ON ESTIMATING MULTINOMIAL PROBABILITIES BY POOLING INCOMPLETE SAMPLES*

CHOOICHIRO ASANO**

(Received Sept. 23, 1963; revised Nov. 13, 1964)

1. Introduction

In the statistical analysis of observations from multinomial distribution it is sometimes necessary to combine, or pool, two or more sets of observations in order to estimate the probabilities involved.

Kastenbaum (1958) considered the pooling of incomplete samples, while Watson (1956), Batschelet (1960), and Geppert (1961) considered the pooling of truncated samples with complete samples. The purpose of this paper is to give a generalized and somewhat simplified treatment of the estimation problem for the multinomial distribution.

First we shall reformulate the problem as considered by the last three authors mentioned above and shall discuss the main properties of the maximum likelihood estimates. Then, in section 3, we generalize the problem, no longer requiring the presence of a complete sample. The main objective will be to find necessary and sufficient conditions for the parameters to be estimable. Finally, in section 4, we study "nested" and "chained" samples. Here explicit formulas for the estimates are available.

2. Formulation of the problem and its maximum likelihood solution

Suppose a multinomial distribution with a finite number of cells, in which individuals are observed. A sample covering all cells is called a complete sample, and we denote the sample space, that is the set of all cells, by Ω_1. We also consider a truncated sample which is observed in a subset Ω_2 of Ω_1.

---

* This study was supported in part by the contract NSF-9968, National Science Foundation; in part by the contract Nonr 2249(05), (Nr 301-579), the Office of Naval Research, Department of the Navy, with the Catholic University of America.

** This paper was written while the author was a Research Associate in the Statistical Laboratory, Department of Mathematics, The Catholic University of America, during the academic year, 1961-62.

Let p_i be the probability in ith cell for a complete sample space Ω_1, where $\sum_{\Omega_1} p_i=1$. Then the conditional probability for a truncated sample is given by $p_i/\sum_{\Omega_2} p_j$ for $i \in \Omega_2$. Let $n_i^{(t)}$ be the frequency for the tth set of observations, $t=1, 2$. Since $\sum_{\Omega_1} p_i=1$, we assume that a special probability p_0 in $\Omega_1 \cap \Omega_2$ is linearly dependent on the other p_i's.

With this notation the likelihood function L_1 is given by

$$L_1=\frac{N_1!\, N_2!}{\prod_{\Omega_1} n_i^{(1)}! \prod_{\Omega_2} n_i^{(2)}!} \prod_{\Omega_1} p_i^{n_i^{(1)}} \prod_{\Omega_2}\left(\frac{p_i}{\sum_{\Omega_2} p_i}\right)^{n_i^{(2)}}, \tag{2.1}$$

where $N_1=\sum_{\Omega_1} n_i^{(1)}$ and $N_2=\sum_{\Omega_2} n_i^{(2)}$.

We then have the following theorems.

THEOREM 1. *The likelihood estimates of the p_i's are given by*

$$\hat{p}_i=n_i^{(1)}/N_1 \qquad \text{for} \quad i \in \Omega_1 \cap \Omega_2^c \tag{2.2}$$

$$\hat{p}_i=(n_i^{(1)}+n_i^{(2)})\Big/N_1\left(1+\frac{N_2}{\sum_{\Omega_2} n_i^{(1)}}\right) \qquad \text{for} \ i \in \Omega_2 . \tag{2.3}$$

This theorem is directly obtained by solving the maximum likelihood equations and is also a special case of the results in Batschelet [1960] and Geppert [1961].

THEOREM 2. *All of the estimates defined by (2.2) and (2.3) are unbiased, consistent and jointly sufficient. And their variances and covariances are as follows:*

$$V\{\hat{p}_i\}=p_i(1-p_i)/N_1 \qquad \text{for} \ i \in \Omega_1 \cap \Omega_2^c , \tag{2.4}$$

$$V\{\hat{p}_i\}=\frac{p_i}{N_1}(1-p_i)+\frac{N_2^2}{N_1^2}\frac{p_i}{\sum_{\Omega_2} p_i}\left(\frac{p_i}{\sum_{\Omega_2} p_i}-1\right) \times\left\{\frac{1}{N_2}-\sum_{n_1=0}^{N_1}\frac{1}{N_1+N_2-n_1}\binom{N_1}{n_1}\Big(\sum_{\Omega_1\cap\Omega_2^c} p_i\Big)^{n_1}\Big(\sum_{\Omega_2} p_i\Big)^{n_2}\right\} \qquad \text{for} \ i \in \Omega_2 . \tag{2.5}$$

