TECHNICAL REPORT
34-72
業　績　報　告　書

APPROXIMATE DISTRIBUTION OF THE MAXIMUM OF C−1 χ^2-STATISTICS (2×2) DERIVED FROM 2×C CONTINGENCY TABLE

2×C分割表から生ずるC−1個のχ^2統計量（2×2）の最大値の漸近的分布

NARIAKI SUGIURA, Sc.D.　杉浦成章
MASANORI OTAKE, A.B.　大竹正徳

ABCC

ATOMIC BOMB CASUALTY COMMISSION

国立予防衛生研究所－原爆傷害調査委員会

JAPANESE NATIONAL INSTITUTE OF HEALTH OF THE MINISTRY OF HEALTH AND WELFARE

TECHNICAL REPORT SERIES
業 績 報 告 書 集

The ABCC Technical Reports provide the official bilingual statements required to meet the needs of Japanese and American staff members, consultants, advisory groups, and affiliated government and private organizations. The Technical Report Series is in no way intended to supplant regular journal publication.

ABCC 業績報告書は，ABCC の日米専門職員，顧問，諮問機関ならびに政府および民間の関係諸団体の要求に応ずるための日英両語による公式報告記録であって，業績報告書集は決して通例の誌上発表論文に代わるものではない.

TECHNICAL REPORT
34-72
業　績　報　告　書

APPROXIMATE DISTRIBUTION OF THE MAXIMUM OF C−1 χ^2-STATISTICS (2×2) DERIVED FROM 2×C CONTINGENCY TABLE

2×C分割表から生ずるC−1個のχ^2統計量(2×2)の最大値の漸近的分布

NARIAKI SUGIURA, Sc.D.　杉浦成章
MASANORI OTAKE, A.B.　大竹正徳

ABCC
ATOMIC BOMB CASUALTY COMMISSION
HIROSHIMA AND NAGASAKI, JAPAN

A Cooperative Research Agency of
U.S.A. NATIONAL ACADEMY OF SCIENCES — NATIONAL RESEARCH COUNCIL
and
JAPANESE NATIONAL INSTITUTE OF HEALTH OF THE MINISTRY OF HEALTH AND WELFARE

with Funds Provided by
U.S.A. ATOMIC ENERGY COMMISSION
U.S.A. NATIONAL CANCER INSTITUTE
U.S.A. NATIONAL HEART AND LUNG INSTITUTE
U.S.A. ENVIRONMENTAL PROTECTION AGENCY
JAPANESE NATIONAL INSTITUTE OF HEALTH

原 爆 傷 害 調 査 委 員 会
広島および長崎

米国学士院一学術会議と日本国厚生省国立予防衛生研究所
との日米共同調査研究機関

米国原子力委員会，米国癌研究所，米国心臓・肺臓研究所
米国環境保護庁および日本国厚生省国立予防衛生研究所
の研究費による

ACKNOWLEDGMENT

謝　辞

Thanks go to Dr. Toranosuke Ishimaru, ABCC for kindly providing the data in Table 3.

表3の資料を提供していただいたABCCの石丸寅之助博士に謝意を表する.

A paper based on this report was published in the following journal:

本報告に基づく論文は下記の雑誌に発表した.

Communications in Statistics 1:9-16, 1973

CONTENTS
目 次

Approved 承認 *17 August 1972*

APPROXIMATE DISTRIBUTION OF THE MAXIMUM OF C − 1 χ^2-STATISTICS (2×2) DERIVED FROM 2×C CONTINGENCY TABLE

2×C分割表から生ずるC−1個のχ^2統計量(2×2)の最大値の漸近的分布

NARIAKI SUGIURA, Sc.D. (杉浦成章)[1]; MASANORI OTAKE, A.B. (大竹正徳)[2]

Department of Mathematics, Faculty of Science, Hiroshima University, Consultant to ABCC [1] and Department of Statistics, ABCC [2]

広島大学理学部数学教室, ABCC顧問,[1] およびABCC統計部[2]

FOREWORD

Data in a medical survey often involve the presence or absence of an attribute such as a specific disease. Such an attribute is then related to one or more factors to determine its association with these factors. In this process, the data are frequently grouped or classified into various rubrics.

