数量化理論と応用例 (III)*

林　知己夫

(1957 年 3 月受付)

Theory and Examples of Quantification (III)

CHIKIO HAYASHI

The present paper is a continuation from the previous papers (The proceedings of the Institute of Statistical Mathematics, Vol. 2, No. 1, 1954 and Vol. 4, No. 2, 1956).

We treat the problems based on the theory described in the paper "Approach for Quantifying Qualitative Data from the Mathematico-Statistical Point of View", the Annals of the Institute of Statistical Mathematics, Vol. II, No.1, 1950. In the case where the statistical independence of predicators (variables) was assumed, numerical computation had been done by using the actual data which was published in the book "Research on Parole Prediction" (in Japanese, 1954, Tokyo University Press).

Here, numerical computation by the same data was done in the general case. However, as a result, we found that the condition of mutual orthogonality of the predicators was approximately fulfilled and the correlation coefficients between them were small, being nearly equall in both outside groups. Therefore the obtained results were almost equivalent.

The method of quantifying qualitative items to maximize the correlation ratio (sucess rate) may imply that it will quantify them to get mutually orthogonal as possible as it can.

Furthermore, another method of quantification will be shown. In the previous papers, we adopted the method to maximize the success rate of prediciton on the min.-max. theory. Here, we consider to require the weights to maximize it in the case where the value Q (the rate of parole failure) is a priori known.

It is proved that the weights are quite the same in both methods, though the dividing point is different and the obtained success rate is different (c. f. 30-31 p.).

Institute of Statistical Mathematics.

本研究も従前の報告につづくものである. ここで論ずる問題は C. Hayashi, Approach for Quantifying Qualitative Data from the Mathematico-Statistical Point of View (Annals

---

* なおこの計算は草野美恵子嬢が行った.

of the Institute of Statistical Mathematics, Vol II, No. 1, 1950）に関する問題である．この応用例の一端は，西村克彦，林知己夫，仮釈放の研究（東大出版）に述べられてあるので，ここで繰返さないが，用いられている数値等は，該研究に於て得られているものである．

この方法は，R個の定質的要因（定量的のものであつても，これをそのまま用いず，量的なものをいくつかの分類に分ち質的なものとして取扱つた，これは統計的手法を用いるに当つて生のままの量を操作することが妥当であるか否か決定できないためである．この理由については数量化の応用例，本誌第2巻第1号 の序に詳しく述べておいた）が与えられたとき，これを用いて outside variable が二つの分類として与えられる場合（これを A，B とする）の弁別を行おうとするものである．

このとき二段の量化を考えた．まづR個の各個の要因を別々にとりあげ，A，B 両群での分散が同一，両群の平均の差（常に‘A 群の平均－B 群の平均’と考える）が最大になるように要因のカテゴリーに数量を与えた．このとき各アイテムにおける分散も同一としておく．これが第一段の量化となる．

このとき各要因を一応確率変数と見做し，X_1，$X_2 \cdots X_R$ とする．しかる後

$$\sum_{j=1}^{R}\alpha_j X_j = x,\quad \alpha_j \text{ は } \sum_{j=1}^{R}\alpha_j = 1$$

なるウエイトつきの和を考える，この x を用いて A，B 両群の弁別を考えることにした．前論文においては x が両群でガウス分布に近似されるとし，且つ X_j が互に独立なものとして，ウエイト α_j ($j=1, 2 \ldots R$) を求めるために計算を実行した．ここでは，前者の条件はそのままにし後者の制限を外し，——前者の条件のみたされることは実査によつても明らかにせられる——(X_k, X_j) が相関をもつものとして計算を実行してみた．

まづ (X_k, X_j) の相関が A，B 両群で差があるかどうかを検討するために，要因を層別してランダムに抽出し，その結果得られたいくつかの要因をとりあげ夫々の相関係数を比較してみた．

なおこの場合は第1類の要因（事前に判明している要因）のみをとりあげておくことにする．

要因間	成功群	失敗群
父母 × 学歴	－0.003	－0.067
（逮捕 × 裁判）×（罪質 × 犯数）	0.084	0.031
犯因性(16～20)×再犯素地(1～5)	－0.176	－0.266