$$\mathrm{Cov}\{\hat{p}_i\hat{p}_j\}=-p_ip_j/N_1 \qquad \text{for} \ i\neq j, \quad i, j \in \Omega_1 \cap \Omega_2^c , \tag{2.6}$$

$$\text{(2.7)}\qquad \mathrm{Cov}\{\hat{p}_i\hat{p}_j\} = -\frac{p_ip_j}{N_1} + \frac{N_2p_ip_j}{N_1^2(\sum_{\Omega_2} p_i)^2} \times \left\{1-\sum \frac{N_2}{N_1+N_2-n_1}\binom{N_1}{n_1}\Big(\sum_{\Omega_1\cap\Omega_2^c} p_i\Big)^{n_1}\Big(\sum_{\Omega_2} p_i\Big)^{n_2}\right\}$$

for $i\neq j,\quad i, j\in\Omega_2$,

$$\text{(2.8)}\qquad \mathrm{Cov}\{\hat{p}_i\hat{p}_j\} = -p_ip_j/N_1 \qquad \textit{for } i\in\Omega_2,\quad j\in\Omega_1\cap\Omega_2^c,$$

where $n_1=\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}$.

In general, the superiority of estimates obtained by using combined data is that they have smaller variances and also smaller covariances than estimates obtained by neglecting such combined data. The differences in values of variance and of covariance are expressed by the second term of the right-hand-side of (2.5) and of (2.7), respectively.

PROOF.

(i) Unbiasedness. The unbiasedness of $\hat{p}_i$ for $i\in\Omega_1\cap\Omega_2^c$ follows obviously from (2.1) and (2.2).

Now, in order to show the unbiasedness of $\hat{p}_i$ for $i\in\Omega_2$, let us study the likelihood function L_1 in (2.1), which is based on the product of two separate multinomial distributions; that is, one ordinary multinomial distribution and a conditional multinomial distribution. Using this fact, let us introduce the following function.

$$\text{(2.9)}\qquad L_2 = \frac{N_1!}{\{\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}\}!\,\{\sum_{\Omega_2} n_i^{(1)}\}!}\;\frac{(N_2+\sum_{\Omega_2} n_i^{(1)})!}{\prod_{\Omega_2}((n_i^{(1)}+n_i^{(2)})!)}\;\frac{\{\sum_{\Omega_1\cap\Omega_2^c} p_i\}^{\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}}}{\{\sum_{\Omega_2} p_i\}^{N_2}}\prod_{\Omega_2} p_i^{(n_i^{(1)}+n_i^{(2)})}$$

where

$$\text{(2.10)}\qquad \sum_{\substack{(\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)},\ \sum_{\Omega_2} n_i^{(1)})\\ \sum_{\Omega_1} n_i^{(1)}=N_1}}\ \ \sum_{\substack{(\cdots,\ \sum_{t=1}^{2} n_i^{(t)},\ \cdots)\\ \sum_{\Omega_2}(\sum_t n_i^{(t)})=N_2+\sum_{\Omega_2} n_i^{(1)}}} L_2 = 1$$

which is a modified form of L_1 over different summations and which in effect simplifies the underlying calculation.

Hence the expectation of $\hat{p}_i$ for $i\in\Omega_2$ is expressed by

$$\text{(2.11)}\qquad E\{\hat{p}_i\} = \sum\sum \frac{\sum_{\Omega_2} n_i^{(1)}\cdot\{\sum_{t=1}^{2} n_i^{(t)}\}}{N_1(N_2+\sum_{\Omega_2} n_i^{(1)})} L_2 \qquad \text{for } i\in\Omega_2$$

and the main part concerning summations becomes equivalent to (2.10) in case we put $N_1' \equiv N_1 - 1$ and $\{\sum_{\Omega_2} n_i^{(1)}\}' \equiv \{\sum_{\Omega_2} n_i^{(1)}\} - 1$. Thus we obtain the unbiasedness.