An important step in the classification of data is the determination of how the events or characteristics should be grouped. Otake and Jablon proposed a data reduction method which identifies the point of maximum χ^2 value where the total range of data may be divided into two parts. The procedure involves the calculation of χ^2 values for the possible 2 X 2 that can be formed by a step-by-step combination of columns. If the maximum χ^2 value is less than a predetermined value, it is presumed that these groups are homogeneous and may therefore be combined. If the maximum χ^2 value is greater than or equal to a specific value, this is taken as an indication that each group is meaningful, statistically speaking.

The computation of the exact critical point of Otake and Jablon is cumbersome to perform. Therefore, Sugiura and Otake have proposed the use of an approximate value. This approximation when applied to leukemia data gave reasonable results.

序文

医学調査資料は，しばしば特定疾患のような属性の有無と関係がある．このような属性は，一要因またはそれ以上の要因との関係が考えられるので，これらの要因と属性との関連性を明確にする．この過程での資料はしばしば種々の項目に群別されるか，または分類される．

資料分類のうえで重要な役割は，事象や特性をいかにグループするかを決定することである．大竹とJablonは，資料の全領域に対して2分割したときのχ^2統計量の最大値点を示すデータ削減法を提案した．この方法は，特性づけられた各列を漸次結合することによって作成される可能な2×2表のχ^2統計量を計算することである．χ^2統計量の最大値がある判定値より小さければ，これらの群は同じ程度の変動であるとみなし，一つの群に統合できると考える．もしχ^2統計量の最大値がある判定値よりも大きいか等しい場合，統計的にいって各群は意味のある徴候を示唆すると考える．

大竹とJablonの正確な基準点に対する計算はきわめてやっかいである．したがって，杉浦と大竹は近似的基準値の使用を提言した．この近似値を白血病資料に応用した場合，妥当な結果が得られた．

SUMMARY

In a 2 × c contingency table, let χ_1^2 be the χ^2-statistic for 2 × 2 table composed of 1st column vs the sum of 2nd, 3rd, . . . and c th column. Let χ_2^2 be the χ^2-statistic for 2 × 2 table composed of 1st plus 2nd columns vs the sum of 3rd, 4th, . . . and c th columns. Finally in this way χ_{c-1}^2 can be defined by the χ^2-statistic for 2 × 2 table composed of the sum of the first c-1 columns vs c th column. In this paper, it is shown that the asymptotic distribution of $T = \max[\chi_1^2, \ldots, \chi_{c-1}^2]$ is expressed in terms of the multivariate normal probability of c-1 dimensional cube for large sample sizes. Approximately conservative critical point of T is obtained in (2.14). Application to the procedure by Otake and Jablon, where the columns are ordered with respect to a numerical variable, for regrouping a 2 × c table proposed is stated in relation to Leukemia data at ABCC.

要　約

2×c 分割表において，第1列対第2列＋第3列＋…＋第c列から作られる2×2表のχ^2統計量をχ_1^2とする．次に同じ2×c表において，第1列＋第2列対第3列＋…＋第c列から作られる2×2表のχ^2統計量をχ_2^2とする．以下同様にして，第1列＋…＋第c－1列対第c列から作られるχ^2統計量をχ_{c-1}^2とする．この論文では，統計量$T = \max[\chi_1^2, \chi_2^2, \ldots, \chi_{c-1}^2]$の漸近分布が近似的にc－1次元正規分布の正方形領域における積分で表わされることを示し，統計量Tの近似的基準点を計算した．2×c分割表のグループ削減に関して，大竹とJablonがデータ削減の方法を提案したときに用いたABCCの白血病データに適用して述べる．ただし，列項目は変量に関して順位づけられているとする．

