

IMPROVED MONTE CARLO ESTIMATION OF STATISTICAL SIGNIFICANCE
FOR TESTS OF TREND IN RATES OR PROPORTIONS

修正 Monte Carlo 推定による比率又は割合の傾向検定に
おける統計的有意性

KENNETH J. KOPECKY, Ph.D.

DALE L. PRESTON, Ph.D.



RADIATION EFFECTS RESEARCH FOUNDATION
財団法人 放射線影響研究所

A Cooperative Japan - United States Research Organization

日米共同研究機関

RERF TECHNICAL REPORT SERIES

放影研業績報告書集

The RERF Technical Reports provide the official bilingual statements required to meet the needs of Japanese and American staff members, consultants, and advisory groups. The Technical Report Series is not intended to supplant regular journal publication.

放影研業績報告書は、日米専門職員、顧問、諮問機関の要求に応えるための日英両語による公式報告記録である。業績報告書は通例の誌上発表論文に代わるものではない。

The Radiation Effects Research Foundation (formerly ABCC) was established in April 1975 as a private nonprofit Japanese Foundation, supported equally by the Government of Japan through the Ministry of Health and Welfare, and the Government of the United States through the National Academy of Sciences under contract with the Department of Energy.

放射線影響研究所(元 ABCC)は、昭和50年4月1日に公益法人として発足したもので、その経費は日米両政府の平等分担により、日本は厚生省の補助金、米国はエネルギー省との契約に基づく米国学士院の補助金とをもって運営されている。

IMPROVED MONTE CARLO ESTIMATION OF STATISTICAL SIGNIFICANCE
FOR TESTS OF TREND IN RATES OR PROPORTIONS修正 Monte Carlo 推定による比率又は割合の傾向検定に
おける統計的有意性KENNETH J. KOPECKY, Ph.D.¹; DALE L. PRESTON, Ph.D.²*Department of Epidemiology & Statistics¹ and Computer Center²*疫学統計部¹ 及びコンピューターセンター²

SUMMARY

Asymptotic significance levels of tests for monotone trends in rates or proportions can be profoundly anticonservative when applied to small numbers of events and when distributions of exposure to risk are highly skewed. In such cases Monte Carlo (MC) estimation of observed levels of significance ("p-values") can be very useful. We describe a simple technique of importance sampling (IS) which can greatly improve the efficiency of MC estimation in this setting. Implementation of the IS technique is described, and the variance of the IS estimator is derived. It is shown that, in many situations likely to occur in practice, the variance is substantially less than that of a simple MC estimator proposed earlier. Generalizations beyond the case of survival data without ties are described, and the use of IS is illustrated with data regarding mortality among atomic bomb survivors.

INTRODUCTION

Tests of trends in rates or proportions, such as the procedures described by Mantel,¹ Tarone,² or Beebe et al,³ or certain applications of Cox's proportional hazards regression model,⁴ are widely used as statistical tools for epidemiologic research and analyses of survival data. Most such tests are conditional procedures, and therefore the resulting distributions of the test statistics are usually data dependent. They are

要 約

比率又は割合における単調傾向のテストの漸近的有意レベルは、少数の事象に適用する場合及び危険への曝露の分布に大幅な歪みがある場合には、大変非保守的になる。そのような場合には、Monte Carlo (MC) 推定として観測された有意水準("p 値")が大変有効である。このような状況において MC 推定の有効性を大幅に改善し得る簡単な方法、重要性抽出 (IS) について述べた。IS 手法適用設定について述べ、IS 推定量の分散を導き出した。実際起こりそうな多くの場合において、分散は以前提案された単純な MC 推定量のものより実質的に少ないことを示した。Tie のない生存データ例を越える一般化について述べ、IS の効用を、原爆被爆者の死亡率に関するデータを用いて示した。

緒 言

Mantel,¹ Tarone,² Beebe ら³ が述べている方法、又はある種 Cox の比例危険回帰モデル⁴ の適用など、比率又は割合の傾向検定は、疫学的研究及び生存データ解析の統計的手段として広く使用されている。このような検定のほとんどは条件付きの方法であるので、結果として得られる検定統計量の分布は通常

also quite often computationally intractable. For these reasons evaluation of the statistical significance of a trend test result is usually based on large-sample theory; in particular observed levels of significance (p-values) are frequently calculated from asymptotic normal or χ^2 distributions. The accuracy of such approximations must of course be considered in each particular application.

Hoel and Jennrich⁵ demonstrated by an example that the asymptotic χ^2 distribution of the trend test statistic described by Mantel¹ and Tarone² can, if used without caution, provide severely inaccurate results. In particular, if the number of events is not large and if the distribution of exposure to the factor of interest is highly skewed, then the distribution of the statistic will also be quite skewed. As a remedy for the resulting inaccuracy, Beebe et al⁶ and Tokunaga et al⁷ utilized Edgeworth series approximations to correct trend test p-values. However, as illustrated below, this approach can also lead to inaccurate results. Hoel and Jennrich⁵ proposed two simple procedures which are based on MC integration of probability functions and which provide accurate assessments of the statistical significance of a trend test result. Their second method, which they called the "simulation test", was recommended, since unlike the first method it is appropriate in situations of unequal censoring. In this paper we describe a modification of the simulation test which exploits the IS technique and which can in realistic situations greatly improve the efficiency of the MC procedure.

In the following section we define the general test for trend and illustrate the potential for inaccuracy in both the usual asymptotic distribution theory and the approach based on Edgeworth series approximations. The MC estimator of Hoel and Jennrich⁵ is then described, and the improved estimator is defined and its variance derived. The use of the improved estimator is illustrated with data from a survey of mortality and radiation exposure among A-bomb survivors of Hiroshima and Nagasaki. We conclude with some additional remarks, including a brief comment on the justification for the improved procedure. The Appendix provides a listing of two short FORTRAN subroutines by which the procedure can be implemented.

データに依存する。それらの分布はまた計算上の処理も困難であることが多い。これらの理由によって、傾向検定結果の統計的有意性の評価は通常大標本理論に基づいており、特に有意性の観測レベル(p値)はしばしば漸近正規分布又は χ^2 分布から計算される。言うまでもなく、これらの近似値の精度については、特定の各々の応用状況に即して考慮する必要がある。

Hoel 及び Jennrich⁵ は、1例を用いて、Mantel¹ 及び Tarone² が報告した傾向検定統計量の漸近的 χ^2 分布が不注意に使用されると極めて不正確な結果をもたらす得ることを示した。特に、事象数が多くはなく、興味のある因子への曝露の分布に大きな歪みがある場合、統計量の分布にもかなりの歪みが起こる。このような不正確な結果の修正法として、Beebeら⁶ 及び徳永ら⁷ は、Edgeworth級数近似を使用して傾向検定のp値を修正した。しかし、下記に示すとおり、この方法も不正確な結果をもたらす可能性がある。Hoel 及び Jennrich⁵ は、確率関数のMC積分に基づくもので傾向検定結果の統計的有意性を正確に評価できる2通りの簡単な方法を提案した。彼らが“シミュレーション検定”と称する第2の方法は、第1の方法とは異なり、不均等な censoring の場合にも適当であることから、この方法を勧めている。本報告では、IS手法を開発し、実際の状況でMC手法の効率を大きく改善し得るシミュレーション検定の修正について述べる。

次の項では、一般的傾向検定を定義し、通常の漸近的分布理論及びEdgeworth級数近似に基づく方法によって不正確な結果がもたらされる可能性を示す。次いで、Hoel 及び Jennrich⁵ のMC推定量について述べ、修正推定量を定義し、その分散を導き出す。広島・長崎における原爆被爆者の死亡率及び放射線被曝の調査から得られたデータを用いて修正推定量の使用を例示する。文末の付記には、修正した手法の妥当性に関する簡単な注釈や他の幾つかの注意を示す。付録は、この手法を実行するための二つの短いFORTRANサブルーチンを示したものである。

Tests of Trend

Notation

We consider data which are organized into I tables of dimensions $2 \times J$. The index $i=1, \dots, I$ identifies strata and/or risk sets. Following Mantel,¹ for example, we might have I separate contingency tables. For analysis of survival data a risk set may be defined in the usual way for each distinct failure time t_1, \dots, t_I (tied or discrete failure times can easily be accommodated in the procedure described below). More generally, for survival data which involve K strata, we can define stratum-specific risk sets associated with each stratum's distinct failure times t_{k1}, \dots, t_{kI_k} , for $k=1, \dots, K$; then $I = I_1 + \dots + I_K$. For simplicity, the term "strata" will be used in this paper, regardless of context.

The index $j=1, \dots, J$ identifies levels or categories of the exposure of interest. Let d_{ij} denote a score (classmark, dose, etc.) assigned to exposure group j in stratum i . We assume without loss of generality that $d_{i1} < \dots < d_{iJ}$ for all i . Frequently the same scores apply to all strata, i.e., $d_{1j} = \dots = d_{Ij} \equiv d_j$ for all j .

The first row of each $2 \times J$ table contains the numbers of events (deaths, diagnoses, etc.) in the exposure categories: n_{ij} for $j=1, \dots, J$. Let $n_{i.} = n_{i1} + \dots + n_{iJ}$. The second row contains the "risk quantities" R_{ij} for $j=1, \dots, J$. These may be either numbers of subjects at risk, in which case the tables are equivalent to true $2 \times J$ contingency tables as considered by Mantel.¹ Alternatively, for an analysis of event rates, such as that described by Beebe et al.,⁶ the R_{ij} might be person-years at risk. For convenience, let $R_i = (R_{i1}, \dots, R_{iJ})$ and define $\pi_{ij} = R_{ij}/R_i$ for all i and j .