(ii) Variances and covariances. To prove (2.4) and (2.6) we have to calculate the following expectations:

$$E\{\hat{p}_i^2\} = p_i^2 - \frac{p_i^2}{N_1} + \frac{p_i}{N_1} \qquad \text{for } i \in \Omega_1 \cap \Omega_2^c, \tag{2.12}$$

$$E\{\hat{p}_i \hat{p}_j\} = \frac{N_1 - 1}{N_1} p_i p_j \qquad \text{for } i \neq j, \quad i, j \in \Omega_1 \cap \Omega_2^c, \tag{2.13}$$

and we obtain immediately (2.4) and (2.6) since each $\hat{p}_i$ is unbiased.

In order to get (2.5) and (2.7) we have to make use of the function (2.9) instead of (2.1).

$$E\{\hat{p}_i^2\} = \sum\sum \frac{(\sum_{\Omega_2} n_i^{(1)})^2}{N_1^2 (N_2 + \sum_{\Omega_2} n_i^{(1)})^2} \left\{ (\sum_{t=1}^{2} n_i^{(t)})(\sum_{t=1}^{2} n_i^{(t)} - 1) + (\sum_{t=1}^{2} n_i^{(t)}) \right\} L_2 \tag{2.14}$$

$$= \left(1 - \frac{1}{N_1}\right) p_i^2 + \frac{N_2}{N_1} p_i^2 (\sum_{\Omega_2} p_i)^{-1} \sum \frac{1}{N_2 + n_2} \frac{(N_1 - 1)!}{n_1!\,(n_2 - 1)!} (\sum_{\Omega_1 \cap \Omega_2^c} p_i)^{n_1} (\sum_{\Omega_2} p_i)^{n_2 - 1}$$

$$+ \frac{p_i}{N_1} - \frac{N_2}{N_1} p_i \sum \frac{1}{N_2 + n_2} \frac{(N_1 - 1)!}{n_1!\,(n_2 - 1)!} (\sum_{\Omega_1 \cap \Omega_2^c} p_i)^{n_1} (\sum_{\Omega_2} p_i)^{n_2 - 1} \qquad \text{for } i \in \Omega_2.$$

$$E\{\hat{p}_i \hat{p}_j\} = \sum\sum \frac{(\sum_{\Omega_2} n_i^{(1)})^2 (\sum_{t=1}^{2} n_i^{(t)})(\sum_{t=1}^{2} n_j^{(t)})}{N_1^2 (N_2 + \sum_{\Omega_2} n_i^{(1)})^2} L_2 \tag{2.15}$$

$$= \left(1 - \frac{1}{N_1}\right) p_i p_j + \frac{N_2}{N_1^2} \frac{p_i p_j}{(\sum p_i)^2}$$

$$\times \left\{ 1 - \sum \frac{N_2}{N_1 + N_2 - n_1} \binom{N_1}{n_1} (\sum_{\Omega_1 \cap \Omega_2^c} p_i)^{n_1} (\sum_{\Omega_2} p_i)^{n_2} \right\}$$

and $V\{\hat{p}_i\} = E\{\hat{p}_i^2\} - p_i^2$, $\mathrm{Cov}\{\hat{p}_i \hat{p}_j\} = E\{\hat{p}_i \hat{p}_j\} - p_i p_j$, for $i \neq j$, $i, j \in \Omega_2$.

The result given by (2.8) is obtained by applying both formulas (2.1) and (2.9) in the following way:

$$E\{\hat{p}_i \hat{p}_j\} = \sum_{\Omega_1, \Omega_2} \frac{n_j^{(1)} (n_i^{(1)} + n_j^{(2)}) (\sum_{\Omega_2} n_i^{(1)})}{N_1^2 (N_2 + \sum_{\Omega_2} n_i^{(1)})} L_2 \tag{2.16}$$

$$=p_j\sum\sum\frac{(n_i^{(1)}+n_j^{(2)})(\sum_{\Omega_2}n_i^{(1)})}{N_1(N_2+\sum_{\Omega_2}n_i^{(1)})}L_2',$$

where L_2' denotes an expression obtained from L_2 by replacing N_1 with N_1-1 and $\sum n_i^{(1)}$ with $\sum_{\Omega_1\cap\Omega_2^c}n_i^{(1)}-1$, and then

$$E\{\hat{p}_i\hat{p}_j\}=\frac{N_1-1}{N_1}p_ip_j\sum\sum L_2''=\frac{N_1-1}{N_1}p_ip_j \tag{2.17}$$