INTRODUCTION

In a 2 × c contingency table

緒　言

次の2×c分割表

$$
(1.1) \qquad \begin{array}{cccc|c}
X_{11} & X_{12} & \cdots & X_{1c} & n_1 \\
X_{21} & X_{22} & \cdots & X_{2c} & n_2 \\
\hline
X_{.1} & X_{.2} & \cdots & X_{.c} & n
\end{array},
$$

we shall define c-1 χ^2-statistics (2 × 2) in the summary as

において要約で述べたc－1個の2×2分割表から得られるχ^2統計量を作ると

$$
(1.2) \qquad \chi_1^2 = \frac{n^3(X_{11} - n_1 X_{.1}/n)^2}{n_1 n_2 X_{.1}(n - X_{.1})}, \qquad \chi_2^2 = \frac{n^3[X_{11} + X_{12} - n_1(X_{.1} + X_{.2})/n]^2}{n_1 n_2 (X_{.1} + X_{.2})(n - X_{.1} - X_{.2})},
$$

$$
\cdots \chi_{c-1}^2 = \frac{n^3\left(\Sigma_{j=1}^{c-1} X_{1j} - n_1 \Sigma_{j=1}^{c-1} X_{.j}/n\right)^2}{n_1 n_2 \left(\Sigma_{j=1}^{c-1} X_{.j}\right)\left(n - \Sigma_{j=1}^{c-1} X_{.j}\right)}
$$

The purpose of this paper is to give an approximate critical point for the distribution of $T = \max[\chi_1^2, \cdots, \chi_{c-1}^2]$, which was used in the first step of Otake and Jablon's procedure.[1] Since the exact distribution of T is not tractable, an asymptotic distribution is investigated for large number of observations n under the following conditions:

である．この論文の目的は，大竹とJablonの方法[1]の最初の段階で用いた統計量$T = \max[\chi_1^2, \cdots, \chi_{c-1}^2]$の分布に対する近似的基準点を求めることである．統計量Tのexact分布はn，cが小さい場合を除いて求めにくいので，われわれは

$$
(1.3) \qquad n_i/n = \rho_i > 0, \quad i = 1,2; \quad X_{.j}/n = \gamma_j > 0, \quad j = 1,2,\cdots,c.
$$

It is given in Theorem 1 of Section 2. The accuracy of the approximation is examined for numerical examples in Section 3. Some conservative critical point of T which is simpler, is obtained.

とおき，$n\to\infty$のとき，Tの漸近分布を求めることを考える．Tの漸近的分布は第2節の定理1に示している．第3節ではこの近似値の精度が数値例を用いて検討される．きわめて簡単な例題についてTのconservativeな基準点が求められる．

APPROXIMATE DISTRIBUTION OF THE MAXIMUM

最大値の漸近分布

Whether we have c common binomial populations or 2 common multinomial populations in (1.1), the conditional probability of $X_{11}, X_{12}, \cdots, X_{1c\text{-}1}$ for given marginals n_1, n_2 and $X_{\cdot j}$ $j = 1,2,\cdots,c$, is expressed by

分割表(1.1)において，c個の二項分布があると考えてもまた二つの多項分布があると考えても，母集団が同一という仮説のもとでは，周辺度数n_1, n_2, $X_{\cdot j}$ ($j=1, 2, \cdots, c$)を一定にしたときのX_{11}, X_{12}, $\cdots$, X_{1c-1}の条件付分布は

$$(2.1)\qquad P(X_{1j}, 1\leqslant j\leqslant c\text{-}1 \mid n_i, X_{\cdot\ell}) = \frac{n_1!\,n_2!}{\Pi_{j=1}^{c}(X_{1j}!\,X_{2j}!)} \Big/ \frac{n!}{\Pi_{j=1}^{c} X_{\cdot j}!}$$

The conditional moments are given by

で与えられる．次の条件付積率が求められる．

$$(2.2)\qquad \begin{aligned} &E(X_{1j} \mid n_i, X_{\cdot\ell}) = n_1 X_{\cdot j}/n, \\ &E(X_{1j}(X_{1j}-1) \mid n_i, X_{\cdot\ell}) = n_1(n_1-1) X_{\cdot j}(X_{\cdot j}-1)/n(n-1), \\ &E(X_{1i}X_{1j} \mid n_i, X_{\cdot\ell}) = n_1(n_1-1) X_{\cdot i}X_{\cdot j}/n(n\text{-}1) \qquad (i\neq j). \end{aligned}$$