Finally, let j_{im} denote the exposure category of the subject who suffered the m -th event in stratum i , for $i=1, \dots, I$ and $m=1, \dots, n_{i.}$

Logrank Trend Test

The statistic for the logrank test for trend is simply

$$S = \sum_{i=1}^I \sum_{j=1}^J n_{ij} d_{ij} = \sum_{i=1}^I \sum_{m=1}^{n_{i.}} d_{ij_{im}} ,$$

傾向検定

記号

$2 \times J$ 次元の I 個の表を構成するデータについて考える。指標 $i=1, \dots, I$ は層及び(又は)リスク集合を示す。例えば Mantel¹の方法によると、 I 個の別々の分割表が得られている状況に対応するであろう。生存データの解析については、各々の相異なる failure time t_1, \dots, t_I について通常の方法でリスク集合が定義される(tieのある場合又は離散値の failure time の場合も容易に下記の方法に組み入れることができる)。より一般的には、 K 個の層からなる生存データに関しては、各層 k , $k=1, \dots, K$ において相異なる failure time t_{k1}, \dots, t_{kI_k} に対応する層別リスク集合を定義することができる。したがって、 $I = I_1 + \dots + I_K$ となる。簡易化のために本報告では、文脈にかかわらず「層」という語を用いる。

指標 $j=1, \dots, J$ は興味の対象である被曝レベル又は被曝カテゴリーを示す。 d_{ij} は i 層の被曝群 j に割り当てたスコア(階級値, 線量等)を示すものとする。普遍性を失うことなく、すべての i について $d_{i1} < \dots < d_{iJ}$ であると仮定する。しばしば、このスコアはすべての層に適用される。すなわち、すべての j について $d_{1j} = \dots = d_{Ij} \equiv d_j$ である。

各 $2 \times J$ 表の第1行には被曝カテゴリーにおける事象(死亡, 診断等)数を含む。すなわち各 j , $j=1, \dots, J$ について n_{ij} とする。 $n_{i.} = n_{i1} + \dots + n_{iJ}$ とする。第2行には"危険量" R_{ij} (ただし $j=1, \dots, J$) が含まれる。これらは at risk での対象者数であってもよいが、その場合、それらの表は、Mantel¹ が考えた真の $2 \times J$ 分割表と等しくなる。また、Beebeら⁶ が報告しているような事象の比率の解析の場合には、 R_{ij} は at risk での人年としてもよい。便宜上、 $R_i = (R_{i1}, \dots, R_{iJ})$ とし、すべての i 及び j について、 $\pi_{ij} = R_{ij}/R_i$ と定義する。

最後に、 j_{im} は i 層における第 m 番目の事象を経験した対象者の被曝カテゴリーを示す。ただし、 $i=1, \dots, I$ 及び $m=1, \dots, n_{i.}$ である。

Logrank 傾向検定

Logrank 傾向検定の統計量は簡単に

that is, the sum of the scores of the $n_{i.}$ subjects who suffered the event.^{1,2} A large value of S is evidence that the rate or proportion of events increases with the level of exposure. The widely-used conditional test of significance may be described as follows.⁵ Conditional on each $n_{i.}$, the observed total number of deaths in the i -th stratum, and on each R_i , the values of j_{i1}, \dots, j_{in_i} are considered to be a realization of a random selection from the set $\{1, \dots, J\}$. If the R_{ij} represent subjects at risk, then the sampling is "without replacement", i.e., the probability function is

$$\Pr(j_{i1}, \dots, j_{in_i} | n_{i.}; R_i) = \left\{ \prod_{j=1}^J \frac{R_{ij}!}{(R_{ij} - n_{ij})!} \right\} / \left\{ \frac{R_i!}{(R_i - n_{i.})!} \right\} \quad [1]$$

If, on the other hand, the R_{ij} represent person-years at risk, then the sampling is "with replacement" and

$$\Pr(j_{i1}, \dots, j_{in_i} | n_{i.}; R_i) = \prod_{j=1}^J \pi_{ij}^{n_{ij}} \quad [2]$$

In either case, the selection is assumed to be conditionally independent between strata, so

$$\Pr(j_{11}, \dots, j_{In_I} | n_{1.}, \dots, n_{I.}; R_1, \dots, R_I) = \prod_{i=1}^I \Pr(j_{i1}, \dots, j_{in_i} | n_{i.}; R_i)$$

and the observed level of significance or p-value, calculated directly, is

$$\begin{aligned} P_D &= \text{Prob}(S \geq S_{\text{obs}} | n_{1.}, \dots, n_{I.}; R_1, \dots, R_I) \\ &= \sum_{S \geq S_{\text{obs}}} \Pr(j_{11}, \dots, j_{In_I} | n_{1.}, \dots, n_{I.}; R_1, \dots, R_I) \end{aligned} \quad [3]$$

where S_{obs} is the observed value of S . Direct computation of P_D , which involves convolutions of discrete probability functions under the sampling described above, is frequently infeasible. As Hoel and Jennrich⁵ pointed out, P_D is an approximation of the true level of significance, since the assumption of conditional independence which leads to [3] is generally not in fact true. Frequently, however, the approximation is quite good.

と書ける。すなわちこれは事象を経験した対象者 $n_{i.}$ 人のスコアの合計である。^{1,2} S の値が大きいことは、事象の比率又は割合が被曝レベルとともに増加することを示している。広く用いられている条件付き有意性検定は下記のとおりである。⁵ 第 i 番目の層における観察総死亡数 $n_{i.}$, また各 R_i を条件において、値 j_{i1}, \dots, j_{in_i} は、集合 $\{1, \dots, J\}$ からの無作為抽出の実現値と考えられる。 R_{ij} が at risk での対象者を表すならば、抽出は“置き換えなし”である。すなわち確率関数は次のようになる。

一方、 R_{ij} が at risk での人年を表すならば、抽出は“置き換えあり”で次のようになる。

どちらの場合にも、抽出は各層間では条件付きで独立していると考えられるので、

であり、観察有意水準、すなわち p 値を直接計算すると次のとおりになる。

ただし、 S_{obs} は S の観察値である。 P_D を直接計算するためには、上記の抽出の際の離散確率関数のたたみ込みが必要であるが、これはしばしば実行不可能である。Hoel 及び Jennrich⁵ が指摘したとおり、[3] へ導く条件付き独立に関する仮定は実際は一般的には真でないので、 P_D は真の有意レベルに関する近似である。しかし、この近似はかなり良いことが多い。

The asymptotic approximation of P_D is derived by noting that

$$E = \sum_{i=1}^I n_i \cdot \left\{ \sum_{j=1}^J d_{ij} \pi_{ij} \right\}$$

is the conditional expectation of S , while the appropriate conditional variance is either

$$V = \sum_{i=1}^I \frac{n_i \cdot (R_i - n_i)}{(R_i - 1)} \left\{ \sum_{j=1}^J d_{ij}^2 \pi_{ij} - \left[\sum_{j=1}^J d_{ij} \pi_{ij} \right]^2 \right\} \quad [4]$$

or

$$V = \sum_{i=1}^I n_i \cdot \left\{ \sum_{j=1}^J d_{ij}^2 \pi_{ij} - \left[\sum_{j=1}^J d_{ij} \pi_{ij} \right]^2 \right\}, \quad [5]$$

depending on whether R_{ij} represents subjects at risk or person-years at risk, respectively. The standardized statistic $(S-E)/V^{1/2}$ is, in most applications, assumed to be approximately normally distributed. That is, the p -value of the test is given by

$$P_N = 1 - \Phi \left\{ (S_{\text{obs}} - E)/V^{1/2} \right\}, \quad [6]$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Frequently the corresponding two-tailed result is expressed in terms of the χ^2 distribution of $(S_{\text{obs}} - E)^2/V$.

In epidemiologic or occupational health studies it is not uncommon to find that the distribution of exposure to a particular risk factor is quite skewed. For example, skewness to the right occurs when most of a population suffers little or no exposure to a hazardous agent while a small subset receives very high exposures. In such cases the distribution of S , or its standardized version in [6], is likely to retain a marked skewness in a study with a small or even moderate number of events. Therefore, for example, use of the large-sample approximation P_N in a test for a positive trend with data skewed toward high exposures is likely to be anticonservative, markedly so if the skewness is pronounced and the number of deaths is small or moderate.

The use of Edgeworth series approximations to correct for nonnormality of the distribution of

P_D の漸近近似については

が S の条件付き期待値であり、適切な条件付き分散は、 R_{ij} が at risk の対象者を表す場合又は at risk の人年を表す場合によって、それぞれ次のようになることに気を付けることにより導くことができる。

又は

ほとんどの適用の場合、正規化統計量 $(S-E)/V^{1/2}$ は近似的な正規分布となっていると考えられる。すなわち、検定の p 値は

である。ただし $\Phi(\cdot)$ は標準正規分布の分布関数である。対応する両側検定結果はしばしば $(S_{\text{obs}} - E)^2/V$ の χ^2 分布で表される。

疫学的又は職業的健康調査において、特定の危険因子への曝露の分布に大幅な歪みが起こることは珍しくない。例えば、集団のほとんどが危険因子へほとんど又は全く曝露しておらず、小集団が極めて高い線量に被曝している場合には、右側への歪みが起こる。このような場合には、 S の分布あるいは [6] のようにその正規化を行ったものでは、少数の事象あるいはある程度の数の事象があっても、それらを伴う研究では著しい歪みを持続しがちである。したがって、例えば高線量被曝への歪みを示すデータによって正の傾向を調べる検定で、大標本近似 P_N を使用すれば非保守的になりやすく、歪みが著しく死亡例数が少ないか又はある程度しか考えられない場合には、この傾向は著しくなる。

検定統計量分布の非正規性を修正するために、Edgeworth 級数近似を使用することは、十分に確立

a test statistic is a well-established technique.⁸ This approach has been adopted in analyses of morbidity among A-bomb survivors, for whom the distribution of radiation doses is very heavily skewed to the right. Beebe et al⁶ provided a table of skewness-corrected p-values as a function of the normal approximation, P_N , and the total number of deaths, $n_{..}$. Tokunaga et al⁷ reported p-values that were corrected for the skewness and kurtosis of the trend statistic. As shown in the following example, however, there is a possibility that even the use of such corrections may fail to provide an accurate assessment of the significance of a trend test result.

Example

Hoel and Jennrich⁵ illustrated their procedure with data concerning mortality from multiple myeloma in relation to radiation dose received by A-bomb survivors of Hiroshima and Nagasaki. The collection of these data is described elsewhere.⁹ A particularly interesting subset of the data is summarized in Table 1. Among Hiroshima females aged 20-34 years at the time of the bomb (ATB), two deaths from multiple myeloma occurred during the follow-up period 1950-78. One was a woman in the 10-49 rad dose group, the other in the 400+ rad group. Thus $I = 2$. The dose category scores are simply the average doses based on all survivors in each category, and $S_{obs} = d_3 + d_8 = 546$. Also $E = 46.6$ and from either [4] or [5], $V = 9408.8$, so the normal approximation of the one-sided p-value for an increasing trend is $P_N = 1 - \Phi(5.15) = 1.4 \times 10^{-7}$. Hoel and Jennrich⁵ argued heuristically that this value overstates the real statistical significance of the observed trend.