for $i\in\Omega_2,\quad j\in\Omega_1\cap\Omega_2^c$,

where L_2'' is obtained from L_2 by replacing N_i, $\sum_{\Omega_1\cap\Omega_2^c}n_i^{(1)}$ and $\sum_{\Omega_2}n_i^{(1)}$ with N_1-2, $\sum_{\Omega_1\cap\Omega_2^c}n_i^{(1)}-1$ and $\sum_{\Omega_2}n_i^{(1)}-1$, respectively. Thus we obtain (2.8) from $\text{Cov}\{\hat{p}_i\hat{p}_j\}=E\{\hat{p}_i\hat{p}_j\}-p_ip_j$.

(iii) Consistency and sufficiency. Consistency is evident since the estimates are unbiased and the variances and covariances tend to zero if $N\to\infty$.

The likelihood function L_1 can be expressed as a function of the estimates $\hat{p}_i$'s and the parameters p_i's only. This will prove the sufficiency of the estimates.

Finally, the set of sufficient statistics given by (2.1) can be mapped to the present estimates in a unique way. This completes the proof of theorem 2.

Now we want to know whether or not the estimates, $\hat{p}_i$'s, are efficient. For this purpose we need the information matrix in R. A. Fisher's sense.

THEOREM 3. *The information matrix of our estimates is given by the following elements.*

$$\sigma^{ii}=N_1\left(\frac{1}{p_i}-\frac{1}{p_0}\right)\qquad \textit{for } i\neq 0,\ i\in\Omega_1\cap\Omega_2^c, \tag{2.18}$$

$$\sigma^{ii}=\frac{N_1+N_2}{p_i}-\frac{N_1}{p_0}-\frac{N_2}{(\sum_{\Omega_2}p_i)^2} \tag{2.19}$$

for $i\neq j,\ i,j\neq 0,\ i,j\in\Omega_1\cap\Omega_2^c$,

$$\sigma^{ii}=\frac{N_1}{p_0}\qquad \textit{for } i\in\Omega_2, \tag{2.20}$$

$$\sigma^{ij}=\frac{N_1}{p_0}-\frac{N_2}{(\sum_{\Omega_2}p_i)^2}\qquad \textit{for } i\neq j,\ i,j\in\Omega_2, \tag{2.21}$$

$$\sigma^{ij}=\sigma^{ji}=\frac{N_1}{p_0} \qquad \textit{for } i\neq 0,\ i\in\Omega_1\cap\Omega_2^c,\ j\in\Omega_2\,. \tag{2.22}$$

PROOF. σ^{ij} is defined to be $\sigma^{ij}=-E\{\partial^2\log L/\partial p_i\partial p_j\}$. The calculations are straightforward and need no special explanation. However, we replaced the probability p_0 of a special cell in $\Omega_1\cap\Omega_2^c$ by $1-\sum' p_i$ where $\sum'$ denotes the sum over all p_i's except p_0.

Now we are going to compare the information matrix with the inverse variance-covariance matrix. Formulas (2.4) through (2.8) provide the elements of the variance-covariance matrix related to all $\hat{p}_i$'s ignoring $\hat{p}_0$. The inversion of this matrix is laborious but it is easy to see that the elements do not correspond to the elements of the information matrix. From this we conclude that the estimates are not efficient.

3. Generalization of the problem

Assume that we are given t samples which are truncated in different ways. To each sample we associate a sample space Ω_i, $i=1, 2, \cdots, t$. For each Ω_i we denote the probability of the jth cell by $p_j^{(i)}$, $j=1, 2, \cdots k_i$. In a similar way $n_j^{(i)}$ denotes the observed frequency of this cell. Values of $n_j^{(i)}$'s are stochastically independent for different values of i. We assume that each sample space has at least one cell in common with some other sample space. Furthermore, without loss of generality, we assume $\Omega_i\neq\Omega_{i'}$ if $i\neq i'$. If this were not true we could simply combine the i'th observations into a single space Ω_i.

