散の信頼区間を求めてみると，信頼係数 $1-\alpha=0.95$ として

$$\sum_{i=1}^{k} s_i^2=26.34,\ \phi=\sum_{i=1}^{k}(n_i-1)=6$$

$$\chi_{0.025}^2=14.45,\ \chi_{0.975}^2=1.237$$

よって信頼区間は

$$\left[\frac{(n-1)\sum_{i=1}^{k} s_i^2}{\chi_{0.025}^2(6)},\ \frac{(n-1)\sum_{i=1}^{k} s_i^2}{\chi_{0.975}^2(6)}\right]=[5.47,\ 63.88]$$

となる。

3.4　母平均の差の検定

測量観測値の例により，母平均の区間推定，母分散の区間推定，母分散の比の検定を行なってきたが，次の例により母平均の差の検定を行なってみる。

A，B両川産の粗骨材の比重試験を行なった結果，次の測定値を得た。A川産の粗骨材の比重が大きいといわれているが，どうか。つまり真の値に差があるといえるか。
（注：一般に骨材の比重が大きいほど品質が勝れている。）

A；2.63　2.61　2.64　2.63　2.62　2.60　2.64　2.63
B；2.62　2.60　2.64　2.61　2.62　2.61　2.60　2.62

最初に母分散の均一性の検定を行なってみるとこの場合は両側検定となり

$$X_{A,0}=2.625,\quad X_{B,0}=2.615$$

偶然にも

$$s_A^2=s_B^2=200\times10^{-6}$$
$$\therefore F=s_A^2/s_B^2=1$$

危険率 $\alpha=0.05,\ 0.01$ とし，$\alpha/2$ に対する限界値は

$$F_0\ (7.7\ ;\ 0.025)=4.99$$
$$F_0\ (7.7\ ;\ 0.005)=8.89$$
$$\therefore F<F_0\ (7.7\ ;\ 0.025)$$

有意でない。すなわちA，B両川の粗骨材の比重のバラツキの間に相違はない。

次に母分散が等しい場合における，母平均の差の検定，すなわち，両川の粗骨材比重の真の値が等しいか否かの検定を行なってみる。いま分散が未知でも等しいことが検定された場合には，帰無仮説 H_0；$\mu_A=\mu_B$ に対して

$$t=\frac{X_{A,0}-X_{B,0}}{\sqrt{\dfrac{(n_A+n_B)\{(n_A-1)s_A{}^2+(n_B-1)s_B{}^2\}}{n_An_B(n_A+n_B-2)}}}$$

が自由度（n_A+n_B-2）の t-分布；$t(n_A+n_B-2)$ に従うことより母平均の差の検定が行なわれる。この場合の仮説は H_0；$\mu_A=\mu_B$，H_1；$\mu_A\neq\mu_B$ である。

$$n_A=n_B=8,\ X_{A,0}=2.625,\ X_{B,0}=2.615$$
$$s_A{}^2=s_B{}^2=200\times10^{-6}$$

よって上式より t を計算すれば $t=1.414$ となる。

自由度 $\phi=14$，危険率 $\alpha=0.05$ に対する $t_{\alpha/2}=2.145$，$\alpha=0.01$ に対する $t_{\alpha/2}=2.977$ であり

$$t<t_{\alpha/2}\ (14\,;\,0.025)$$

よって，両川の粗骨材比重の真値に差がないという仮説 H_0 は5％の危険率でも棄却できない。

4　交通事故発生への確率論の適用

4.1　交通事故データの基礎的考察

交通事故は，道路交通環境，人，車輛などの多要因により，きわめて不確定な要素を含む偶発的現象である。一件ごとにその原因を説明しようとしても明らかでない。しかしながら，事故全体を大量観察すると，偶然性，かつ不規則な変化の内に，高い秩序に基づいた法則が生じている。いま，ある路線または，区間を走行する自動車を考えると，前述の不確定要素を含む現象の内に常に事故の起きる危険性をもっている。このような不確定性現象の表現，およびデータ解析には，確率論を基礎とした統計解析によらなければならない。1969年 O. E. C. D. の主催した "Symposium on the Use of Statistical Method in the Analysis of Road Accidents" は事故発生と，その多要因の関連を見出し，確率論に基づく統計的手法により事故発生予測を行わんとするものである。