された方法の一つである。⁸ 放射線量分布が右側へ大幅に歪んでいる原爆被爆者の罹患率の解析には、この方法が使用されてきた。Beebeら⁶は、正規近似 P_N 及び総死亡例数 $n_{..}$ の関数としての歪みを補正した p 値の表を呈示している。徳永ら⁷は傾向統計量の歪度及び尖度について補正した p 値を報告している。しかし、下記の例に示すとおり、このような補正を用いても傾向検定結果の有意性を正確に評価できない可能性がある。

例

Hoel 及び Jennrich⁵ は、多発性骨髄腫による死亡率と広島・長崎の原爆被爆者が受けた放射線量との関係に関するデータを用いて、彼らの方法を示した。これらのデータの収集については他に記述がある。⁹ 特に興味深いデータの一部を表 1 に要約した。原爆時年齢が 20～34 歳であった広島的女性では、1950—78 年の追跡調査期間中、多発性骨髄腫による死亡が 2 例認められた。その 1 人は 10～49 rad 線量群、他の 1 人は 400+ rad 群の女性であった。したがって $I = 2$ である。線量カテゴリーのスコアは、単に各カテゴリーの全被爆者に基づく平均線量であり、 $S_{obs} = d_3 + d_8 = 546$ となる。また、 $E = 46.6$ であり、[4] 又は [5] から $V = 9408.8$ となるので、増加傾向の片側 p 値の正規近似は $P_N = 1 - \Phi(5.15) = 1.4 \times 10^{-7}$ となる。Hoel 及び Jennrich⁵ は、この値が観察傾向の真の統計的有意性を誇張することを経験的に論じている。

TABLE 1 EXAMPLE OF LOGRANK TEST FOR TREND (MULTIPLE MYELOMA MORTALITY, 1950-78; HIROSHIMA FEMALES AGED 20-34 ATB)

表 1 Logrank 傾向検定例 (多発性骨髄腫による死亡率, 1950～78年; 原爆時年齢20～34歳の広島的女性)

	Radiation dose category in rad							
	0	1-9	10-49	50-99	100-199	200-299	300-399	400+
Score	0	4	22	70	141	242	343	524
Risk set for death No.1	3972	2303	1636*	475	250	97	59	95
Risk set for death No.2	3904	2270	1610	469	247	95	58	92*

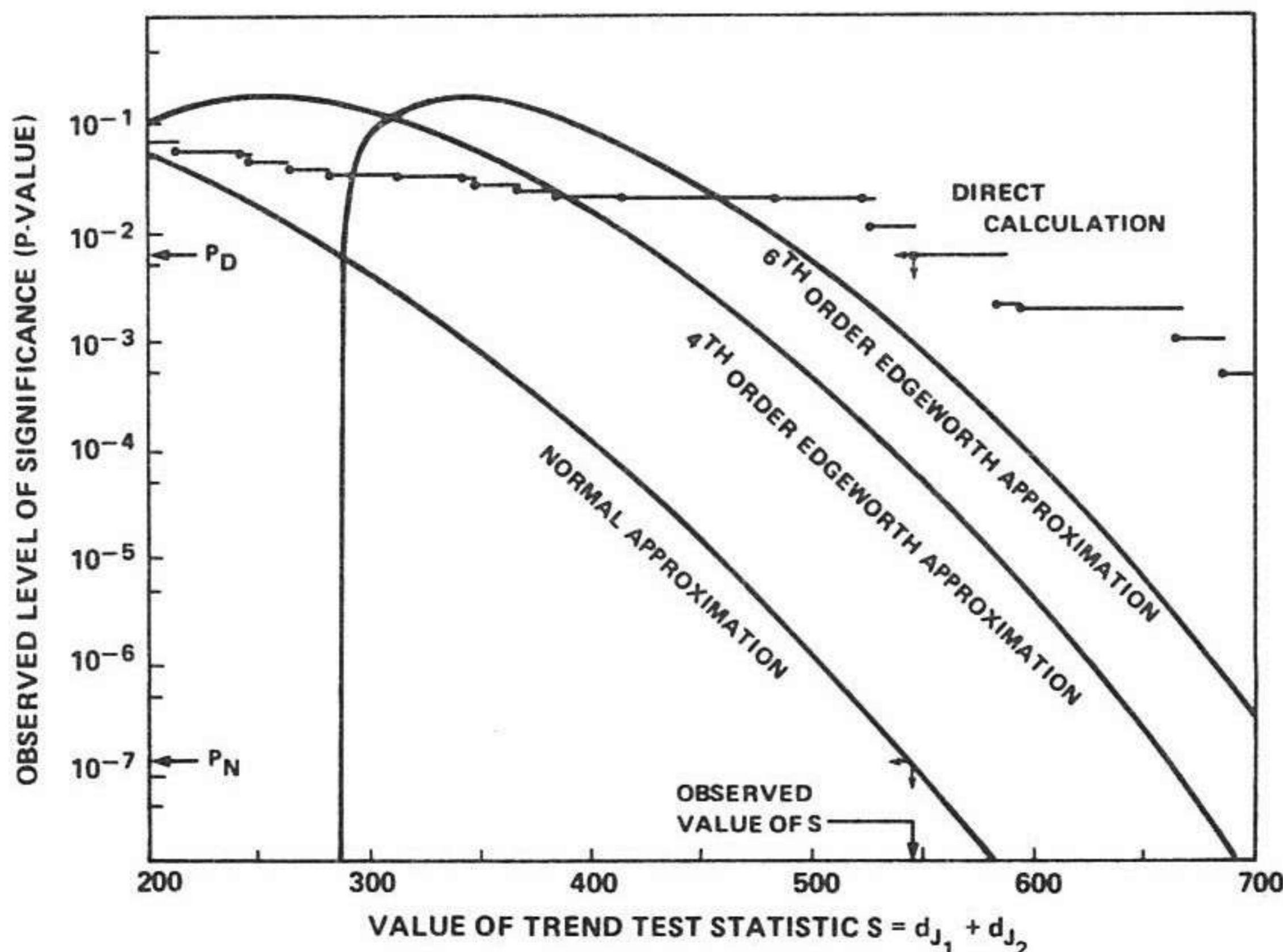
*Dose categories in which deaths occurred. 死亡例が認められた線量カテゴリー

For this simple example with $n_{..} = 2$ deaths, it is easy to calculate P_D directly as a function of S . Figure 1 illustrates this function over a range of S such that P_D lies between about 0.001 and 0.1. This is the range of greatest interest, within which a poor approximation of P_D may lead to an erroneous conclusion. For $S_{obs} = 546$ we find that $P_D = 0.0063$. In fact the normal approximation is very poor over most of the range plotted. Although fourth- and sixth-order Edgeworth series approximations provide more accurate assessments of significance than P_N for some values of S , they can still err by more than one order of magnitude. Moreover, such approximations are not monotonically nonincreasing and can take negative values; this makes it difficult to know whether they provide a conservative or anticonservative result for a particular problem.

死亡例数 $n_{..} = 2$ であるこの簡単な例では、 S の関数として直接に P_D を計算することは容易である。図1は、 P_D が約0.001と0.1の間になるような S の範囲におけるこの関数を示す。この範囲では、 P_D の近似の悪さのために誤った結論に導くこともあるので、極めて興味深い範囲である。 $S_{obs} = 546$ の場合、 $P_D = 0.0063$ となる。事実、正規近似はプロットした範囲のほとんどにわたって極めて悪い。第4及び第6次 Edgeworth 級数近似では、幾つかの S の値に関して、 P_N よりも正確な有意性評価が得られるが、その場合でさえも一桁以上の誤りが起こり得る。更に、このような近似は単調非増加ではなく負の値もとり得る。このことによって、特定の問題に関してそれらの近似が保守的な結果をもたらすか非保守的な結果をもたらすか知ることが困難となっている。

FIGURE 1 ACCURACY OF APPROXIMATE P-VALUES:
MORTALITY FROM MULTIPLE MYELOMA, 1950-78 (HIROSHIMA FEMALES AGED 20-34 ATB)

図1 近似P値の精度：多発性骨髄腫による死亡率，1950～78年
(原爆時年齢20～34歳の広島的女性)



Monte Carlo Procedures

Simple Procedure

In order to obtain accurate assessments of the statistical significance of trends, Hoel and Jennrich⁵ proposed that MC integration be utilized to estimate P_D . This was first suggested as a general approach by Barnard, in discussion of a paper by Bartlett,¹⁰ and has been investigated primarily as a kind of randomized procedure for tests conducted at fixed critical levels.¹¹⁻¹³ Kellerer and Chmelevsky¹⁴ used such a procedure to estimate exact p-values for two-sample rank tests with censored data. With the wide availability of high-speed computers, MC integration has become a well-known technique.^{15,16}

In fact, Hoel and Jennrich⁵ proposed two procedures: the first based on a distribution of permutations of the assignments of scores, d_{ij} , to events, and the second based on the discrete sampling scheme as described above. This presentation is concerned with the second procedure, which they called the "simulation test", since it is appropriate in situations of unequal censoring.

The procedure of Hoel and Jennrich⁵ is conceptually and computationally quite simple. A realization of S is produced by random or pseudorandom sampling of values of j_{im} from the discrete conditional distributions described above. Let $S^{(k)}$ denote the k -th such realization, for $k=1, \dots, K$. The k -th realization is a "hit" if $S^{(k)} \geq S_{obs}$, and the estimate of P_D is simply the proportion of hits among the K replications:

$$\hat{P}_{HJ} = K^{-1} \sum_{k=1}^K I(S^{(k)} \geq S_{obs}),$$

where $I(\cdot)$ is the indicator function. Clearly \hat{P}_{HJ} is an unbiased estimate of P_D , that is $E(\hat{P}_{HJ}) = P_D$. Also

$$V(\hat{P}_{HJ}) = K^{-1} (P_D - P_D^2) . \quad [7]$$

Improved Procedure

A more efficient estimator of P_D can be obtained by employing the variance-reducing technique known as IS.¹⁶ In this approach, the sampling for each of the K replications is weighted to increase the probability of a hit, and each hit is accordingly down-weighted in the final estimator, \hat{P}_{IS} . As shown below, a simple application of this technique can lead to quite useful reductions in the variance of the MC estimator.

Monte Carlo 法

簡単な方法

傾向の統計的有意性を正確に評価するために, Hoel 及び Jennrich⁵ は P_D の推定に MC 積分を用いることを提案した. これは最初に Bartlett¹⁰ の報告の考察部分で Barnard による一般的方法として提案され, 主として固定された有意レベルで実施される検定について一種のランダム化の方法として研究されている.¹¹⁻¹³ Kellerer 及び Chmelevsky¹⁴ は, censored されたデータを用いて 2 標本順位検定の正確な p 値を推定するためにこの方法を用いた. 高速コンピューターが広範に利用されるようになり, MC 積分は広く知られる方法となった.^{15, 16}

事実, Hoel 及び Jennrich⁵ は 2 通りの方法を提案している. 第 1 の方法はスコア d_{ij} の事象への割り当ての順列分布に基づき, 第 2 の方法は上記の離散標本抽出法に基づく. この第 2 の方法は彼らのいわゆる "シミュレーション検定" であり, 不均等な censoring のある場合に妥当であるので, 本報ではこの方法について述べる.

Hoel 及び Jennrich⁵ の方法は概念的にも計算上でもかなり簡単である. S の実現は上記の条件付き離散分布から j_{im} 値を無作為又は擬似乱数的に抽出することによって求められる. $S^{(k)}$ は第 k 番目のこのような実現を示すものとし, $k=1, \dots, K$ とする. $S^{(k)} \geq S_{obs}$ ならば第 k 番目の実現は "命中" であり, P_D の推定値は単に K 個の反復における命中率となる:

ただし, $I(\cdot)$ は指標関数である. 明らかに \hat{P}_{HJ} は P_D の不偏推定値である. すなわち $E(\hat{P}_{HJ}) = P_D$ である. また, 次のようになる.

改訂方法

IS として知られる分散減少法を用いれば, P_D のより有効な推定量が得られる.¹⁶ この方法では, K 回の反復のうちの各々の抽出を加重して, 命中の確立を増加させ, その結果最終推定量 \hat{P}_{IS} において各命中を下向加重する. 下記に示すとおり, この方法を単純に応用すれば, MC 推定量の分散にかなり有益な減少がもたらされる.