Under the limiting condition (1.3), we can write the means, variances and correlations for given marginals as

極限(1.3)のもとで，一定の周辺度数に対する条件付での平均，分散，相関係数は

$$(2.3)\qquad \begin{aligned} &E(X_{1j} \mid n_i, X_{\cdot\ell}) = \rho_1\gamma_j n, \\ &V(X_{1j} \mid n_i, X_{\cdot\ell}) = \rho_1\rho_2\gamma_j(1-\gamma_j)n + 0(1), \\ &\rho(X_{1j}, X_{1k} \mid n_i, X_{\cdot\ell}) = -[\gamma_j\gamma_k/(1-\gamma_j)(1-\gamma_k)]^{1/2}. \end{aligned}$$

Normalizing the statistic X_{ij}, we shall put

として表わされる．そこで，確率変数X_{ij}を標準化して

$$(2.4)\qquad Y_j = (X_{1j} - n\rho_1\gamma_j)/[\rho_1\rho_2\gamma_j(1-\gamma_j)n]^{1/2}, \quad j=1,2,\cdots,c\text{-}1.$$

Then $Y = (Y_1, \cdots, Y_{c\text{-}1})'$ has zero mean vector and covariance matrix $\Sigma + 0(n^{-1})$, where for Kronecker's δ_{ij},

とおけば，$Y=(Y_1, \cdots, Y_{c-1})'$は平均0，共分散行列$\Sigma+0(n^{-1})$，ただしKroneckerの$\delta_{ij}$に対して

$$(2.5)\qquad \Sigma = (\delta_{ij} - (1-\delta_{ij})[\gamma_i\gamma_j/(1-\gamma_i)(1-\gamma_j)]^{1/2}).$$

Further we can prove the following:

をもつ c−1 次元の確率変数になる．Y についてさらに次の結果が成り立つ．

Lemma 1. Under the limiting condition (1.3) the statistic $Y = (Y_1, \cdots, Y_{c-1})'$ has asymptotically normal distribution $N(0, \Sigma)$ as $n \to \infty$.

Lemma 1．極限(1.3)のもとで $n \to \infty$ のときに与えられる確率変数 $Y = (Y_1, \cdots, Y_{c-1})'$ は漸近的に平均0，共分散行列 Σ の正規分布に従う．

Proof. By (2.4) and the Stirling's formula, each factorial in the exact probability (2.1) can be evaluated as

(証明)．(2.4)と Stirling の公式とにより，exact 確率(2.1)の各階乗は

$$\log n! = \log\sqrt{2\pi} + (n+\frac{1}{2}) \log n - n + 0(n^{-1}),$$

$$\log \Pi_{j=1}^{c} X_{.j}! = c \log\sqrt{2\pi} + \Sigma_{j=1}^{c} (n\gamma_j + \frac{1}{2}) \log n\gamma_j - n + 0(n^{-1}),$$

$$\log(n_1!n_2!) = 2\log\sqrt{2\pi} + (n+1)\log n - n + \Sigma_{i=1}^{2} (n\rho_i + \frac{1}{2}) \log \rho_1 + 0(n^{-1}), \tag{2.6}$$

$$\log \Pi_{j=1}^{c} (X_{1j}!\, X_{2j}!) = 2c\log\sqrt{2\pi} - n + \Sigma_{j=1}^{c} (n\gamma_j + 1)\log n\gamma_j + \Sigma_{i=1}^{2} (n\rho_i + \frac{c}{2}) \log \rho_i + \frac{1}{2} \Sigma_{j=1}^{c-1} (1-\gamma_j) Y_j^2 + \frac{1}{2\gamma_c} (\Sigma_{j=1}^{c-1} \sqrt{\gamma_j(1-\gamma_j)}\, Y_j)^2 + 0(n^{-1}),$$

