⑪ Dosage-Mortality Curve and Systematic Statistics

阪大理学部数学教室　　小 川 潤 次 郎

動物の毒薬に対する抵抗力を調べるとき，一定の強さの薬物で，それ以下でわ死なないが，それ以上でわ死ぬとき，その強さをその動物の致死量（ Lethal Dose ）と云う。 つまり抵抗力わ致死量によつて測られる訳である。 過去の研究によれば薬物の濃度 D の対数 $x = \log D$ でこの致死量を表わすとき，動物の抵抗力の分布わ正規分布で表現されることが知られているが，こゝで注意を要する点わ如何なる実験においても死んだ動物の致死量というものわ知ることが出来ないのである。

例えば一定の強さの薬物に対して 10 匹の試験動物の内 6 匹が死んだとしても，これで分ることは死んだ動物の致死量がその強さ以下であつたということ丈である。

従つて実験的に得られる曲線わ累積頻度のヒストグラム丈である。 これの母集団分布曲線を Dosage-Mortality Curve と名づけるのである。

今問題になっている Dosage-Mortality Curve を

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} \exp.-(t-m)^2/2\sigma^2 dt \qquad (1)$$

とおくとき，母数 m を半数致死量と名づけるのである。 結局致死量の研究わ母数 m, σ の研究ということになる。

このような問題わしばしば起るのであつて，増山元三郎氏の原子爆弾被

害調査（同氏著：推計学の話，朝日新聞社刊，１２６頁～１２９頁参照）の場合にもこれが問題になった。　そして増山氏が上記の書物で述べられている m，σ の推定法わ近代的な統計学の立場から見て不充分な方法であることわ同氏も知っていられる筈である。

又，原爆被害調査をやっていられる当時 —— 正確な日時わ忘れたが —— 増山氏との対話のとき，もっと一般的な問題として，普通の場合のように標本値の一つ一つの値が観測されるのでなく，或値までの累積頻度が観測される場合の正規母集団の母数 m，σ の推定とか検定とかいう問題わ如何にすべきかというような話が出たと記憶している。この論文わ以上のような問題に対する一つの回答を述べようと思う。　以下述べる理論わ数学的厳密さをもって云えば，大標本の場合しか適用されない。

従って，増山氏の原爆調査のように８００人も調べるなら丁度よい適用例である。

しかし，少数例の場合にも或程度良い近似を与えるのであることも後で注意する。

先づ，増山氏の用いられた方法の批判から初めよう。

総頻度を N として観測の行われた強度を $x_1 < x_2 < \cdots\cdots < x_{k+1}$ とする。　そして，x_i なる強さ迄死ぬ数を

$$n_i = f_1 + f_2 + \cdots\cdots + f_i\,, \qquad i = 1, 2, \cdots\cdots, k+1$$

勿論

$$f_1 + f_2 + \cdots\cdots + f_{k+1} = N$$

従って x_i 迄の累積相対頻度わ

$$p_i = \frac{f_1 + f_2 + \cdots\cdots + f_i}{N} = \frac{n_i}{N} \tag{2}$$

である。　このとき増山氏わ（１）で

$$u = \frac{x - m}{\sigma} \tag{3}$$

なる変換をすると

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} \exp. -\frac{u^2}{2}\, du \qquad (4)$$

となることを利用して，

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_i-m}{\sigma}} \exp. -\frac{u^2}{2}\, du \qquad (5)$$

とおき， p_i が與えられたとき確率積分表から

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_i} \exp. -\frac{u^2}{2}\, du \qquad (6)$$

となる u_i を求めると（5）から

$$\frac{x_i - m}{\sigma} = u_i \qquad (7)$$

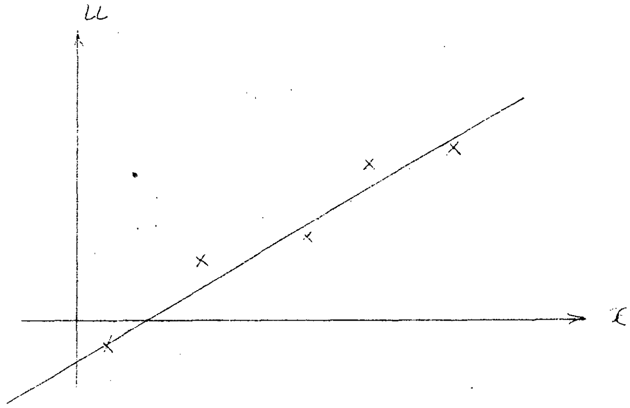
となるべきであるから，このような $k+1$ コの点 $(x_1, u_1), (x_2, u_2),$ ------ ，(x_{k+1}, u_{k+1}) を (x, u) 平面上にプロットすれば，一直線上にある筈である。