Note that a necessary condition for a hit ($S \geq S_{obs}$) is that $d_{ij_{im}} \geq n^{-1} S_{obs}$ for some $i=1, \dots, I$ and $m=1, \dots, n_i$. If we define cut points

$$c_i = \min \{c | d_{ic} \geq n^{-1} S_{obs}\}, i=1, \dots, I,$$

then the necessary condition can be written as

$$j_{im} \geq c_i \text{ for some } i=1, \dots, I \text{ and } m=1, \dots, n_i. \quad [8]$$

ある適当な $i, i=1, \dots, I$ 及びある適当な $m, m=1, \dots, n_i$ に関して $j_{im} \geq c_i$.

For the current purpose, IS is performed by requiring that the condition in [8] be true for each MC replication $k=1, \dots, K$. This is accomplished by a two-step procedure:

Step 1. Select a stratum index ℓ at random from among $\{1, \dots, I\}$ according to a specified probability distribution $\{\alpha_1, \dots, \alpha_I\}$.

Step 2. Generate randomly the exposure categories j_{11}, \dots, j_{In_I} . If $i \neq \ell$ select j_{i1}, \dots, j_{in_i} according to the probability function [1] or [2] as appropriate. However, for the ℓ -th stratum, select $j_{\ell 1}$ from among $\{c_\ell, \dots, J\}$ according to the conditional distribution $\{\pi'_{\ell, c_\ell}, \dots, \pi'_{\ell, J}\}$ with $\pi'_{\ell, j} = \pi_{\ell, j} / (\pi_{\ell, c_\ell} + \dots + \pi_{\ell, J})^*$; then select $j_{\ell 2}, \dots, j_{\ell n_\ell}$ from among $\{1, \dots, J\}$ according to the appropriate unconditional relative frequencies (when sampling without replacement, the selection of $j_{\ell 1}$ is taken into account in selecting $j_{\ell 2}, \dots, j_{\ell n_\ell}$).

The probability function for selection of j_{11}, \dots, j_{In_I} under this scheme is

$$\Pr_{IS}(j_{11}, \dots, j_{In_I}) = \Pr(j_{11}, \dots, j_{In_I} | n_1, \dots, n_I; R_1, \dots, R_I) \prod_{i=1}^I \frac{\alpha_i I(j_{i1} \geq c_i)}{1 - F_i},$$

where $F_i = \pi_{i,1} + \dots + \pi_{i,c_i-1}$. It follows¹⁶ that

$$\hat{P}_{IS} = K^{-1} \sum_{k=1}^K \frac{I(S^{(k)} \geq S_{obs})}{\left[\prod_{i=1}^I \frac{\alpha_i I(j_{i1} \geq c_i)}{1 - F_i} \right]}$$

is an unbiased estimator of P_D .

命中 ($S \geq S_{obs}$) の必要条件は、ある適当な $i, i=1, \dots, I$ 及びある適当な $m, m=1, \dots, n_i$ が存在し、そのとき $d_{ij_{im}} \geq n^{-1} S_{obs}$ が成立することであることに留意すべきである。切断点を次のように定義すれば、

必要条件は次のように書き表すことができる。

今回の目的のために、[8]の条件が各MC反復 $k=1, \dots, K$ について真であることを必要条件としてISを実行する。これは2段階の方法を用いて実施する。

段階1. 特定の確率分布 $\{\alpha_1, \dots, \alpha_I\}$ に従って $\{1, \dots, I\}$ から無作為に層指数 ℓ を選択する。

段階2. 被曝カテゴリー j_{11}, \dots, j_{In_I} を無作為に作成する。 $i \neq \ell$ ならば、適宜に確率関数 [1] 又は [2] に従って j_{i1}, \dots, j_{in_i} を選択する。しかし、第 ℓ 番目の層については、 $\pi'_{\ell, j} = \pi_{\ell, j} / (\pi_{\ell, c_\ell} + \dots + \pi_{\ell, J})^*$ を伴う条件付き分布 $\{\pi'_{\ell, c_\ell}, \dots, \pi'_{\ell, J}\}$ に従って $\{c_\ell, \dots, J\}$ から $j_{\ell 1}$ を選択する。次いで、適当な条件付きでない相対頻度に従って $\{1, \dots, J\}$ から $j_{\ell 2}, \dots, j_{\ell n_\ell}$ を選定する (置き換えを伴わない抽出を行うとき、 $j_{\ell 2}, \dots, j_{\ell n_\ell}$ の選定の際には $j_{\ell 1}$ の選定を考慮する)。

この方法における j_{11}, \dots, j_{In_I} の選択の確率関数は、次のとおりである。

ただし $F_i = \pi_{i,1} + \dots + \pi_{i,c_i-1}$ である。その結果、

は P_D の不偏推定量となる。¹⁶

* For brevity, we ignore the possibility that $\pi_{\ell, c_\ell} + \dots + \pi_{\ell, J} = 0$. If this is true for some strata, then those strata are excluded from selection in Step 1.
簡略化のために、 $\pi_{\ell, c_\ell} + \dots + \pi_{\ell, J} = 0$ である可能性を無視する。これが幾つかの層について真であるならば、それらの層は段階1の選択から除外する。

The procedure described above is slightly different from and more general than that described by Kopecky,¹⁷ and is intended to provide an easy computational scheme that can accommodate data with ties or grouping ($n_{ij} > 1$) and conditional hypergeometric sampling. The Appendix contains listings of two simple FORTRAN subroutines which can be used to implement this IS method.

Example (continued)

The procedure of Hoel and Jennrich⁵ and the IS procedure are illustrated in Figure 2 for the multiple myeloma data of Table 1. The conditional sampling model for S, which is simulated for the estimator \hat{P}_{HJ} , is equivalent to selecting a cell at random from an 8x8 table. To each cell is assigned the score S, the sum of the column and row scores. We may define an empirical rejection region as those cells for which $S \geq S_{obs} = 546$. The columns, corresponding to the dose categories for the first death, are assigned relative frequencies (probabilities) $\pi_{11} = R_{11}/R_{1.} = 3972/8887 = 0.4469$, π_{12} , etc., and a column j_1 is selected at random accordingly. Similarly a row j_2 is selected independently according to the relative frequencies for the second risk set. A hit occurs if the selected cell lies in the empirical rejection region.

上記の方法は、Kopecky¹⁷の報告した方法とは少し異なってもっと一般的であり、tie 又は分類を伴うデータ ($n_{ij} > 1$) を扱うことができる簡単な計算方法及び条件付き超幾何標本抽出をもたらすように意図されたものである。付録には、このIS法の実施に用いられる2通りの簡単なFORTRAN サブルーチンを示した。

例 (続き)

図2は、表1の多発性骨髄腫に関するHoel及びJennrich⁵の方法並びにIS手法を示したものである。推定量 \hat{P}_{HJ} についてシミュレーションを行うSの条件付き抽出モデルは、8x8表から無作為にその要素を選択することに等しい。各欄に列及び行スコアの合計であるスコアSを割り当てる。先験的棄却域 (empirical rejection region) は、 $S \geq S_{obs} = 546$ である列として定義することもできる。最初の死亡の線量カテゴリーに相当する列に、相対頻度(確率) $\pi_{11} = R_{11}/R_{1.} = 3972/8887 = 0.4469$, π_{12} 等を割り当て、それによって列 j_1 を無作為に選定する。同様に、第2番目のリスク集合の相対頻度に従って行 j_2 を別個に選択する。選択した列が先験的棄却域にあるならば、命中が起こる。

FIGURE 2 SCHEMATIC REPRESENTATION OF MONTE CARLO SAMPLING FOR EXAMPLE OF HOEL AND JENNRICH

図2 Hoel 及び Jennrich の例に関する Monte Carlo 標本抽出の図式

DOSE GROUP	REL. FREQ.	DOSE GROUP							
		1	2	3	4	5	6	7	8
		RELATIVE FREQUENCY							
		.4469	.2591	.1841	.0534	.0281	.0109	.0066	.0107
1	.4464	0	4	22	70	141	242	343	524
2	.2596	4	8	26	74	145	246	347	528
3	.1841	22	26	44	92	163	264	365	546 19.7
4	.0536	70	74	92	140	211	312	413	594 5.7
5	.0282	141	145	163	211	282	383	484	665 3.0
6	.0109	242	246	264	312	383	484	585 0.7	766 1.2
7	.0066	343	347	365	413	484	585 0.7	686 0.4	867 0.7
8	.0105	524	528	546 19.4	594 5.6	665 3.0	766 1.1	867 0.7	1048 1.1

Note: Score in upper left of each cell is the sum of the row and column dose classmarks. The empirical rejection region consists of those cells below and to the right of the heavy line. Number in lower right of cells in the empirical rejection region is $10^4 \times$ probability of the cell under independent sampling from marginal distributions.

注 各セルの左上部の数字は、行及び列の線量階級点数の合計である。先験的棄却域は、太い線の下側及び右側の各小区画からなる。先験的棄却域における各セル右下部の数字は、周辺分布からの独立な標本抽出における小区画の $10^4 \times$ 確率である。

For the IS procedure, we note that $c_1 = c_2 = 7$; that is, in order for $S = d_{j_1} + d_{j_2}$ to equal or exceed S_{obs} , it is necessary that at least one of the two deaths occur in dose group 7 or 8. In effect, the shaded region in Figure 2 is excluded from the sampling.

This can be accomplished by selecting $\ell = 1$ or 2 at random according to probabilities α_1 and $\alpha_2 = 1 - \alpha_1$. If $\ell = 1$, then the column j_1 is chosen from $\{7, 8\}$ according to probabilities $\pi'_{17} = 59/(59+95)$ and $\pi'_{18} = 95/(59+95)$; the row j_2 is chosen from $\{1, \dots, 8\}$ according to the row relative frequencies in Figure 2. The procedure is analogous if $\ell = 2$.

Variance and Efficiency of \hat{P}_{IS}

Notice that the set of all $J^{n..}$ possible outcome vectors $(j_{11}, \dots, j_{In_I})$ can be partitioned into 2^I equivalence classes with

$$(j_{11}, \dots, j_{In_I}) \sim (j'_{11}, \dots, j'_{In_I})$$

defined by the condition that $I(j_{i1} \geq c_i) = I(j'_{i1} \geq c_i)$ for all $i=1, \dots, I$. Denote these classes Q_ℓ , $\ell=0, \dots, 2^I-1$, where Q_0 is the class of outcomes with $I(j_{i1} \geq c_i) = 0$ for all i . Q_0 is simply the set of outcomes which are excluded under the IS procedure. Obviously if $(j_{11}, \dots, j_{In_I}) \sim (j'_{11}, \dots, j'_{In_I})$, then

$$\sum_{i=1}^I \frac{\alpha_i I(j_{i1} \geq c_i)}{1 - F_i} = \sum_{i=1}^I \frac{\alpha_i I(j'_{i1} \geq c_i)}{1 - F_i}$$

Let $W(Q)$ denote the value of this expression on equivalence class Q . Then the variance of the \hat{P}_{IS} is

$$V_{IS}(\hat{P}_{IS}) = K^{-1} \left(\sum_{\ell=1}^{2^I-1} \left\{ \frac{\text{Prob}[S \geq S_{obs} \text{ and } (J_{11}, \dots, J_{In_I}) \in Q_\ell | n_{1..}, \dots, n_{I..}; R_1, \dots, R_I]}{W(Q_\ell)} \right\} - P_D^2 \right) \quad [9]$$

Notice that

$$\sum_{\ell=1}^{2^I-1} \text{Prob}[S \geq S_{obs} \text{ and } (J_{11}, \dots, J_{In_I}) \in Q_\ell | n_{1..}, \dots, n_{I..}; R_1, \dots, R_I] = P_D$$

Also, in the situations of interest (skewed exposure distribution and few deaths), we can expect that $W(Q_\ell) > 1$. Therefore, comparing [7] and [9], we can anticipate that $V(\hat{P}_{HJ}) > V_{IS}(\hat{P}_{IS})$ in such cases.