We consider a decomposition $\{\omega_u\}$ of $\bigcup_{i=1}^{t}\Omega_i$ which is such that $\omega_u\cap\omega_{u'}=\phi$ if $u\neq u'$. Such a decomposition is $\{\omega_u\}=\{\bigcap_{i=1}^{t}\Omega_i^{r_i}\}$, where $r_i=0, 1$, subject to the condition $1\leqq\sum_{i=1}^{t} r_i\leqq t$, and where $\Omega_i^1\equiv\Omega_i$ and $\Omega_i^0\equiv\Omega_i^c$. Thus, for $t=2$, $\{\omega_u\}=\{\Omega_1\cap\Omega_2,\ \Omega_1^c\cap\Omega_2,\ \Omega_1\cap\Omega_2^c\}$. Under these conditions, number of ω_u's, M, is such that $t+1\leqq M\leqq 2^t-1$.

We thus have

$$\sum_{u=1}^{M}\omega_u=\sum_{\substack{r_i=0,\,1\\ 1\leqq\Sigma r_i\leqq t}}\ \bigcap_{i=1}^{t}\Omega_i^{r_i}=\bigcup_{i=1}^{t}\Omega_i(\equiv\Omega)\,, \tag{3.1}$$

with $\omega_u\cap\omega_{u'}=\phi$ when $u\neq u'$.

Now we wish to indicate explicitly what value u takes for each ω_u that falls in Ω_i for any i, $i=1, 2, \cdots, t$. To do this, we can show that each Ω_i will contain at most 2^{t-1} such ω_u's so that

$$u=1, 2, \cdots, h_1\leqq 2^{t-1} \qquad \text{for } i=1\,, \tag{3.2}$$

while

(3.3) $$u=h_1+1,\ h_1+2,\ \cdots,\ h_1+h_2\leqq h_1+2^{t-2} \qquad \text{for } i=2\,.$$

In general, for $i=1, 2, \cdots, t$,

(3.4) $$u=\sum_{k=1}^{i-1} h_k+1,\ \sum_{k=1}^{i-1} h_k+2,\ \cdots,\ \sum_{k=1}^{i-1} h_k+h_i\,,$$

where $h_k\leqq 2^{t-k}$ and $h_0=0$.

Now let p_{uv}'s, $v=1, 2, \cdots, l_u$, be the probabilities associated with the cells in an arbitrary ω_u of our decomposition. Let $P_u=\sum_{v=1}^{l_u} p_{uv}$, where $\sum_{u=1}^{M} P_u=1$. Hence, we have $\sum_{u=1}^{M} l_u-1$ parameters to be estimated. Since maximum likelihood estimation is invariant, we may, and indeed shall, assume that P_1 is a linearly dependent parameter. Here we denote

(3.5) $$n_{uv}^{(i)}\equiv(n_j^{(i)}\,|\,\Omega_i\supset\omega_u\ni j)\,,\quad n_{u\cdot}^{(i)}\equiv\sum_{v=1}^{l_u} n_{uv}^{(i)}\,,\quad N_i\equiv\sum_{j=1}^{k_i} n_j^{(i)},$$

where the subscript of v of $n_{uv}^{(i)}$ is denoted the newly-defined number of cell in ω_u instead of the j of $n_j^{(i)}$.

Thus we obtain under the foregoing conditions the following general result.

THEOREM 4. *Estimates of the p_{uv}'s are obtained from the combined sample as follows*:

(3.6) $$\hat{p}_{uv}=\frac{\sum\limits_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} n_{uv}^{(i)}}{\sum\limits_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} n_{u\cdot}^{(i)}}\hat{P}_u$$

where $u=1, 2, \cdots, M\leqq 2^t-1$ and $v=1, 2, \cdots, l_u$, and where

(3.7) $$\hat{P}_1=1-\sum_{u=2}^{M}\hat{P}_u$$

and $\hat{P}_2, \hat{P}_3, \cdots, \hat{P}_M$ are obtained as the solutions of the following: $M-1$ simultaneously equations:

(3.8) $$\sum_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}}\left[\frac{n_{u\cdot}^{(i)}}{P_u}-\frac{N_i}{\sum\limits_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} P_u}\right]=0 \qquad \textit{for } u=1, 2, \cdots, M.$$

PROOF. The likelihood function L_1 is given by

(3.9) $$L_1=\frac{\prod\limits_{i=1}^{t} N_i\,!}{\prod\limits_{i=1}^{t}\prod\limits_{\{j|\Omega_i\ni j\}} n_j^{(i)}\,!}\prod_{i=1}^{t}\prod_{\{j|\Omega_i\ni j\}}\left[\frac{p_j^{(i)}}{\sum\limits_{\{j|\Omega_i\ni j\}} p_j^{(i)}}\right]^{n_j^{(i)}}.$$