だから，この直線の勾配と切片で m，σ が判るとした。

しかし，これで（5）わ明らかに標本値と母集団値とを混同したものであって，増山氏の所謂“集計学的”なやり方である。

このようなやり方が如何に旧くさく悪いかということわ増山氏御自身で大声叱呼されているのであるから私わこゝでわ述べない。

実事わ点 (x_i, u_i) が完全に一直線上にのることわないし，特に



両端の附近でわしばしば相当にはづれる。

この理由わ後で判明するであろう。　しからばこのとき，如何なる直線を引くかゞ問題になる。

このような問題に対してわ，F. Mosteller[1]の云うSystematic Statistics を用いると至極明快な解決が得られ，上の直線を引く問題わ Systematic Statistics による The Best Linear Unbiased Estimate[2]の問題として簡単に解かれることになる。

即ち正規母集団 $N(m, \sigma)$ から抽出された大さ N の任意標本を順序づけ最小のものを1番とし，最大なものを N 番とするときこれを次のようにかく

$$x(1) < x(2) < x(3) < \cdots\cdots < x(N)$$

このとき k コの数 $n_1, n_2, \cdots\cdots, n_k$ をとり

$$\lim_{N\to\infty} \frac{n_i}{N} = p_i$$

とすれば $x(n_1), x(n_2), \cdots\cdots, x(n_k)$ の同時極限分布の密度

が次のように與えられることを利用す。

$$(2\pi)^{-\frac{k}{2}}\ \frac{\sigma^2}{n}\times\prod_{i=1}^{k+1}(p_i-p_{i-1})\times\exp\left[-\frac{n}{2\sigma^2}\left\{\sum_{i=1}^{k}\frac{p_{i+1}-p_{i-1}}{(p_{i+1}-p_i)(p_i-p_{i-1})}f_i^2\cdot\left(x(n_i)-m-u_i\sigma\right)^2\right.\right.$$

$$\left.\left.-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{p_i-p_{i-1}}\left(x(n_i)-m-u_i\sigma\right)\left(x(n_{i-1})-m-u_{i-1}\sigma\right)\right\}\right] \tag{8}$$

但し，こゝで

$$p_0=0, \qquad p_{k+1}=1$$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

$$p_i=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u_i}e^{-\frac{t^2}{2}}dt \qquad i=1,2,----,k$$

$$f_i=f(u_i) \qquad i=1,2,----k$$

である.

上の致死量の問題でわ，

$$x_1=x(n_1),\quad x_2=x(n_2),\ \cdots,\ x_k=x(n_k)$$

と考えればよい。従つて (x, u) 平面上にプロットされた点に直線を当嵌める問題わ（8）から，m, σ の J. Neyman の意味[(3)]における最良線形不偏推定値を求めることになる。

それわ従つて拡張された"最小自乗法"に関する Markoff の定

理"[4]によつて求められる。

$$K_1 = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})^2}{p_i - p_{i-1}}, \qquad K_2 = \sum_{i=1}^{k+1} \frac{(f_i u_i - f_{i-1} u_{i-1})^2}{p_i - p_{i-1}}$$

$$K_3 = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})(f_i u_i - f_{i-1} u_{i-1})}{p_i - p_{i-1}} \qquad (9)$$

$$\Delta = K_1 K_2 - K_3^2$$

とおけば，

$$\hat{m} = \frac{1}{\Delta} \left\{ K_2 \sum_{i=1}^{k+1} \frac{f_i - f_{i-1}}{p_i - p_{i-1}} (f_i x(n_i) - f_{i-1} x(n_{i-1})) - K_3 \sum_{i=1}^{k+1} \frac{f_i u_i - f_{i-1} u_{i-1}}{p_i - p_{i-1}} (f_i x(n_i) - f_{i-1} x(n_{i-1})) \right\} \qquad (10)$$

$$\hat{\sigma} = \frac{1}{\Delta} \left\{ -K_3 \sum_{i=1}^{k+1} \frac{f_i - f_{i-1}}{p_i - p_{i-1}} (f_i x(n_i) - f_{i-1} x(n_{i-1})) + K_1 \sum_{i=1}^{k+1} \frac{f_i u_i - f_{i-1} u_{i-1}}{p_i - p_{i-1}} (f_i x(n_i) - f_{i-1} x(n_{i-1})) \right\}$$