となり相関係数は孰れも低く且つその中で比較的大きいところでも符号が一致している．

そこでほぼ両群で相関係数が等しいとみて大過のないことが確かめられたのでその仮定の下に計

	父母	妻×年令	学歴	罪質×犯数	資産×動機	逮捕×裁判	釈放後の保護者	再犯素地 1～5	犯因性 6～10	犯因性 16～20	犯因性 21～25
父母		－0.0114	0.1222	0.2124	0.1642	0.0040	0.1487	－0.1508	0.0322	－0.0652	－0.0435
妻×年令			0.2074	0.1807	0.1814	0.1600	0.4473	0.0510	0.1534	0.0446	0.2497
学歴				0.1716	0.1703	－0.0165	0.1953	0.0044	0.1702	－0.1880	－0.0116
罪質×犯数					0.2136	0.0771	0.1865	0.2005	0.0685	－0.0483	0.1062
資産×動機						0.0661	0.2385	0.1268	－0.0201	－0.0084	0.0742
逮捕×裁判							0.1348	0.0032	0.0035	－0.0092	0.1746
釈放後の保護者								0.1391	0.2108	－0.0085	0.2647
再犯素地 1～5									0.1924	－0.1497	0.2919
犯因性 6～10										0.1402	0.2398
犯因性 16～20											0.0298
犯因性 21～25											

算を実行してみることにした.

この場合も前の論文に示したと同様に相関係数は全般的に低く，我々の数量化の一つの特性と言えるのかもしれない.（これについては今後検討の要がある). 試みに相関表をかかげておくと前頁のようになる.

A 群での分布を $f_A(x)$，平均を m_A，分散を $\sigma_A{}^2$

B 群での分布を $f_B(x)$，平均を m_B，分散を $\sigma_B{}^2$

とする. なお

$$\sigma^2=\sigma_A{}^2=\sigma_B{}^2=\sum_k\sum_j\alpha_k\alpha_j\rho_{kj}$$

ρ_{kj} は (X_kX_j) の相関係数，σ_k，σ_j は夫々 X_k，X_j の分散であるが $\sigma_k{}^2=\sigma_j{}^2$ としてあるのでいまこれを1としておく，これは何等制限を与えるものではない.

分割点は mini.－max. 理論によれば

$$\int_{x_0}^{\infty}f_A(x)dx=\int_{-\infty}^{x_0}f_B(x)dx$$

で与えられるので,

$$x_0=\frac{m_A\sigma_B+m_B\sigma_A}{\sigma_A+\sigma_B}$$

したがつて判断的中率 P は両分布がガウス分布とされるとき

$$P=\frac{1}{\sqrt{2\pi}}\int_{-\frac{m_A-m_B}{\sigma_A+\sigma_B}}^{\infty}e^{-\frac{t^2}{2}}dt$$ によつて与えられる.

したがつて下端の $\dfrac{m_A-m_B}{\sigma_A+\sigma_B}$ を最大にするように α_j $(j=1,\ 2,\ \cdots\cdot R)$ をもとめればよい.

これによると

$$\sum_j\alpha_j'\rho_{kj}=l_k\qquad k=1,\ 2,\cdots\cdots R$$

但し l_k は k item における A, B 両群の平均の値，常に $l_k\geqq 0$ としておく，また α'_j は

$$\frac{\sum\sum\alpha_k\alpha_j\rho_{kj}}{\sum\alpha_kl_k}=1$$ として一応寸法を定めたものである.

を解けばよいことになる. 参考のため l_k をかかげておくと，次のようになつた. k の順番は前に相関表をかかげた順序である. 0.38, 0.78, 0.46, 0.88, 0.94, 0.38, 1.06, 1.00, 0.64, 0.66, 0.84. かうしてこう前と比較のためウエイトの和が1になるようにしてウェイト α をさだめることにする.

この結果 α として，次のものを得た. これと前の結果を比較してみると，次のようで全般的に order 的な大差のないことが示された.

	独立とした場合	相関ある場合		独立した場合	相関ある場合
父母	0.047	0.072	保護者	0.132	0.106
妻×年令	0.097	0.037	再犯素地 1～5	0.125	0.198
学歴	0.047	0.069	犯因性 6～10	0.080	0.019
犯数×罪質	0.110	0.081	〃 16～20	0.082	0.198
資産	0.117	0.110	〃 21～25	0.105	0.065
逮捕×裁判	0.057	0.043			

なおこの種の数量化を行うとき各要因間には相関が少くなるようにされているのではないかと思われる(数量化理論と応用例 II 参照). このことについては今後研究を要する.