さて，事故発生の基礎的考察を行なってみると，いま路線またはある区間を走行する自動車全体の集団を考え，一台当りの事故の危険率を p とすれば，交通量，道路交通環境を同一とした場合，ある期間にこの部分を走行した自動車の総数の内，事故を起こす割合，すなわち危険率は，道路を自動車が走るという試行において，事故の発生する確率と定義される。V を延べ交通量とし，一台ごとに独立の走行をすると仮定する。このとき発生する事故件数が j になる確率 $p(j)$ は，理論的に

$$p(j)={}_VC_jp^j\ (1-p)^{V-j}$$

交通事故においては，p はきわめて小さく，V は非常に大きい。すなわち交通事故は稀現象である。しかし V が非常に大きいので，事故件数としては相当の数となる。この場合事故件数は次の確率分布，すなわち Poisson distribution で近似できる。事故件数が x になる確率 $p(x)$ は

$$p(x)=e^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,\ 1,\ 2,\ \cdots\cdots$$

$$\lambda=pV$$

で表わされる。

4.2　交通事故による死者発生頻度の確率分布関数への適用

交通事故による死者の発生頻度は，前述の事故件数よりはるかに稀現象であり，離散型確率分布に従うと推定される。つまり日々の死者発生頻度は，不確定現象であり，明日のことは全く予測できないが，一年間を通じて大量観察した場合一定の高い秩序に従うものと予測される。以下昭和44年中に発生した北陸三県（福井，石川，富山）の死者発生頻度について，各種離散型確率分布関数の適合度の検定を行なってみる。表－3はその頻度表である。

表－3　北陸三県死者発生頻度（昭44年中）

死者（人/日）＼月別	1	2	3	4	5	6	7	8	9	10	11	12	計
0	17	11	12	7	14	7	9	5	9	9	2	10	112
1	9	12	12	16	7	11	13	8	11	6	9	7	121
2	5	3	5	6	4	8	7	8	6	6	12	8	78
3		2	1	1	3	3	1	5	3	6	5	4	34
4			1		2	1	1	2	0	4	1	2	14
5					1			3	1				5
6													
7													
8											1		1
計	31	28	31	30	31	30	31	31	30	31	30	31	365

注：福井144人（134件），石川139人（132件），富山185人（184件）

4.2.1　Binomial Distribution への適合

事象 A が x 回起こる確率 $p(x)$ は

$$p(x)=\frac{n!}{x!(n-x)!}p^xq^{n-x}={}_nC_x\,p^xq^{n-x}=\binom{n}{x}p^xq^{n-x}\quad(x=0,\ 1,\ 2,\ \cdots\cdots,\ n)$$

binomial distribution の平均値 μ と分散 σ^2 は

$$\mu=np$$

$$\sigma^2=npq=np(1-p)$$

$p(x)$ について

$$p(x+1)=\frac{(n-x)p}{(x+1)(1-p)}p(x)$$

いま，母集団の期待値 μ の推定値 $\bar{x}$ は

$$\bar{x}=\frac{\sum f_x \cdot x}{\sum f_x}=1.28$$

$n=8$ に対する確率 p の推定値は

$$p=\frac{\bar{x}}{n}=0.16$$

したがってこの場合の確率の値を計算すると

$p(0)=0.24788$

$p(1)=0.37771,\quad p(5)=0.00348$

$p(2)=0.25181,\quad p(6)=0.00033$

$p(3)=0.09593,\quad p(7)=0.00002$

$p(4)=0.02284,\quad p(8)=0.00000$

いま試料の観測値を f_x，これに対する理論値を F_x とすると，求めたデータが，仮定した母集団分布にどの程度適合しているかは χ^2 goodness-of-fit test による。すなわち

$$\chi_0{}^2=\sum_{x=1}^{n}\frac{(f_x-F_x)^2}{F_x}=\sum\frac{f_x{}^2}{F_x}-2\sum f_x+\sum F_x$$

$$\sum f_x=\sum F_x=n$$

$$\therefore \chi_0{}^2=\sum\frac{f_x{}^2}{F_x}-n$$

$\chi_0{}^2$ の値と χ^2 表で有意水準 α，自由度 $\phi=n-2$ の χ^2 の値を比較し小ならば仮定した母集団分布を否定できず大ならば仮定母集団分布しないといえる。

表－4は，死者発生頻度と binomial distribution による理論頻度の計算結果を比較したものである。あてはまりのよさを χ^2 goodness-of-fit test すれば

$$\chi_0{}^2=\sum\frac{f_x{}^2}{F_x}-n=28.269$$

$$\chi^2\ (\phi=4,\ \alpha=0.005)=14.86<\chi_0{}^2$$

0.5％有意水準でも棄却される。すなわち binomial distribution していない。

表－4　死者発生頻度と Binomial Distribution による理論頻度の比較

死者数 x(人/日)	死者頻度 f_x		理論頻度 F_x		$f_x{}^2/F_x$
0	112		90.48		138.638
1	121		137.86		106.202
2	78		91.91		66.195
3	34		35.01		33.019
4	14		8.34		23.501
5	5	6	1.27	1.40	25.714
6	0		0.12		
7	0		0.01		
8	1		0.00		
合　計	365		365.00		393.269