IS手法の場合は $c_1 = c_2 = 7$ であることに留意する。すなわち、 $S = d_{j_1} + d_{j_2}$ が S_{obs} と等しいか又はそれ以上であるためには、線量群 7 又は 8 において 2 例の死亡のうち少なくとも 1 例は生じることが必要である。図 2 の斜線部分は事実上抽出から除外した。

これは、確率 α_1 及び $\alpha_2 = 1 - \alpha_1$ に従って $\ell = 1$ 又は 2 を無作為に抽出することによって実行できる。 $\ell = 1$ ならば、確率 $\pi'_{17} = 59/(59+95)$ 及び $\pi'_{18} = 95/(59+95)$ に従って $\{7, 8\}$ から列 j_1 を選択する。図 2 における行の相対頻度に従って $\{1, \dots, 8\}$ から行 j_2 を選択する。 $\ell = 2$ のときも、同様の操作を行う。

\hat{P}_{IS} の分散及び効率

$J^{n..}$ 個からなる可能なすべてのベクトル事象 $(j_{11}, \dots, j_{In_I})$ の集合は、すべての $i=1, \dots, I$ に対して $I(j_{i1} \geq c_i) = I(j'_{i1} \geq c_i)$ であるという条件で同値関係

を定義することにより 2^I 個の同値類に分割できるとに留意する。これらの同値類 Q_ℓ を、 $\ell=0, \dots, 2^I-1$ として示す。ただし Q_0 は、すべての i に対して $I(j_{i1} \geq c_i) = 0$ となる事象の類である。 Q_0 は単に IS 手法においては除外される事象の集合であるにすぎない。明らかに、 $(j_{11}, \dots, j_{In_I}) \sim (j'_{11}, \dots, j'_{In_I})$ であるならば、

である。 $W(Q)$ を同値類 Q に関するこの式の値を示すものとすれば、その結果、 \hat{P}_{IS} の分散は

となる。ただし次のことに留意する。

また、興味のある状況(歪みのある被曝分布及び少ない死亡例数)においては、 $W(Q_\ell) > 1$ であることが期待される。したがって、[7]と[9]を比較すると、このような場合には $V(\hat{P}_{HJ}) > V_{IS}(\hat{P}_{IS})$ であることが予想される。

Example (continued)

For the estimator of Hoel and Jennrich⁵ applied to the data of Table 1, recalling that $P_D = 0.0063$, we find that $V(\hat{P}_{HJ}) = 0.00626K^{-1}$. For the IS estimator, assume $\alpha_1 = \alpha_2 = 0.5$, and let

$$\begin{aligned} Q_1 &= \{j_1, j_2 \mid 7 \leq j_1 \leq 8 \text{ and } 1 \leq j_2 \leq 6\} \\ Q_2 &= \{j_1, j_2 \mid 1 \leq j_1 \leq 6 \text{ and } 7 \leq j_2 \leq 8\} \\ Q_3 &= \{j_1, j_2 \mid 7 \leq j_1, j_2 \leq 8\} \end{aligned}$$

Then

$$\begin{aligned} W(Q_1) &= (0.5 \times 1) / .0173 + (0.5 \times 0) / .0172 = 28.85 \\ W(Q_2) &= 29.15 \\ W(Q_3) &= 58.00. \end{aligned}$$

Therefore

$$V_{IS}(\hat{P}_{IS}) = 0.000173K^{-1} = 0.0275 V(\hat{P}_{HJ}).$$

The efficiency of the IS estimator \hat{P}_{IS} , in terms of variance, is thus about 36, relative to the simple estimator \hat{P}_{HJ} . Since the simple procedure requires the generation of $n_{..} = 2$ random numbers per replication, while \hat{P}_{IS} requires $n_{..} + 1 = 3$, a more realistic measure of efficiency is perhaps $n_{..} V(\hat{P}_{HJ}) / [(n_{..} + 1) V_{IS}(\hat{P}_{IS})]$, or about 24 for this example.

Hoel and Jennrich⁵ gave three estimates \hat{P}_{HJ} for these data, each based on $K = 1,000$ replications: 0.004, 0.005, and 0.009. Ten values of \hat{P}_{IS} , also based on 1,000 replications each, ranged between .0056 and .0070, with empirical variances between 1.58×10^{-7} and 1.87×10^{-7} .

Application to A-Bomb Survivor Mortality Data

The IS procedure described above has been employed in a survey† of mortality among A-bomb survivors of Hiroshima and Nagasaki. That survey is similar in content and format to the earlier reports of Beebe et al.^{3,6} and the more recent report of Kato and Schull.⁹ In the current report routine analyses of the radiation dose response are based on data for 91,231 survivors who were within 10,000 m of the hypocenter in either city ATB, and for whom a radiation dose can be calculated according to the T65D dosimetry system described by Milton and Shohoji.¹⁸ The data were organized into $I=160$ strata according to city, sex, age at

例 (続き)

表1のデータに適用した Hoel 及び Jennrich⁵ の推定量に関しては, $P_D = 0.0063$ であることを考えて, $V(\hat{P}_{HJ}) = 0.00626K^{-1}$ であることを認めた. IS 推定量については, $\alpha_1 = \alpha_2 = 0.5$ と仮定し,

とする. そのとき,

となる. したがって,

である. その結果, 分散に関しては, IS 推定量 \hat{P}_{IS} の効率は, 単純な推定量 \hat{P}_{HJ} に対して約36になる. \hat{P}_{IS} の場合には反復1回につき $n_{..} + 1 = 3$ 個の乱数の作成が必要であるのに対して, 単純な方法では $n_{..} = 2$ 個が必要であるので, 効率に関するより現実的な基準は, 恐らく $n_{..} V(\hat{P}_{HJ}) / [(n_{..} + 1) V_{IS}(\hat{P}_{IS})]$ であり, すなわちこの例では約24になる.

Hoel 及び Jennrich⁵ は, 各々 $K = 1,000$ 回の反復に基づくこれらのデータに関する3個の推定値 \hat{P}_{HJ} , すなわち 0.004, 0.005, 及び 0.009を示した. また, 各々 1,000回の反復に基づく \hat{P}_{IS} の10個の値は .0056から.0070まで, 実験的分散は 1.58×10^{-7} から 1.87×10^{-7} までの範囲内であった.

原爆被爆者死亡率データへの適用

上記の IS 手法は, 広島・長崎の原爆被爆者の死亡率調査†に用いられている. その調査は内容及び形式において, 以前の Beebe ら^{3,6}の報告及び最近の加藤及び Schull⁹の報告に類似する. 本報告においては, 放射線量反応の通常的な分析は, 両市いずれにおいても原爆時に爆心地から 10,000m 以内におり, その被曝線量を, Milton 及び正法地¹⁸が報告した T65D 線量測定法に従って計算できる 91,231人の被爆者に関するデータに基づいて行った. このデータは, 都市, 性, 被爆時年齢(5種の分類群)及び追跡調査期間(8区間)に応じて $I=160$ 個の層で編成した.

†Currently in preparation. 現在準備中

exposure (five categories) and period of follow-up (eight intervals). Eight radiation dose categories, the same as for the example above (Table 1), were assigned city-specific average doses as scores. The skewness of the distribution of radiation doses in this population is similar to that illustrated in Table 1; the skewness of the standardized statistics $(S_{\text{obs}} - E)/V^{1/2}$ was approximately $5n_{..}^{-1/2}$. A total of 38 causes of death or combinations of causes of death were analyzed separately, with the trend test of Mantel¹ and Tarone² calculated for the entire data set and for each period of follow-up. The analysis was based on person-years at risk.

For each test the normal approximation P_N was calculated. If $P_N < 0.10$ and $n_{..} \leq 6$ then P_D was calculated; if $P_N < 0.10$ and $7 \leq n_{..} \leq 249$, then \hat{P}_{IS} was calculated. The IS estimator in each case was based on $K = 1,000$ replications, and calculated using the subroutines in the Appendix. The random number generator WUNI used in subroutine TRENDMC is of the multiplicative congruential type with $X_{i+1} = 1664525(X_i \text{ mod } 2^{32})$. Stratum selection probabilities were defined by

$$\alpha_i = \begin{cases} 1/I^* & \text{if } n_{i.} > 0 \\ 0 & \text{if } n_{i.} = 0 \end{cases}$$

where I^* is the number of strata with $n_{i.} > 0$. (R_{ij} was positive for all i and j .)

The results are summarized in Figures 3 and 4. P_D was calculated for three tests with $n_{..} > 1$: one with five deaths ($P_N = 1.5 \times 10^{-9}$, $P_D = .001$) and two with three deaths each ($P_N = 5.9 \times 10^{-6}$, $P_D = .010$; and $P_N = 9.6 \times 10^{-5}$, $P_D = .024$). As for the example of multiple myeloma described above, the extreme inaccuracy of P_N for small values of $n_{..}$ is apparent. \hat{P}_{IS} was calculated for 74 tests of trend with $n_{..}$ ranging between 8 and 225. For 14 of these, $\hat{P}_{\text{IS}} = 0.0$, i.e., there were no hits among 1,000 replications; for all 14 tests, $P_N < 1.6 \times 10^{-5}$. The remaining 60 values of \hat{P}_{IS} ranged between 7.3×10^{-5} and 0.1004 (Figure 3). The vertical bars in Figure 3, taken from Beebe et al.,⁶ are the ranges of asymptotic p-values adjusted by fourth-order Edgeworth expansions, as functions of the total number of deaths $n_{..}$, for $P_N = .001, .0025, .005, .01, .025, \text{ and } .05$. The top of each bar corresponds to $n_{..} = 10$, and the bottom to $n_{..} = 200$.