なお σ_A，σ_B について実際の値を計算してみると

$$\sigma_A = 0.366$$

$$\sigma_B = 0.343$$

となつているのは前の分析とあわせ満足すべき一つの手掛りである．また両群で $f_A(x)$, $f_B(x)$ がガウス分布をしているか否かを検討するために χ^2 検定を行つてみたところ，前報告の如くやはりかなりよいあてはまりが認められた．

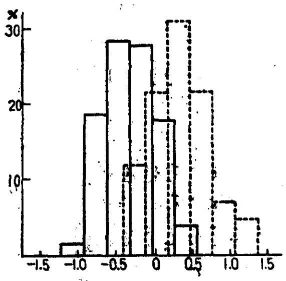
$f_A(x)$ については，

$$\chi_k^2 = 0.35.\ \ P_r\{\chi^2 \geqq \chi_k^2\} = 0.85.\ \ 自由度\ 2$$

$f_B(x)$ については

$$\chi_k^2 = 1.41.\ \ P_r\{\chi^2 \geqq \chi_k^2\} = 0.50.\ \ 自由度\ 2$$

であつた．A, B グループの分布をつくつてみると下図のようになつている．



ここでは mini.-max. 理論を用いて最もよい α_j を決定することを考えたが，成功率 Q_A が判明している場合 α_j はどう求められるかを検討してみることにしよう．結論から云えば，やはり全く mini.-max. の場合と同様にして α_j を求めてよい，ただ分割点を変えるだけでよいと云うことが示された．したがつてここで述べたもの，或は前掲書の α_j の決定方法はいづれの場合にもオプティマムなものとなつている．即ち失敗率を Q_B とする（成功率 Q_A, $Q_A+Q_B=1$）と

$$P = Q_A\int_{x_0}^{\infty} f_A(x)dx + Q_B\int_{-\infty}^{x_0} f_B(x)dx.$$

したがつて $Q_Af_A(x_0)=Q_Bf_B(x_0)$ として分点 x_0 がさだめられる．

ここで的中率を計算してみると，

	min.-max. の場合		Q_B が判明している場合(Q_B=0.35)	
	独立とした場合	相関を考えに入れた場合	独立とした場合	相関を考えに入れた場合
分　　点	− 0.034	− 0.0132	− 0.160	− 0.114
的　中　率	0.81	0.83	0.82	0.85

となり，独立を仮定したときと相関を考えに入れた場合とで大した差は見られない．

なお以下でさきに示した α_j の求め方がいずれの場合（mini.-max. のとき，また Q の判明しているとき）でもオプティマムとなつていることを示してみよう．さて分点 x_0 は $\sigma_A=\sigma_B=\sigma$ としてあるから

$$x_0 = \frac{(m_A^2-m_B^2)+2S\sigma^2}{2(m_A-m_B)}$$

但し　　$S = \log Q_B - \log Q_A$

として求められる．さてここで

P が最大になるような $\alpha_j(j=1, 2\cdots R)$ を求めるために $\partial P/\partial\alpha=0$ をつくればよい．

$$P = Q_A\frac{1}{\sqrt{2\pi}}\int_{\frac{x_0-m_A}{\sigma}}^{\infty} d\Phi(x) + Q_B\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x_0-m_B}{\sigma}} d\Phi(x)$$

但し
$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$
であるから
$$Q_Af_A(x_0)\frac{\partial\left(\dfrac{x_0-m_A}{\sigma}\right)}{\partial\alpha}=Q_Bf_B(x_0)\frac{\partial\left(\dfrac{x_0-m_B}{\sigma}\right)}{\partial\alpha}$$
を解くことになるが，$Q_Af_A(x_0)=Q_Bf_B(x_0)$ であるから
$$\frac{\partial\left(\dfrac{x_0-m_A}{\sigma}\right)}{\partial\alpha}=\frac{\partial\left(\dfrac{x_0-m_B}{\sigma}\right)}{\partial\alpha}$$
をとけばよい，

ところが
$$\frac{x_0-m_B}{\sigma}=\frac{M}{2}+\frac{S}{M}$$
$$\frac{x_0-m_A}{\sigma}=-\frac{M}{2}+\frac{S}{M}$$
但し
$$\frac{m_A-m_B}{\sigma}=M$$
であるから，上の式は
$$\frac{\partial M}{\partial\alpha}=0$$
をとけばよいことになり

Q が判明しているときでも，全く min.-max. の時と一致する.

(統計数理研究所)