4.2.2 Poisson Distribution への適合

binomial distribution において試行回数 n がきわめて大きく，出現確率 p が小さく $np=\lambda$ が定数であるとき，この分布は次の式で表わされる。

$$p(x)=e^{-\lambda}\frac{\lambda^x}{x!}, \quad x=0,\ 1,\ 2,\ \cdots\cdots$$

Poisson distribution の平均値 μ, 分散 σ^2 は

$$\mu=\lambda$$
$$\sigma^2=\lambda$$

この分布は交通流解析によく用いられ，たとえば，道路のある地点を一定時間に K 台の自動車が通過する度数は λ を適当にとって

$$f(K)=Ne^{-\lambda}\frac{\lambda^k}{K!}, \quad \text{ただし } N \text{ は全度数}$$

で表わすことができる。

Poisson distribution の数値計算は次のように行なう。

$$p(x+1)=\frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}=\frac{e^{-\lambda}\lambda^x}{x!}\cdot\frac{\lambda}{x+1}=\frac{\lambda}{x+1}p(x)$$

いま北陸三県における死者数 f_x, 理論度数を F_x とすれば，

$$F(x)=Ne^{-\lambda}\frac{\lambda^x}{x!}=365e^{-\lambda}\frac{\lambda^x}{x!}$$

$$\lambda=\frac{\sum f_x\cdot x}{\sum f_x}=1.28$$

この場合の確率の値を計算すると

$$p(0)=e^{-\lambda}=0.27804$$

$$p(x+1)=\frac{\lambda}{x+1}p(x) \quad \text{より}$$

$p(1)=0.35589$, $p(5)=0.00796$
$p(2)=0.22777$, $p(6)=0.00170$
$p(3)=0.09718$, $p(7)=0.00031$
$p(4)=0.03110$, $p(8)=0.00005$

表一5は，死者発生頻度と Poisson distribution による理論頻度の計算結果を比較したものであ

る。

χ^2 goodness-of-fit test を行なえば

$$\chi_0{}^2=\Sigma\frac{f_x{}^2}{F_x}-n=4.195$$

$$\chi^2(\varphi=4,\ \alpha=0.25)$$
$$=5.29>\chi_0{}^2$$

よって25％有意水準で Poisson distribution しているといえる。

表—5　死者発生頻度と Poisson Distribution による理論頻度の比較

死者数 x(人/日)	死者頻度 f_x		理論頻度 F_x		$f_x{}^2/F_x$
0	112		101.48		123.611
1	121		129.90		112.710
2	78		83.14		73.178
3	34		35.47		32.591
4	14		11.35		17.269
5	5	6	2.91	3.66	9.836
6	0		0.62		
7	0		0.11		
8	1		0.02		
合　計	365		365.00		369.195

4．2．3　Polya-Eggenberger's Distribution への適合

この確率変数は負の binomial distribution とも定義され，病気発生などの伝播現象のみならず，独立な現象の蓄積の近似モデルともなり，次式で表わされる。

$$p(X=k)=\frac{h(h+d)(h+2d)\cdots\cdots(h+\overline{k-1}\,d)}{k!}(1+d)^{-\frac{h}{d}-k}$$

$$\begin{cases}\text{期待値}\quad E(X)=h\\ \text{分　散}\quad \sigma^2(X)=h(1+d)\end{cases}\qquad\left(\begin{array}{l}h,\ d>0\\ k=0,\ 1,\ 2,\ \cdots\cdots\end{array}\right)$$

$d=0$ のときは明らかに binomial distribution である。また $d\to0$ のとき $p(k)\to e^{-h}h^k/k!$ (Poisson distribution) となる。

$$\begin{cases}d>0\ \text{のとき正の伝播（起こる確率が増大する傾向をもつ）}\\ d<0\ \text{のとき負の伝播（起こる確率が減少する傾向をもつ）}\end{cases}$$

$$c(x)=\frac{p(x+1)}{p(x)}(x+1)=\frac{\sigma^2-h}{\sigma^2}x+\frac{h}{\sigma^2},\quad \sigma^2=h(1+d)$$

$$c(x)\doteqdot\frac{s^2-\bar{x}}{s^2}x+\frac{\bar{x}^2}{s^2}$$

$$\bar{x}=h,\quad d=\frac{s^2}{h}-1$$

$$(x+1)p(x+1)=c(x)\cdot p(x)$$

$$x=0,\ p(1)=c(0)\cdot p(0)$$

$$x=1,\ 2p(2)=c(1)\cdot p(1)$$

$$\therefore\ p(2)=c(1)\cdot p(1)/2$$

以下順次確率計算を行なえばよい。

(1) 観測日数 $N=365$日

(2) 1日の平均死者数 h (Poisson distribution の λ に相当する) は

$$h=\frac{\Sigma f_x\cdot x}{\Sigma f_x}=1.28$$

(3) 分布の標本分散 s^2 は

$$s^2=\frac{1}{N}[f_0h^2+f_1(1-h)^2+f_2(2-h)^2+f_3(3-h)^2+f_4(4-h)^2$$
$$+f_5(5-h)^2+f_6(6-h)^2+f_7(7-h)^2+f_8(8-h)^2]$$
$$=\frac{1}{365}[545.056]=1.4933$$