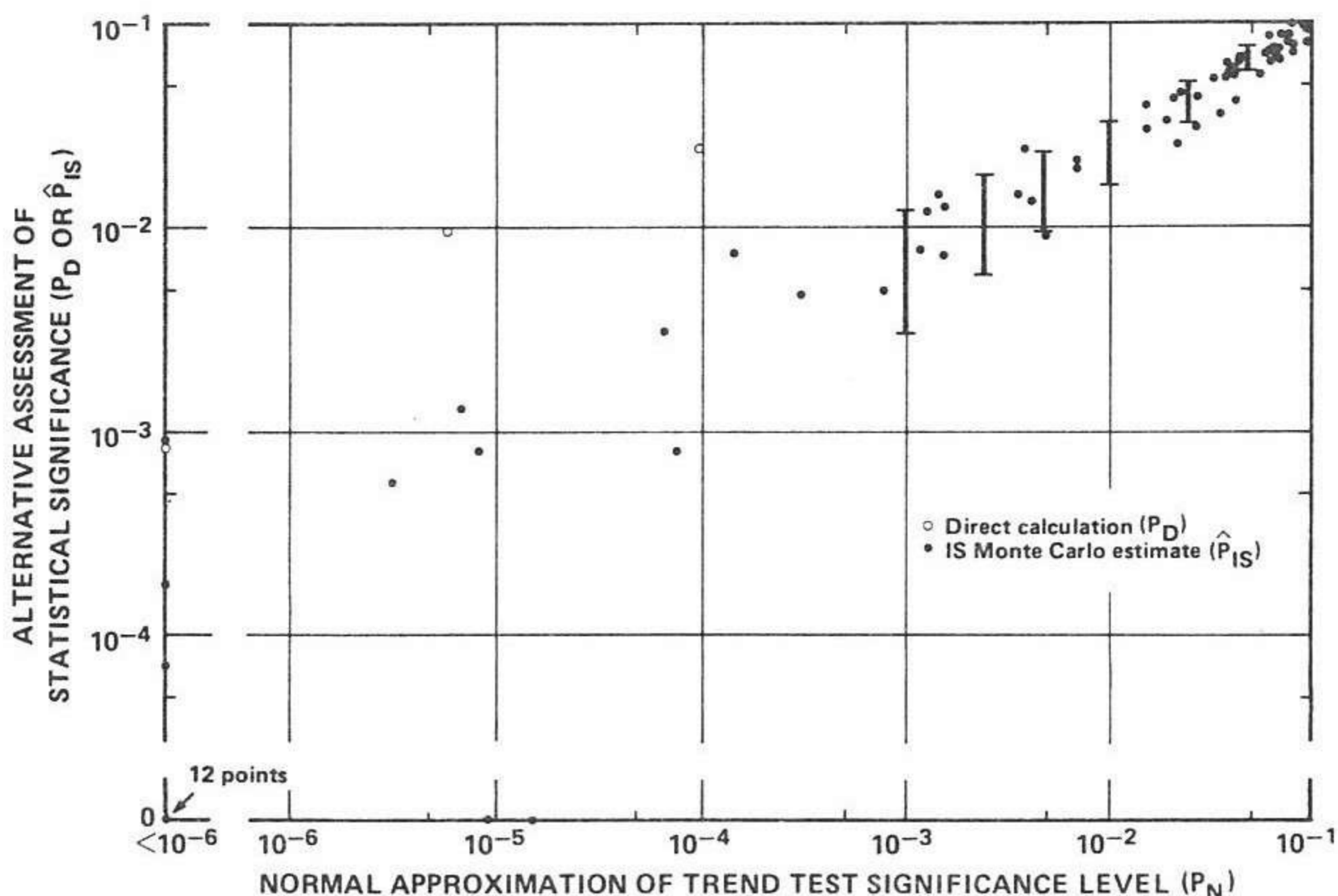
上記の例(表1)の場合と同様に、8種の放射線量群に都市別平均線量をスコアとして割り当てた。この集団における放射線量分布の歪みは表1に示すものと近似しており、正規化統計量 $(S_{\text{obs}} - E)/V^{1/2}$ の歪度は約 $5n_{..}^{-1/2}$ であった。全データ集合及び各追跡調査期間について計算した Mantel¹ 及び Tarone² の傾向検定を用いて、合計38種の死因又は死因群の組み合わせを別々に解析した。解析は at risk の人年に基づいて行った。

各検定においては、正規近似 P_N を計算した。 $P_N < 0.10$ 及び $n_{..} \leq 6$ であるならば、 P_D を計算し、 $P_N < 0.10$ 及び $7 \leq n_{..} \leq 249$ ならば \hat{P}_{IS} を計算した。各場合の IS 推定量は $K = 1,000$ 回の反復を基にし、付録のサブルーチンを用いて計算した。サブルーチン TRENDMC で用いた乱数発生機構 WUNI は、 $X_{i+1} = 1664525(X_i \text{ mod } 2^{32})$ となる乗法合同型である。層選択確率は、

によって定義した。ただし、 I^* は $n_{i.} > 0$ となる層の数である。(すべての i 及び j に対して R_{ij} は正であった)

その結果は図3及び4に要約した。 $n_{..} > 1$ となる三つの検定について P_D を計算した。一つは5例の死亡 ($P_N = 1.5 \times 10^{-9}$, $P_D = .001$)、二つは各々3例の死亡 ($P_N = 5.9 \times 10^{-6}$, $P_D = .010$; 及び $P_N = 9.6 \times 10^{-5}$, $P_D = .024$) を用いたものである。上記の多発性骨髄腫例に関しては、 $n_{..}$ の値が小さいとき P_N が極めて不正確であることは明らかである。 $n_{..}$ が8から225までの範囲内にある74の傾向検定について \hat{P}_{IS} を計算した。これらのうち14については $\hat{P}_{\text{IS}} = 0.0$ 、すなわち1,000回の反復で命中は認められなかった。これら14の検定ではいずれも $P_N < 1.6 \times 10^{-5}$ であった。残り60件の \hat{P}_{IS} 値は 7.3×10^{-5} から 0.1004までの範囲内であった(図3)。図3内の縦線は Beebe ら⁶ からの引用で、 $P_N = .001, .0025, .005, .01, .025$ 及び $.05$ に対して総死亡数 $n_{..}$ の関数として、第4次 Edgeworth 展開により調整した漸近的 p 値の範囲である。各縦線の上端は $n_{..} = 10$ に相当し、下端は $n_{..} = 200$ に相当

FIGURE 3 RESULTS OF 77 TESTS FOR TREND OF CAUSE-SPECIFIC MORTALITY BY RADIATION DOSE: P_D OR \hat{P}_{IS} vs P_N
 図3 死因別死亡率の放射線量別傾向に関する77の検定の結果: P_D 又は \hat{P}_{IS} 対 P_N



The empirical standard errors of the estimates \hat{P}_{IS} are shown in Figure 4. For the tests with $\hat{P}_{IS} > .001$ a reasonably tight upper envelope is obtained by noting that the empirical standard errors were less than $0.04 \hat{P}_{IS}^{1/2}$.

In order to further evaluate the performance of the IS estimator, \hat{P}_{HJ} was calculated for each trend test for which \hat{P}_{IS} was calculated. This was done by using the subroutines in the Appendix after setting the IS cutpoint $c_i = 1$ for all strata. For 20 of the 74 tests, $\hat{P}_{HJ} = 0.0$; these included all 14 tests with $\hat{P}_{IS} = 0.0$, as well as 6 tests with \hat{P}_{IS} ranging between 7.3×10^{-5} and .0009. An estimate of the efficiency of \hat{P}_{IS} relative to \hat{P}_{HJ} was calculated for each test as

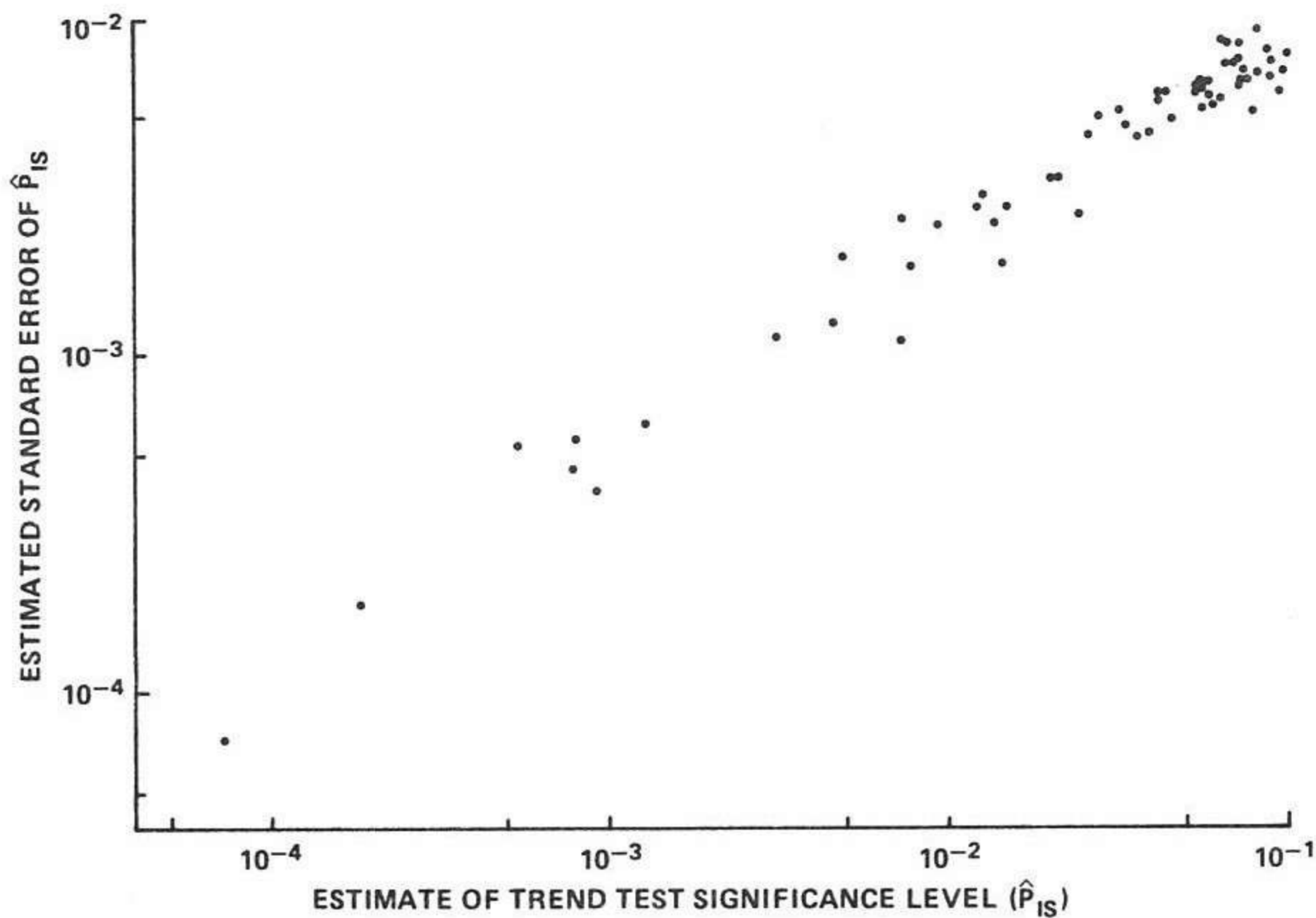
する。推定値 \hat{P}_{IS} の実験的標準誤差を図4に示した。 $\hat{P}_{IS} > .001$ を示す検定については、実験的標準誤差が $0.04 \hat{P}_{IS}^{1/2}$ 未満であったことに留意することによって、適度にきつい上限が得られる。