which yields

として与えられ，これを使って，次の評価を得る．

$$\log P(X_{1j}, 1 \leqslant j \leqslant c\text{-}1 \mid n_i, X_{.\ell}) = -(c\text{-}1)\log\sqrt{2\pi} - \frac{1}{2}\Sigma_{j=1}^{c} \log n\gamma_j + \frac{1}{2}\log n - \frac{1}{2}(c\text{-}1)\log\rho_1\rho_2 - \frac{1}{2}\Sigma_{j=1}^{c-1}(1-\gamma_j)Y_j^2 - \frac{1}{2\gamma_c}(\Sigma_{j=1}^{c-1}\sqrt{\gamma_j(1-\gamma_j)}\, Y_j)^2 + 0(n^{-1/2}). \tag{2.7}$$

Noting that $dX_{1j} = [\rho_1\rho_2\gamma_j(1-\gamma_j)\, n]^{1/2}\, dY_j$ and putting

であることを注意して，

$$A = (\delta_{ij}(1-\gamma_i) + [\gamma_i\gamma_j(1-\gamma_i)(1-\gamma_j)]^{1/2}/\gamma_c) \tag{2.8}$$

with $|A| = (\Pi_{j=1}^{c-1}\sqrt{1-\gamma_j})/\sqrt{\gamma_c}$, the left-hand side of (2.7) can be written by

とおけば

でもって，(2.7)の左側の式は

$$(2\pi)^{-(c-1)/2}\, |A|^{1/2} \exp[-\frac{1}{2} Y'AY]\, dY + 0(n^{-1/2}). \tag{2.9}$$

It can easily be verified that $A^{-1} = \Sigma$, which completes the proof.

として得られる．$A^{-1} = \Sigma$ であることは容易にわかる．

In order to derive the limiting distribution of T, we shall put $Z_i = \pm\sqrt{\chi_i^2}$ and $\delta_i = \Sigma_{j=1}^{i} \gamma_j$ for $i = 1, 2, \cdots, c\text{-}1$. Then from (1.2), $Z_i = (\Sigma_{j=1}^{i} X_{1j} - \rho_1 n\delta_i)/[\rho_1\rho_2\delta_i(1-\delta_i)n]^{1/2}$, which can be regarded as a linear transformation from $(Y_1, \cdots, Y_{c-1})'$ namely, $(Z_1, \cdots, Z_{c-1})' = B(Y_1, \cdots, Y_{c-1})'$ where B is given by

T の極限分布を導くために $Z_i = \sqrt{\chi_i^2}$ と $\delta_i = \Sigma_{j=1}^{i} \gamma_j$，$i = 1, 2, \cdots, c-1$ とおくと，(1.2)から $Z_i = (\Sigma_{j=1}^{i} X_{1j} - \rho_1 n\delta_i)/[\rho_1\rho_2\delta_i(1-\delta_i)n]^{1/2}$，は $(Y_1, \cdots, Y_{c-1})'$ の線型変換として表わすことができる．すなわち $(Z_1, \cdots, Z_{c-1})' = B(Y_1, \cdots, Y_{c-1})'$ ただし B は次のようになる．

$$
(2.10)\quad
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
\sqrt{\dfrac{\gamma_1(1-\gamma_1)}{\delta_2(1-\delta_2)}} & \sqrt{\dfrac{\gamma_2(1-\gamma_2)}{\delta_2(1-\delta_2)}} & 0 & \cdots & 0 \\
\sqrt{\dfrac{\gamma_1(1-\gamma_1)}{\delta_3(1-\delta_3)}} & \sqrt{\dfrac{\gamma_2(1-\gamma_2)}{\delta_3(1-\delta_3)}} & \sqrt{\dfrac{\gamma_3(1-\gamma_3)}{\delta_3(1-\delta_3)}} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\sqrt{\dfrac{\gamma_1(1-\gamma_1)}{\delta_{c-1}(1-\delta_{c-1})}} & \sqrt{\dfrac{\gamma_2(1-\gamma_2)}{\delta_{c-1}(1-\delta_{c-1})}} & \sqrt{\dfrac{\gamma_3(1-\gamma_3)}{\delta_{c-1}(1-\delta_{c-1})}} & \cdots & \sqrt{\dfrac{\gamma_{c-1}(1-\gamma_{c-1})}{\delta_{c-1}(1-\delta_{c-1})}}
\end{pmatrix}
$$