Using the notation introduced in (3.5), (3.9) can be rewritten as follows:

$$L_1=\text{const.}\prod_{i=1}^{t}\prod_{\{u|\Omega_i\supset\omega_u\}}\prod_{v=1}^{l_u}\left[\frac{p_{uv}}{\sum_{\{u|\Omega_i\supset\omega_u\}}P_u}\right]^{n_{uv}^{(i)}}. \tag{3.10}$$

We then obtain

$$\log L_1=\log(\text{const.})+\sum_{i=1}^{t}\left[\sum_{\{u|\Omega_i\supset\omega_u\}}\sum_{v=1}^{l_u}n_{uv}^{(i)}\{\log p_{uv}-\log(\sum_{\{u|\Omega_i\supset\omega_u\}}P_u)\}\right] \tag{3.11}$$

and

$$\frac{\partial\log L_1}{\partial p_{uv}}=\sum_{\{i|\Omega_i\supset\omega_u\}}\left[\frac{n_{uv}^{(i)}}{p_{uv}}-\frac{N_i}{\sum_{\{u|\Omega_i\supset\omega_u\}}P_u}\right]=0. \tag{3.12}$$

The likelihood function for P_u's is

$$L_2=\frac{\prod_{i=1}^{t}N_i!}{\prod_{i=1}^{t}\prod_{\{u|\Omega_i\supset\omega_u\}}n_{u\cdot}^{(i)}!}\prod_{i=1}^{t}\prod_{\{u|\Omega_i\supset\omega_u\}}\left[\frac{P_u}{\sum_{\{u|\Omega_i\supset\omega_u\}}P_u}\right]^{n_{u\cdot}^{(i)}} \tag{3.13}$$

from which we get

$$\frac{\partial\log L_2}{\partial P_u}=\sum_{\{i|\Omega_i\supset\omega_u\}}\left[\frac{n_{u\cdot}^{(i)}}{P_u}-\frac{N_i}{\sum_{\{u|\Omega_i\supset\omega_u\}}P_u}\right]=0 \tag{3.14}$$

for $u=2, 3, \cdots, M$,

which is the same as (3.8).

Combining (3.12) and (3.14), we obtain the simple formula (3.6) given in theorem 4. (Q. E. D.)

In general, as t increases, the solution of (3.6) becomes more complicated. However, the estimates, $\hat{P}_u$'s, $u=2, 3, \cdots, M$, can be obtained by iterative solutions of maximum likelihood equations as indicated below.

Let the solutions be $\hat{P}_2, \hat{P}_3, \cdots, \hat{P}_M$. Let $\tilde{P}_{21}, \tilde{P}_{31}, \cdots, \tilde{P}_{M1}$ be approximations to $\hat{P}_2, \hat{P}_3, \cdots, \hat{P}_M$, respectively, obtained in any manner. An easy procedure for obtaining such approximations is to neglect any combination of observations and take as our approximations the frequency ratios of ω_u in Ω_i.

Now by the Taylor-Maclaurin expansion, to the first order of small quantities, improved values for the estimates will be

$$\tilde{P}_{22}=\tilde{P}_{21}+\delta\tilde{P}_{21},\ \tilde{P}_{32}=\tilde{P}_{31}+\delta\tilde{P}_{31},\ \cdots,\ \tilde{P}_{M2}=\tilde{P}_{M1}+\delta\tilde{P}_{M1}, \tag{3.15}$$

where the increments $\delta\tilde{P}_{21}, \delta\tilde{P}_{31}, \cdots, \delta\tilde{P}_{M1}$ are the solutions of

$$\frac{\partial \log L_1}{\partial \tilde{P}_{u1}} + \delta \tilde{P}_{u1} \frac{\partial^2 \log L_1}{(\partial \tilde{P}_{u1})^2} + \sum_{\substack{j=2 \\ j \neq u}}^{M} \delta \tilde{P}_{j1} \frac{\partial^2 \log L_1}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = 0 \tag{3.16}$$

for $u \neq j$, $u=2, 3, \cdots, M$.