であつて，これが Asymptotically に Most efficient であることも云える。[5]

$k=2$ ときわ

$$\hat{m} = \frac{1}{u_2 - u_1} \left\{ u_2 x(n_1) - u_1 \cdot x(n_2) \right\}$$

$$\hat{\sigma} = \frac{1}{u_2 - u_1} \left\{ x(n_2) - x(n_1) \right\}$$

となる。[6]

これで最小自乗法の問題わ片附いたのであるが，m，σ の信頼区域についてわ

$$\frac{n}{\sigma^2}\left\{\sum_{i=1}^{k}\frac{p_{i+1}-p_{i-1}}{(p_{i+1}-p_i)(p_i-p_{i-1})}f_i^2\cdot\left(x(n_i)-m-u_i\sigma\right)^2\right.$$

$$\left.-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{p_i-p_{i-1}}\left(x(n_i)-m-u_i\sigma\right)\left(x(n_{i-1})-m-u_i\sigma\right)\right\} \qquad (12)$$

が自由度 k の χ^2 分布に従うことを利用すればよい。

又，m, σ についての検定は所謂 Linear hypothesis[7] であるから，例えば σ が unspecified のとき

$$H_0 : m = m_0$$

を検定するには

$$F_{k-1}=\frac{n}{\sigma^2}\left\{\sum_{i=1}^{k}\frac{p_{i+1}-p_{i-1}}{(p_{i+1}-p_i)(p_i-p_{i-1})}f_i^2\cdot\left(x(n_i)-m_0-u_i\hat{\sigma}\right)^2\right.$$

$$\left.-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{p_i-p_{i-1}}\left(x(n_i)-m_0-u_i\hat{\sigma}\right)\left(x(n_{i-1})-m_0-u_i\hat{\sigma}\right)\right\}$$

と

$$F_{k-2}=\frac{n}{\sigma^2}\left\{\sum_{i=1}^{k}\frac{p_{i+1}-p_{i-1}}{(p_{i+1}-p_i)(p_i-p_{i-1})}f_i^2\left(x(n_i)-\hat{m}-u_i\hat{\sigma}\right)^2\right.$$

$$\left.-2\sum_{i=1}^{k}\frac{f_i f_{i-1}}{p_i-p_{i-1}}\left(x(n_i)-\hat{m}-u_i\hat{\sigma}\right)\left(x(n_{i-1})-\hat{m}-u_i\hat{\sigma}\right)\right\}$$

が夫々自由度 $k-1$, $k-2$, の χ^2 分布に従い，且つ

$$F_{k-1}-F_{k-2}=F_1$$

わ自由度1の χ^2 分布に従って，F_{k-2} と独立であることから

$$F'_{k-2} = \frac{F_1}{F_{k-2}} \tag{13}$$

が自由度（1, $k-2$）の F 分布に従うことを利用すればよい。[8]

次に F. Garwood[9] の論文について批判を述べる。

今こゝで次のような k 組の実験を考えよう。

番号	$x = \log D$	実験動物の数	死亡数	生残数	生残率
1	x_1	n_1	$n_1 - S_1$	S_1	$q_1 = S_1/n_1$
2	x_2	n_2	$n_2 - S_2$	S_2	$q_2 = S_2/n_2$
3	x_3	n_3	$n_3 - S_3$	S_3	$q_2 = S_3/n_3$
⋮	⋮	⋮	⋮	⋮	⋮
i	x_i	n_i	$n_i - S_i$	S_i	$q_i = S_i/n_i$
⋮	⋮	⋮	⋮	⋮	⋮
k	x_k	n_k	$n_k - S_k$	S_k	$q_k = S_k/n_k$