By Lemma 1, we immediately obtain:

Theorem 1. Under the limiting condition (1.3), the distribution of $(Z_1, \cdots, Z_{c-1})'$ is asymptotically $N(0, \tilde{\Sigma})$, where the covariance matrix $\tilde{\Sigma} = (\tilde{\sigma}_{ij})$ is given by $\tilde{\sigma}_{ij}\ (i \leqslant j) = [\delta_i(1-\delta_j)/\delta_j(1-\delta_i)]^{1/2}$ for $\delta_i = \Sigma_{j=1}^{i}\gamma_j$.

Lemma より次の結果を得る．

定理1．$Z_j=\sqrt{\chi_j^2}$ とおき極限(1.3)のもとで $n\to\infty$ とすれば $Z=(Z_1, \cdots, Z_{c-1})'$ は平均0，共分散行列 $\tilde{\Sigma}=(\tilde{\sigma}_{ij})$ の正規分布に従う．ただし $i\leqq j$ のとき $\tilde{\sigma}_{ij}\ (i\leqq j)=[\delta_i(1-\delta_j)/\delta_j(1-\delta_i)]^{1/2}$，$\delta_i=\Sigma_{j=1}^{i}\gamma_j$ で $\tilde{\Sigma}$ は与えられる．

From Theorem 1, the limiting distribution of T is expressed by

この定理よりTの極限分布として

$$
(2.11)\quad \lim_{n\to\infty} P(T < t) = (2\pi)^{-(c-1)/2}\,|\tilde{\Sigma}|^{-1/2} \int_{-\sqrt{t}}^{\sqrt{t}} \cdots \int_{-\sqrt{t}}^{\sqrt{t}} \exp[-\frac{1}{2} Z'\tilde{\Sigma}^{-1} Z]\,dZ.
$$

Unfortunately the limiting distribution depends on marginal frequencies $X_{.j}$ and the approximate critical point of T should be computed for each given marginals. The following lemma due to Šidák[2] releases this inconvenience.

を得る．この極限分布は周辺度数 $X_{.j}$ を含んでおり，Tの近似的判定値点を求めるには便利でない．ところが幸いにも多次元正規分布については Šidák[2] による次の結果が使える．

Lemma 2. (Šidák) Let $(X_1, \cdots, X_k)$ have normal distribution $N(0, \Sigma)$. Then for any covariance matrix Σ, the following inequality holds:

Lemma 2．(Šidák) $(X_1, \cdots, X_k)$ を平均0の k 次元正規分布に従う確率変数とするとき，共分散行列のいかんにかかわらず次の不等式が成り立つ．

$$
(2.12)\quad P(|X_1| \leqslant x_1, \cdots, |X_k| \leqslant x_k) \geqslant \Pi_{j=1}^{k} P(|X_j| \leqslant x_j).
$$

The above inequality means that the probability for normal distribution in a rectangular region is minimized when the components are independently distributed. Denoting the distribution function of N (0, 1) by $\Phi(x)$, we can imply from Lemma 2,

この不等式については各成分が独立に分布するときに矩形域のk次元正規分布の確率は最小になることを示している．これより $\Phi(x)$ を標準正規分布の分布関数とするとき，極限分布について

$$\lim_{n\to\infty} P(T \geqslant t) \leqslant 1 - (2\Phi(\sqrt{t}) - 1)^{c-1}. \tag{2.13}$$

Putting the right-hand side of (2.13) = 0.05, we have the approximately conservative 5% point of T, which is given by

これより右辺を0.05になるようにtを定めればconservativeな5％点が周辺度数によらずに求められる．$t(\alpha)$ を標準正規分布の両側100 α ％点とするときTの近似的5％点は

$$t\,(1 - 0.95^{\frac{1}{c-1}})^2, \tag{2.14}$$

Where $t(\alpha)$ is the two-sided $100\alpha\%$ point of $\Phi(x)$. For c = 2∿7 the values are computed from Pearson and Hartley.[3]