IS推定量の成果を再評価するために、 \hat{P}_{IS} を計算した各傾向検定について \hat{P}_{HJ} を計算した。これは、すべての層についてIS分割点 $c_i = 1$ を設定した後、付録のサブルーチンを用いて行った。74の検定のうち20件においては $\hat{P}_{HJ} = 0.0$ であったが、この中には、 $\hat{P}_{IS} = 0.0$ を示す14件の検定のすべて、並びに \hat{P}_{IS} が 7.3×10^{-5} から.0009までの範囲内にある6件の検定が含まれていた。各検定ごとに、 \hat{P}_{HJ} に対する \hat{P}_{IS} の効率の推定値を

$$\hat{RE} = \frac{n_{..} \hat{P}_{IS} (1 - \hat{P}_{IS})}{1000(n_{..} + 1) \hat{V}(\hat{P}_{IS})} \times 100\%$$

FIGURE 4 RESULTS OF 60 TESTS FOR TREND OF CAUSE-SPECIFIC MORTALITY BY RADIATION DOSE: ESTIMATED SE OF \hat{P}_{IS} vs \hat{P}_{IS}

図4 死因別死亡率の放射線量別傾向に関する60の検定の結果： \hat{P}_{IS} の推定SE対 \hat{P}_{IS}



where $\hat{V}(\hat{P}_{IS})$ is the empirical variance of \hat{P}_{IS} , and $\hat{P}_{IS}(1-\hat{P}_{IS})/1000$ is an estimate of $V(\hat{P}_{HJ})$ obtained by setting $P_D = \hat{P}_{IS}$ in [7]. The factor $n_{..}/(n_{..}+1)$ accounts for the additional random number required in each MC replication for \hat{P}_{IS} . A stem-and-leaf plot of the values of \hat{RE} for the 60 tests with $\hat{P}_{IS} > 0.0$ is shown in Figure 5. The estimated relative efficiency was greater than 200% for 14 tests, and greater than 150% for 36 tests.

Additional Comments

We have not discussed the choice of the probabilities $\alpha_1, \dots, \alpha_I$ required for Step 1 of the IS procedure. Clearly α_i should be zero for any stratum with

$$n_{i.} = 0 \text{ or } \pi_{i,c_1} + \dots + \pi_{i,J} = 0 \quad [10]$$

Choosing the α_i values in order to minimize $V_{IS}(\hat{P}_{IS})$ requires knowledge of the numerators in the summand of [9], which is not available in real problems. Based on limited experience, we believe that assigning equal probability to all strata for which the condition [10] is false will provide a generally adequate solution.

として計算した。ただし、 $\hat{V}(\hat{P}_{IS})$ は \hat{P}_{IS} の実験的分散であり、 $\hat{P}_{IS}(1-\hat{P}_{IS})/1000$ は[7]において $P_D = \hat{P}_{IS}$ を設定して得た $V(\hat{P}_{HJ})$ の推定値である。係数 $n_{..}/(n_{..}+1)$ は \hat{P}_{IS} の各MC反復において必要な追加乱数の数を補正している。図5は、 $\hat{P}_{IS} > 0.0$ である60件の検定における \hat{RE} 値の茎葉プロットを示したものである。14件の検定においては推定相対効率は200%より大きく、36件の検定においては150%より大きかった。

付記

IS手法の段階1に必要な確率 $\alpha_1, \dots, \alpha_I$ の選択についてはまだ述べていなかった。次の式のいずれの層についても α_i は明らかにゼロである。

$V_{IS}(\hat{P}_{IS})$ を最小限にするために α_i 値を選択するには、[9]の被加数における分子の知識が必要であるが、実際の問題では利用不可能である。限られた経験を基にすると、条件[10]が偽であるすべての層に均等な確率を割り当てれば、概して妥当な解釈が得られるものと考えられる。

FIGURE 5 ESTIMATED EFFICIENCY OF \hat{P}_{IS} RELATIVE TO \hat{P}_{HJ} IN 60 TESTS FOR TREND OF CAUSE-SPECIFIC MORTALITY BY RADIATION DOSE:

STEM-AND-LEAF PLOT OF \hat{R}_E

図5 死因別死亡率の放射線量別傾向に関する60の検定における \hat{P}_{HJ} に対する \hat{P}_{IS} の推定効率: \hat{R}_E の茎葉プロット

STEM (multiples of 100%)	LEAVES (multiples of 10%)
0	899
1	000112222223333344444555566666677777888889
2	003459
3	0258
4	-
5	125
6	-
.	.
.	.
.	.
11	-
12	1

Hastings¹⁹ has sounded a cautionary note about the potential for extreme nonnormality of IS estimators. It is reassuring in this regard that \hat{P}_{IS} is more symmetrically distributed than the binomial estimator \hat{P}_{HJ} , since \hat{P}_{IS} is the mean of K realizations of a discrete random variable which can take as many as 2^I values.

Sound statistical practice demands that if one can choose between two or more unbiased estimators of a parameter, then the estimator with smallest variance must be selected. However the significance of this truism is perhaps unclear when considering MC estimators, since the experimenter has nearly complete control over the number of replications and, therefore, over the variance of the estimator. With modern computing facilities it is relatively easy to conduct very large MC studies, thereby attempting to assure oneself of an estimator with adequate precision. Still, if there is reason to believe that one MC estimator has, for fixed cost, smaller variance than another one, then the first one must be selected.

We have demonstrated above that the IS estimator \hat{P}_{IS} is likely to provide more precise, unbiased estimates of p-values than the simple estimator \hat{P}_{HJ} , when testing for a trend with highly skewed data and a small or moderate number of events. Thus \hat{P}_{IS} is to be preferred over \hat{P}_{HJ} in such situations.

Hastings¹⁹ は IS 推定量の極端な非正規性のポテンシャルについて警告的な意見を述べている。この点、 \hat{P}_{IS} は 2^I 個もの値をとり得る離散確率変数の実現値の K 個の平均であるので、 \hat{P}_{IS} の方が二項推定量 \hat{P}_{HJ} よりも対称的に分布していることは心強い。

正確な統計解析を実施するためには、パラメーターに関して 2 個以上の不偏推定量から選択することができる場合は、最小の分散をもつ推定量を選択しなければならない。しかし、MC 推定量を検討する際には実験者が反復数を、したがってその推定量の分散をもほぼ完全に制御するので、恐らくこの公理の有意性は不明である。現代のコンピューター設備を利用すれば、大規模な MC 研究を実施し、適度に正確な推定量の確認を試みることは比較的容易である。しかし、一定の労力費用のもとで、ある MC 推定量の分散が他の推定量のものより小さいと考える理由がある場合は、前者を選択すべきである。

本報では、大幅な歪みのあるデータ及び少数又は適度な数の事象によって傾向検定を行う場合には、IS 推定量 \hat{P}_{IS} の方が単純な推定量 \hat{P}_{HJ} よりも正確な p 値の不偏推定値をもたらす可能性があることを示した。したがってこのような場合には、 \hat{P}_{HJ} より \hat{P}_{IS} の方が妥当である。

APPENDIX

付 録

SUBROUTINE TRENDMC(ISTRAT,JCLASS,SOBS,KCODE,KREP,I SEED,NTOT,
*P,V,KOUNT)

```

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
C THIS ROUTINE CALCULATES A MONTE CARLO ESTIMATE OF THE OBSERVED
C LEVEL OF SIGNIFICANCE (P-VALUE) OF A LINEAR TREND BASED ON
C STRATIFIED DATA. IT MAY BE USED FOR ANALYSES BASED ON PERSONS
C AT RISK OR ON PERSON-YEARS AT RISK, AND INCORPORATES AN IMPOR-
C TANCE SAMPLING TECHNIQUE.
C
C INPUT DATA
C THROUGH LABELLED COMMON/A/:
C   R(I,J) = NUMBER OF PERSONS OR PERSON-YEARS AT RISK IN
C           STRATUM I AND EXPOSURE GROUP J
C   N(I)   = TOTAL NUMBER OF DEATHS IN STRATUM I (<= 250)
C   S(I,J) = SCORE ASSIGNED TO EXPOSURE GROUP J IN STRATUM I
C   JCUT(I) = IMPORTANCE SAMPLING CUTPOINT FOR STRATUM I
C           (<= JCLASS FOR EACH I)
C   IND(I) = 0 IF STRATUM I IS EXCLUDED FROM TEST OF TREND
C           = 1 IF STRATUM I IS INCLUDED IN TEST OF TREND
C THROUGH CALL STATEMENT PARAMETERS:
C   ISTRAT = NUMBER OF STRATA (<= 250)
C   JCLASS = NUMBER OF EXPOSURE GROUPS (<= 20)
C   SOBS   = OBSERVED VALUE OF TEST STATISTIC, I.E. SUM OF
C           EXPOSURE GROUP SCORES OF OBSERVED DEATHS/CASES
C   KCODE  = 1 FOR MULTINOMIAL SAMPLING (PERSON-YEARS AT RISK);
C           2 FOR HYPERGEOMETRIC SAMPLING (PERSONS AT RISK)
C   KREP   = NUMBER OF MONTE CARLO REPLICATIONS (<= 10,000)
C   ISEED  = SEED FOR RANDOM NUMBER GENERATION
C
C OUTPUT
C   KCODE = -1 IF INVALID VALUE OF ISTRAT, JCLASS, KCODE, OR
C           KREP, OR INVALID ELEMENT OF JCUT OR IND;
C           -2 IF N(1) + ... + N(ISTRAT) < 1 OR >250;
C           -3 IF NO INCLUDED STRATUM HAS BOTH
C           N(I) > 0 AND R(I,JCUT(I))+...+R(I,JCLASS) > 0
C   NTOT  = TOTAL NUMBER OF DEATHS
C   P     = ESTIMATE OF OBSERVED LEVEL OF SIGNIFICANCE
C           (P-VALUE)
C   V     = ESTIMATED VARIANCE OF P
C   KOUNT = NUMBER OF HITS, I.E. REPLICATIONS FOR WHICH PSEUDO-
C           RANDOM REALIZATION, ST, OF TEST STATISTIC SATISFIED
C           ST >= SOBS
C
C EXTERNAL CALLS AND FUNCTIONS
C   WUNI   : RANDOM NUMBER GENERATOR; CALLED IN STEPS 1,
C           2.2, AND 2.3.2
C   DISCRETE : AUXILIARY ROUTINE TO CONVERT CONTINUOUS VARIATE
C           TO ARBITRARY DISCRETE VARIATE
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
DIMENSION RCUM(250,20),RL(250),RCUM1(20),A(250),C(250)
COMMON/A/ R(250,20),N(250),S(250,20),JCUT(250),IND(250)
KOUNT=0
USUM=0.0
UUSUM=0.0

```

(Continue 続<)