各実験わ独立になされたものとする。

上に述べたことから，n_i が充分大であれば

$$x_i = x(n_i - S_i) \qquad i = 1, 2, \cdots, k$$

と考えられ，x_i の強さにおける母集団残率を Q_i 従って母集団死亡率を $P_i = 1 - Q_i$ とすると，S_i の確率わ

$$\frac{n_i!}{S_i!\,(n_i - S_i)!} P_i^{\,n_i - S_i} Q_i^{\,S_i}$$

であるから，　$q_i = \frac{S_i}{n_i}$　については

$$E(q_i) = Q_i, \quad D^2(q_i) = \frac{P_i Q_i}{n_i}$$

であるから，n_iが充分大ならば $q_i = Q_i$ と考えてよい。よって x_i の極限分布わ

$$\frac{1}{\sqrt{2\pi}\sigma} f_i \sqrt{\frac{n_i}{P_i q_i}} \exp. - \frac{f_i^2}{2\sigma^2} \frac{n_i}{P_i q_i} (x_i - m - u_i\sigma)^2 \qquad (14)$$

に従うから，x_1，x_2，-----，x_k の同時極限分布わ

$$(2\pi)^{-\frac{k}{2}} \sigma^{-k} \prod_{i=1}^{k} f_i \sqrt{\frac{n_i}{P_i q_i}} \exp. - \frac{1}{2\sigma^2} \sum_{i=1}^{k} \frac{n_i f_i^2}{P_i q_i} (x_i - m - u_i\sigma)^2 \qquad (15)$$

よって，m，σ の The Best Linear Unbiased Estimate わ

$$S = \sum_{i=1}^{k} \frac{n_i f_i^2}{P_i q_i} (x_i - m - u_i\sigma)^2 \qquad (16)$$

として，

$$\frac{\partial S}{\partial m} = 0 \quad \frac{\partial S}{\partial \sigma} = 0$$

から求められるのである。

しかしこの方法わ，n_i が小さいときわ正しくない。

それわ q_i 従って p_i が Random Variable であるからである。

しかし，

$$P_i = Q_i = \frac{1}{2}, \qquad n_i = 100$$

なら

$$D(q_i) = \frac{1}{20} = 0.05$$

であって，q_i の Coeff. of Variatum は

$$\frac{D(q_i)}{E(q_i)} = \frac{0.05}{0.5} = 0.1$$

であるから他の P_j，Q_j ではもっと小さくなる。

(1950. 12. 7)

註及び参考文献

(1) Frederick Mosteller: On Some Useful "Inefficient" Statistics:
Ann. of Math. Statist. Vol. XVII. No. 4 December, (1946) pp. 377-408

(2) F.N David and J. Neyman: Extension of The Markoff Theorem On Least Squares.
Statistical Research Memoirs, Volume II.
Department of Statistics. University of London. University College.
統数研講究録: Vol. 3, No. 9. pp. 152-163 に小川の紹介がある。

(3) J. Neyman: loc. cit. (2)
J. Neyman: On The Different Two Aspects of Representative Method:
Journ. of Roy. Statist. Soc. Vol. 97 (1934) p. 558

(4) F. N. David and J. Neyman. loc. cit.
小川 潤次郎 Markoff の定理について，統数研講究録 Vol. 5. No. 1 (1949) pp 1-5
Ogawa Junjiro: Note on the Markoff's Theorem on the Least Squares: Osaka Journal of Math. Vol. 2. No. 2 (1950)
増山 元三郎; Markoff の定理について，統数研講究録 Vol. 4 No. 11 (1949)

(5) このような Systematic Statistics について綜合的な結果わ，本論文の応用をも含めて，Osaka Journal of Math. に発表の予定

(6) これわ Mosteller の用いた Statistic である。F. Mosteller

の論文（1）参照

（7） S. Kolodziejczyk : On an Important Class of Statistical Hypotheses, Biometrika. Vol. 27 (1935) p. 161.
S. S. Wilks, Mathematical Statistics. (1943) Chapt. VIII.
小川 潤次郎： 正規回帰の理論及び其応用に就て，統数研彙究録，Vol. 3. No. 21~21 pp. 374－396. (1948)

（8） このTestのPowerについては次の論文参照
P. C. Tang : The power Function of The Analysis of Variance Test With Table And Illustrations of Their Use. Statistical Research Memoirs, Vol. II. (1938)
統数研彙究録，Vol. 5. No. 3. (1949) pp. 97~151 に小川による紹介あり。

（9） F. Garwood : The Application of Maximum Likelihood To Dosage-Mortality Curves. Biometrika, Vol. 32. Part I. (1941)

（受付 1950 12. 10.）