To indicate the first estimates of p_u and p_j, we write $\tilde{p}_{u1}$ and $\tilde{p}_{j1}$ after differentiation, and further each term of (3.16) is given as follows:

$$\frac{\partial \log L_1}{\partial \tilde{P}_{u1}} = \sum_{\{i \mid \Omega_i \supset \omega_u\}} \left[\frac{n_{u\cdot}^{(i)}}{\tilde{P}_{u1}} - \frac{N_i}{\sum \tilde{P}_{u1}} \right], \tag{3.17}$$

$$\frac{\partial^2 \log L_1}{(\partial \tilde{P}_{u1})^2} = - \sum_{\{i \mid \Omega_i \supset \omega_u\}} \left[\frac{n_{u\cdot}^{(i)}}{\tilde{P}_{u1}^2} - \frac{N_i}{(\sum \tilde{P}_{u1})^2} \right], \tag{3.18}$$

$$\frac{\partial^2 \log L_1}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = - \sum_{\{i \mid \Omega_i \ni u,\, j\}} \frac{N_i}{(\sum \tilde{P}_{u1})^2} \qquad \text{for } j \in (\Omega_i \ni u), \tag{3.19}$$

$$\frac{\partial^2 \log L_1}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = 0 \qquad \text{for } j \notin (\Omega_i \ni u). \tag{3.20}$$

An iterative process may be based on the $M-1$ equations given in (3.16), replacing $\tilde{P}_{u1}$, $\tilde{P}_{j1}$ by $\tilde{P}_{u2}$, $\tilde{P}_{j2}$ and solving for increments $\delta \tilde{P}_{u2}$, $\delta \tilde{P}_{j2}$, and so on, until a satisfactorily close approach to $\tilde{P}_u$, $\tilde{P}_j$ is achieved. Thus we obtain the individual estimates of p_{uv} by applying (3.5).

THEOREM 5. *A necessary and sufficient condition for the p_i to be estimable is that every sample space has at least one cell in common with some other of the sample spaces involved.* (*In other words, it is the condition that each sample space somewhere overlaps another sample space.*)

PROOF. The reader should feel that this theorem is natural. To prove the necessity, suppose that t sample spaces are given and are separated to s connected spaces in a sense of overlap. Let a total number of their cells be k. Then, since the sum of the probabilities in cells becomes one, $k-1$ probabilities must be linearly independent. Suppose, on the other hand, we are given s parameter spaces, each of which corresponds to one of the s connected sample spaces. Then the number of linearly independent parameters in each parameter space is less by one than the number of cells in the space. Hence, there exist, at most, $k-s$ linearly independent parameters altogether. So we have $s-1$ degrees of freedom to estimate the probabilities and cannot obtain uniquely the k parameters. From this, we conclude that s should be 1.

To prove the sufficiency, let us use the above notation. If $s=1$, then we actually showed the possibility of estimating all p_{uv}'s in theorem

4. From this, we can conclude to the sufficiency of the estimates. (Q.E.D.)

4. Some particular generalizations

As corollaries to theorem 4, we now give explicit estimates of the p_{uv}'s for some special cases.

4.1 Nested case. Let us consider the case in which the observation space is such that

$$\Omega \equiv \Omega_1 \supset \Omega_2 \supset \cdots \supset \Omega_t .$$

In this case we shall refer to the sample spaces as " nested ". This is a generalization of the case considered in section 2.

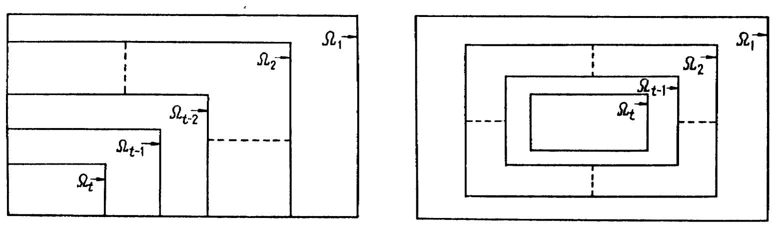

Fig. 1. The combination of contingency tables

In this case, the decomposition $\{\omega_u\}$ of $\bigcup_{i=1}^{t} \Omega_i$ can be defined by

$$\sum_{u=1}^{t} \omega_u = \bigcup_{i=1}^{t} (\Omega_i \cap \Omega_{i+1}^c) = \Omega , \tag{4.1}$$

where $\Omega_{t+1}^c \equiv \Omega$ and the values of u and i are the same. Let p_{uv}'s, $v = 1, 2, \cdots, l_u$, denote the probabilities associated with the cells in ω_u. Then we obtain the following result.