で与えられる．c＝2～7についてPearson and Hartleyの表[3]よりこの値を求めると次のようになる．

TABLE 1 APPROXIMATE 5% POINTS OF T

表1　Tの近似的5％点

c	2	3	4	5	6	7
t	3.84	5.00	5.70	6.22	6.60	6.94

NUMERICAL EXAMPLES AND APPLICATION

数値例と応用

We shall take the following 2 × 3 tables:

次の二つの2×3表について試みよう．

X_{11}	X_{12}	X_{13}	17
X_{21}	X_{22}	X_{23}	13
13	11	6	30

Case 1.

例1

X_{11}	X_{12}	X_{13}	16
X_{21}	X_{22}	X_{23}	7000
5764	998	254	7016

Case 2.

例2

Case 1. This contingency table in Yates[4] was used in another problem by Sugiura and Otake.[5] For given marginals the number of observable tables is 74, several of which give the same value of $T = \max[\chi_1^2, \chi_2^2]$. From (2.1) we can get:

例1．Yates[4]のこの分割表は以前杉浦と大竹[5]が他の問題で用いたものである．周辺度数を一定にしたとき，可能な分割表の個数は74であるが，そのうちいくつかの異なる分割表が同一の$T = \max(\chi_1^2, \chi_2^2)$の値を与える．(2.1)よりexactな確率を計算すれば次の表を得る．

TABLE 2 EXACT DISTRIBUTION OF T

表2　Tのexact分布

t	P (T ≥ t)	t	P (T ≥ t)	t	P (T ≥ t)
30.0	0.835×10^{-8}	7.30	0.0130	1.66	0.429
22.4	0.185×10^{-5}	6.27	0.0272	1.47	0.514
17.5	0.217×10^{-4}	5.74	0.0457	1.03	0.632
15.9	0.110×10^{-3}	4.89	0.0805	0.305	0.806
11.9	0.782×10^{-3}	3.83	0.119	0.222	0.905
10.5	0.241×10^{-2}	3.10	0.181	0.136	1.000
9.81	0.524×10^{-2}	2.17	0.288		

When c = 3, Table 1 gives an approximate critical point of T as 5.00. Table 2 gives the exact probability of P (T ≥ 5.00) = 0.0457, whereas the better formula (2.11) gives P (T ≥ 5.00) = 0.048, based on the table for bivariate normal distribution.[6]

c＝3のとき表1では近似的5％点は5.00である．表2よりexactな確率はP（T≧5.00）＝0.0457を得る．もし第3節の(2.11)を使えばよりよい近似が得られるはずであるが，二次元正規分布に関する表を使って計算すればP（T≧5.00）＝0.048を得る．[6]

Case 2. Similarly the eaxct probability is given by P (T ≥ 5.00) = 0.0306 and the formula (2.11) gives P (T ≥ 5.00) = 0.048. Hence our approximation is reasonable.

例2．同様にして，例2のexactな確率はP（T≧5.00）＝0.0306であり，一方(2.11)の近似はP（P≧5.00）＝0.048を得た．したがって，近似的5％点は実用上十分使えるであろう．

Finally as an application of the distribution of T, we shall consider the following 2 × 7 contingency table given in Otake and Jablon.[1]

最後に，Tの極限分布の1応用例として，大竹とJablon[1]によって示された次の2×7分割表を考えてみよう．

TABLE 3 NUMBER OF LEUKEMIA CASES OBSERVED FOR THE PERIOD 1 OCTOBER 1950-30 SEPTEMBER 1966 AMONG HIROSHIMA MALE SURVIVORS FOR THE EXTENDED LIFE SPAN STUDY SAMPLE AT ABCC AGED 15-39 AT THE TIME OF THE A-BOMB