APPENDIX (Continued) 付 録 (続き)

```

C
C TEST VALIDITY OF INPUT PARAMETERS AND OF ARRAYS N, JCUT, AND IND
C
  IF(ISTRAT.LT.1.OR.ISTRAT.GT.250) GO TO 20
  IF(JCLASS.LT.1.OR.JCLASS.GT.20) GO TO 20
  IF(KCODE.NE.1.AND.KCODE.NE.2) GO TO 20
  IF(KREP.LT.1.OR.KREP.GT.10000) GO TO 20
  INDTEST=0
  DO 10 I=1,ISTRAT
  IF(N(I).LT.0.OR.N(I).GT.250) GO TO 20
  IF(IND(I).LT.0.OR.IND(I).GT.1) GO TO 20
  IF(N(I).GT.0.AND.IND(I).EQ.1.AND.
+ (JCUT(I).LT.1.OR.JCUT(I).GT.JCLASS)) GO TO 20
  INDTEST=INDTEST+IND(I)
10 CONTINUE
  IF(INDTEST.EQ.0) GO TO 20
  GO TO 30
20 KCODE=-1
  RETURN

C
C COMPUTE NTOT = TOTAL NUMBER OF DEATHS, AND PROCEED ONLY IF
C   1 <= NTOT <= 250
C
30 NTOT=0
  DO 40 I=1,ISTRAT
40 NTOT=NTOT+N(I)*IND(I)
  IF(NTOT.LT.1.OR.NTOT.GT.250) THEN
    KCODE=-2
    RETURN
  END IF

C
C COMPUTE ARRAYS RCUM (STRATUM-SPECIFIC PYR OR PERSONS AT RISK,
C CUMULATIVE OVER EXPOSURE GROUPS) AND RL (TOTAL PYR OR PERSONS
C AT RISK IN EXPOSURE GROUPS JCUT(I),...,JCLASS)
C
  DO 60 I=1,ISTRAT
  IF(N(I).EQ.0.OR.IND(I).EQ.0) GO TO 60
  RCUM(I,1)=R(I,1)
  DO 50 J=2,JCLASS
50 RCUM(I,J)=RCUM(I,J-1)+R(I,J)
  IF(JCUT(I).EQ.1) THEN
    RL(I)=RCUM(I,JCLASS)
  ELSE
    RL(I)=RCUM(I,JCLASS)-RCUM(I,JCUT(I)-1)
  END IF
60 CONTINUE

C
C COMPUTE INDIVIDUAL AND CUMULATIVE STRATUM-SELECTION PROBABILITIES
C A(I) AND C(I), I=1,...,ISTRAT
C
  ISTR1=0
  ISTR2=0
  DO 70 I=1,ISTRAT
  IF(RL(I).GT.0.0.AND.N(I).GT.0.AND.IND(I).EQ.1) THEN
    ISTR1=ISTR1+1
    ISTR2=I

```

(Continue 続く)

APPENDIX (Continued) 付 録 (続き)

```

      A(I)=1.0
    ELSE
      A(I)=0.0
    END IF
70  CONTINUE
    IF(ISTR2.EQ.0) THEN
      KCODE=-3
      RETURN
    END IF
    F=1.0/FLOAT(ISTR1)
    A(I)=A(I)*F
    C(I)=A(I)
    IF(ISTRAT.GE.2) THEN
      DO 80 I=2,ISTRAT
        A(I)=A(I)*F
80    C(I)=C(I-1)+A(I)
      END IF
C
C  PERFORM KREP PSEUDORANDOM REPLICATIONS
C
      DO 130 K=1,KREP
        ST=0.0
        DENOM=0.0
C
C  1. SELECT AT RANDOM A STRATUM, LSTRAT, FROM AMONG 1,2,...,ISTRAT
C
        RAND=WUNI(ISEED)
        CALL DISCRETE(RAND,C,1,ISTR2,LSTRAT)
C
C  2. BEGIN LOOP THROUGH STRATA I = 1,...,ISTRAT
C
        DO 120 I=1,ISTRAT
          IF(N(I).LT.1.OR.IND(I).EQ.0) GO TO 120
C
C  2.1 INITIALIZE RL1 AND ARRAY RCUM1 FOR THE CURRENT STRATUM
C
          RL1=RL(I)
          DO 90 J=1,JCLASS
90    RCUM1(J)=RCUM(I,J)
C
C  2.2 GENERATE EXPOSURE CATEGORY FOR FIRST DEATH OF CURRENT STRATUM
C      (IF CURRENT STRATUM WAS SELECTED IN STEP 1, THEN FIRST DEATH
C      WILL BE SELECTED FROM EXPOSURE GROUPS JCUT(1),...,JCLASS)
C
          RAND=WUNI(ISEED)
          IF(I.EQ.LSTRAT) THEN
            RAND=RCUM1(JCUT(I)-1)+RL1*RAND
            CALL DISCRETE(RAND,RCUM1,JCUT(I),JCLASS,JJ)
            DENOM=DENOM+A(I)*RCUM1(JCLASS)/RL1
          ELSE
            RAND=RCUM1(JCLASS)*RAND
            CALL DISCRETE(RAND,RCUM1,1,JCLASS,JJ)
            IF(JJ.GE.JCUT(I)) DENOM=DENOM+A(I)*RCUM1(JCLASS)/RL1
          END IF
          ST=ST+S(I,JJ)
C

```

(Continue 続く)

APPENDIX (Continued) 付 録 (続き)

```

C   2.3 IF NEEDED, GENERATE EXPOSURE CATEGORIES FOR DEATHS 2,...,N(I)
C   OF CURRENT STRATUM
C
C   IF(N(I).EQ.1) GO TO 120
C   DO 110 N1=2,N(I)
C
C   2.3.1 IF SAMPLING FROM HYPERGEOMETRIC, UPDATE RL1 AND RCUM1
C
C   IF(KCODE.EQ.2) THEN
C     DO 100 J1=JJ,JCLASS
100   RCUM1(J1)=RCUM1(J1)-1.0
C     IF(JJ.GE.JCUT(I)) RL1=RL1-1.0
C     END IF
C
C   2.3.2 GENERATE EXPOSURE CATEGORY FOR DEATH NO. N1
C
C   RAND=WUNI(ISEED)
C   RAND=RCUM1(JCLASS)*RAND
C   CALL DISCRETE(RAND,RCUM1,1,JCLASS,JJ)
C   ST=ST+S(1,JJ)
110 CONTINUE
120 CONTINUE
C
C   COMPUTE CONTRIBUTION OF THIS REPLICATION TO USUM AND UUSUM
C
C   U=0.0
C   IF(ST.GE.SOBS) THEN
C     U=1.0/DENOM
C     KOUNT=KOUNT+1
C   END IF
C   USUM=USUM+U
C   UUSUM=UUSUM+U*U
130 CONTINUE
C
C   COMPUTE FINAL ESTIMATE OF P-VALUE AND ITS ESTIMATED VARIANCE
C
C   F=FLOAT(KREP)
C   P=USUM/F
C   V=(UUSUM-USUM*USUM/F)/((F-1.0)*F)
C   RETURN
C   END

```

(Continue 続く)

APPENDIX (Continued) 付 録 (続き)

```

SUBROUTINE DISCRETE(RAND,CDF,JMIN,JMAX,JGRP)
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
C THIS ROUTINE IS AN AUXILIARY FOR SUBROUTINE TRENDMC. IT CONVERTS C
C RAND, A REALIZATION OF A CONTINUOUS RANDOM VARIABLE, INTO JGRP, A C
C REALIZATION OF A DISCRETE RANDOM VARIABLE. THE VECTOR CDF IS C
C ASSUMED TO SATISFY C
C C
C CDF(JMIN) <= CDF(JMIN+1) <= ... <= CDF(JMAX). C
C C
C JGRP IS DEFINED BY THE CONDITION THAT C
C C
C CDF(JGRP-1) < RAND <= CDF(JGRP). C
C C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
DIMENSION CDF(250)
IF(JMIN.EQ.JMAX) THEN
  JGRP=JMIN
  RETURN
END IF
DO 10 J=JMIN,JMAX-1
IF(RAND.LE.CDF(J)) THEN
  JGRP=J
  RETURN
END IF
10 CONTINUE
JGRP=JMAX
RETURN
END

```

REFERENCES

参考文献

1. MANTEL N: Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *J Amer Stat Assn* 58:690-700, 1963
2. TARONE R: Tests for trend in life table analysis. *Biometrika* 62:679-82, 1975
3. BEEBE GW, KATO H, LAND CE: JNIIH-ABCC Life Span Study, Hiroshima-Nagasaki. Report 5. Mortality and radiation dose, October 1950 - September 1966. ABCC TR 11-70 (*Radiat Res* 48:613-49, 1971)
4. COX DR: Regression analysis and life tables. *J Royal Statist Soc B* 34:187-220, 1972
5. HOEL DG, JENNRICH RI: Life table analysis with small numbers of cases: An example - multiple myeloma in Hiroshima and Nagasaki. *J Statist Comput Simul* 20:311-22, 1985 (RERF TR 9-84)
6. BEEBE GW, KATO H, LAND CE: Life Span Study Report 8. Mortality experience of atomic bomb survivors, 1950-74. ABCC TR 1-77 (*Radiat Res* 75:138-201, 1978)
7. TOKUNAGA M, LAND CE, YAMAMOTO T, ASANO M, TOKUOKA S, EZAKI H, NISHIMORI I: Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-80. RERF TR 15-84
8. KENDALL MG, STUART A: *The Advanced Theory of Statistics*. Vol. 1. London, Charles Griffin and Company, Ltd., 1958
9. KATO H, SCHULL WJ: Studies of the mortality of A-bomb survivors. 7. Mortality, 1950-78: Part 1. Cancer mortality. *Radiat Res* 90:395-432, 1982 (RERF TR 12-80)
10. BARTLETT MS: The spectral analysis of point processes. *J Royal Statist Soc B* 25:264-96, 1963
11. HOPE ACA: A simplified Monte Carlo test procedure. *J Royal Statist Soc B* 30:582-98, 1968
12. MARRIOTT FHC: Barnard's Monte Carlo tests: How many simulations? *Appl Statist* 28:75-7, 1979
13. FOUTZ RV: A method for constructing exact tests from test statistics that have unknown null distributions. *J Statist Comput Simul* 10:187-93, 1980
14. KELLERER AM, CHMELEVSKY D: Small-sample properties of censored-data rank tests. *Biometrics* 39:675-82, 1983
15. HAMMERSLEY JM, HANSCOMB DC: *Monte Carlo Methods*. London, Chapman and Hall, 1964
16. RUBINSTEIN RY: *Simulation and the Monte Carlo Method*. New York, John Wiley and Sons, Inc., 1981
17. KOPECKY KJ: Utilization of the Adult Health Study Data: Resources and methods of past and current analyses. In *Atomic Bomb Survivor Data: Utilization and Analysis*. Ed by R.L. Prentice and D.J. Thompson. Philadelphia, SIAM, 1984. pp 197-218
18. MILTON RC, SHOHOJI T: Tentative 1965 radiation dose estimation for atomic bomb survivors, Hiroshima - Nagasaki. ABCC TR 1-68
19. HASTINGS WK: Variance reduction and nonnormality. *Biometrika* 61:143-9, 1974