(4) 弱伝播度 d は

$$d=\frac{s^2}{h}-1=0.17$$

(5) $\frac{h}{d}=7.53$ ($h=1.28$, $d=0.17$)

以下この場合の確率の値を計算する。

$$p(0)=(1+d)^{-\frac{h}{d}}=0.30659$$

$p(1)=0.33542$, $p(5)=0.01229$

$p(2)=0.20785$, $p(6)=0.00373$

$p(3)=0.09593$, $p(7)=0.00105$

$p(4)=0.03669$, $p(8)=0.00028$

表—6は，死者発生頻度と Polya-Eggenberger's distribution による理論頻度の計算結果を比較したものである。

χ^2 goodness-of-fit test を行なえば

$$\chi_0{}^2=\Sigma\frac{f_x{}^2}{F_x}-n=0.271$$

$$\chi^2(\phi=3,\ \alpha=0.95)$$
$$=0.352>\chi_0{}^2$$

となり，95％の高水準で有意であ

表—6 死者発生頻度と Polya-Eggenberger's Distribution による理論頻度の比較

死者数 x(人/日)	死者頻度 f_x		理論頻度 F_x		f_x^2/F_x
0	112		111.91		112.090
1	121		122.43		119.587
2	78		75.87		80.190
3	34		35.01		33.019
4	14		13.39		14.638
5	5	6	4.49	6.33	5.687
6	0		1.36		
7	0		0.38		
8	1		0.10		
合計	365		364.94		365.211

り，先の Poisson distribution よりはるかによく一致している。

5 む す び

一般に不確定要素を含む現象を数学的に解明する場合，実験，調査結果の推計，予測に確率論が応用され，工学，自然科学，社会科学の統計的手法と確率論の結びつきは深いものがある。本論文においては，確率論を基礎とした測定値の推計学的処理，交通事故死者発生頻度の離散型確率分布関数の適合度の検定を試みた。従来測定値や誤差の処理については，観測誤差は Gauss の誤差曲線に従うとして，その精度計算を行なってきた。しかしこれは，測定回数が無限回の場合に適合するものであり，小数回観測データからの最確値，平均二乗誤差などの推定は，不正確な，およその近似値に過ぎず，全く不完全なことが確かめられた。今後における測量観測値の処理については，測定値が一定な確率分布をもつ確率変数であることから，μ，σ^2 の区間推定，母分散の比の検定，母平均の差の検定など，確率分布関数，確率密度関数による推定，検定こそ必要である。測量士国家試験においてもこのような推計学的処理を早急に取入れるべきであろう。

一方，交通事故問題は，いまや交通工学における重要な研究分野であり，その統計的手法に関する研究，特に頻度分析と回帰分析は有力な事故解析手法である。今回，その死者発生頻度について，北陸三県データの確率分布関数への近似を試みたが，従来 Poisson distribution への適合が高いとされてきたが，今度のデータに関する限り，Polya-Eggenberger's distribution への適合度がきわめて高いことが判明した。今後，死者発生頻度，路線，区間，交差点などの事故発生件数について Polya-Eggencberger's distribution への近似モデルの研究を進めたい。さらに事故発生における複合 Poisson distribution の検討も必要である。また死者発生間隔については，伝達情報量の期待値（エントロピー）による情報理論，exponential distribution への近似も試みたい。

参 考 文 献

1） 長浜友治：測量観測精度と風力に関する実験的研究(1)，土木学会第25回年次学術講演会講演集第4部，pp. 219～222，(1970).
2） 長浜友治：水平角測精度におよぼす風力の影響，昭45年度，土木学会関西支部年次学術講演会講演概要，pp. IV-18-1～IV-18-3，(1970).
3） 日本測量協会編：測量士，測量士補国家試験受験テキスト，(1967).
4） 京都大学土木会編：土木計測便覧，丸善，(1970).
5） 松本嘉司：土木解析法1，技報堂，(1971).
6） 日本交通科学協議会編：交通科学研究資料第10集，(1969).

（著者　建設工学科　昭和47年2月24日受理）