表3　寿命調査拡大対象群のうち広島の男の被爆生存者で原爆時15－39歳の者において1950年10月1日から1966年9月30日までの期間中に観察された白血病例

Dose (rad)	< 5	5～	20～	50～	100～	200～	300+	Total
Leukemia	2	0	3	2	2	2	5	16
Not Leukemia	4601	1161	477	271	243	98	149	7000
Total	4603	1161	480	273	245	100	154	7016

If the observed value of $T = \max [\chi_1^2, \chi_2^2, \cdots, \chi_7^2]$ is less than a preassigned constant, we shall conclude that classifying the data by dose is not necessary. If not, we shall take 2×2 table giving the maximum value. In this case $T = \chi_5^2 = 74.01$. The next step in this example is to do the same analysis for the first 5 columns regarding as a 2×5 table and for the 6th column vs. 7th column regarding as a 2×2 table separately. Again we may have additional spliting of each column in the 2×2 table giving χ_5^2, or if not, we shall conclude that only two dose-groups separated at 200 rad are appropriate presentation of the data. In this way we can reach final $2 \times c$ table ($c \leqslant 7$). This is the Otake and Jablon's procedure,[1] if we take 3.84 (critical point of χ^2-distribution with one degree of freedom) as a preassigned constant in all steps. Their procedure is also applied to multiway contingency table and decides an appropriate regrouping of the table. Now we can propose to use the approximate cirtical point of T in Table 1 instead of 3.84, which motivated this investigation. Irrespective of the critical value taken, their procedure for this example gives the same 2×3 table (Case 2), in which first group received radiation dose less than 20 rad, the second group 20-199 rad and the third group 200 rad and over, as in Otake and Jablon.[1]

もし統計量 $T = \max [\chi_1^2, \chi_2^2, \cdots, \chi_6^2]$ の最大値が前もって定めた基準値より小さければ，線量区分による分類は必要でないと結論づける．もしそうでなければ，最大値を与えるところで分割して2×2表を作る．この例では，$T = \chi_5^2 = 74.01$ が最大値を与える．次の段階として，2×2表として分割された一つは2×5表である第1列対第2列…対第5列の群と，一方の群は2×2表である第6列対第7列に対して同様の方法を別々に行なっていく．χ_5^2統計量の最大値によって得られた2×2表の各列はさらに分割することが可能であるが，もし列の分割ができないとすれば，200 radで区分した二つの線量群は特にデータの適切な分類を表示すると結論づける．この方法を続けると最後的には2×c表（c≦7）になる．もし前もって定めた判定値3.84（自由度1のχ^2の5％点）を全段階に用いると大竹とJablon方式[1]になる．かれらはこの方法をさらに多次元分割表に適用し，多次元分割表の適切なグループ結合を示している．この調査研究に対して，杉浦と大竹は3.84の代わりに表1のTの近似的5％点の使用を提案した．提案されたこの判定値によって試みたが，この例題に関しては大竹とJablon[1]と同じ2×3表（例2）を示した．すなわち，最初のグループは20 rad未満の放射線量を受けたグループであり，第2のグループは20－199 radに属するグループで，第3のグループは200 rad以上のグループであった．

REFERENCES
参考文献

1. OTAKE M, JABLON S: An approach to data reduction for special multiple classifications by means of a series of binary splits. Presented at the meeting of Japan Statistical Society, 1971

2. ŠIDÁK Z: Rectangular confidence regions for the means of multivariate normal distributions. J Am Statist Assoc 62 626-633, 1967

3. PEARSON ES, HARTLEY HO: Biometrika Tables for Statisticians. Vol. 1. Cambridge Univ Press, 1966

4. YATES F: Contingency tables involving small numbers and the χ^2 test. Suppl J Roy Statist Soc. 1 217-35, 1934

5. SUGIURA N, OTAKE M: Numerical comparison of improved methods of testing in contingency tables with small frequencies. Ann Inst Statist Math 20 505-17, 1968

6. NATIONAL BUREAU OF STANDARDS. Tables of the Bivariate Normal Distribution Function and Related Functions. Applied Mathematical Series 50. Washington, U.S. Government Printing Office, 1959