COROLLARY 1. *For all $v \in \omega_u$, $u=1, 2, \cdots, t$, the following explicit expression for the maximum likelihood estimate of p_{uv} is obtained from the combined samples:*

$$\hat{p}_{uv} = \frac{n_{uv}}{\sum_{j=1}^{t} \dfrac{n_v^{(j)}}{\{1 - \sum_{r=0}^{j-1} \sum_{\omega_r} \hat{p}_{rv}\}}} \tag{4.2}$$

where $\omega_0 \equiv \phi$, $p_{0v} = 0$, and n_{uv} and $n_v^{(j)}$ are defined in section 4. Furthermore, these estimates are unbiased, consistent and jointly sufficient.

Indeed, we can obtain the explicit estimates of p_{uv}'s by applying successively the above formula (4.2) starting with $u=1$.

For example, in case $t=3$, we get

$$\hat{p}_{1v_1}=\frac{n_{1v_1}}{N_1} \qquad \text{for } v_1 \in \omega_1 , \tag{4.3}$$

$$\hat{p}_{2v_2}=\frac{n_{2v_2}}{N_1\left\{1+\dfrac{n^{(2)}_{v_2}}{N_1\left\{1-\dfrac{\sum\limits_{\omega_1} n^{(1)}_{v_1}}{N_1}\right\}}\right\}} \qquad \text{for } v_2 \in \omega_2 , \tag{4.4}$$

and

(4.5)

$$\hat{p}_{3v_3}=\frac{n_{3v_3}}{N_1\left\{1+\dfrac{n^{(2)}_{v_3}}{N_1\left\{1-\sum\limits_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1}\right\}}+\dfrac{n^{(3)}_{v_3}}{N_1\left\{1-\sum\limits_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1}-\sum\limits_{\omega_2}\dfrac{n_{2v_2}}{N_1\left\{1+\dfrac{n^{(2)}_{v_2}}{N_1\left\{1-\sum\limits_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1}\right\}}\right\}}\right\}}\right\}}$$

for $v_3 \in \omega_3$.

The latter half of this corollary is proved in quite the same manner as was theorem 2. However, the following functions are applied in place of L_1 and L_2 as given by (2.1) and (2.9) respectively:

$$L_1=\prod_{j=1}^{t}\left[\frac{N_j\,!}{\prod\limits_{\Omega_j \ni i} n^{(j)}_i\,!}\prod_{\Omega_j \ni i}\left(\frac{p_i}{\sum\limits_{\Omega_j} p_i}\right)^{n^{(j)}_i}\right] , \tag{4.6}$$

$$L_2=\prod_{j=1}^{t-1}\left[\frac{(N_j+\sum\limits_{\Omega_{j-1}\cap\Omega^c_j} n^{(j-1)}_i)\,!}{(\sum\limits_{\Omega_{j+1}} n^{(j)}_i)\,!\,(\sum\limits_{\Omega_j\cap\Omega^c_{j+1}} n^{(j)}_i)\,!}\;\frac{(\sum\limits_{\Omega_j\cap\Omega^c_{j+1}} p_i)^{\sum\limits_{\Omega_{j+1}} n^{(j)}_i}}{(\sum\limits_{\Omega_j \ni i} p_i)^{N_j}}\right] \tag{4.7}$$

$$\times\frac{(N_t+\sum\limits_{\Omega_{t-1}\cap\Omega^c_t} n^{(t-1)}_i)\,!}{\sum\limits_{\Omega_t \ni i} p_i}\;\frac{\prod\limits_{\Omega_t \ni i} p_i^{\sum_{j=1}^{t} n^{(j)}_i}}{\prod\limits_{\Omega_t \ni i}(n_i^{\sum_{j=1}^{t} n^{(j)}_i})\,!} .$$

4.2 Chained case. Let us consider the estimation problem in which the observation spaces are linked like a chain. That is,

$$\begin{aligned}\Omega_i \cap \Omega_j &\neq \phi, \quad \text{if } j=i-1,\ i,\ i+1 ,\\ &=\phi, \quad \text{otherwise.}\end{aligned}$$

This case is a generalization of Kastenbaum (1958) and the case is also described in chapter 6 of Li (1961).

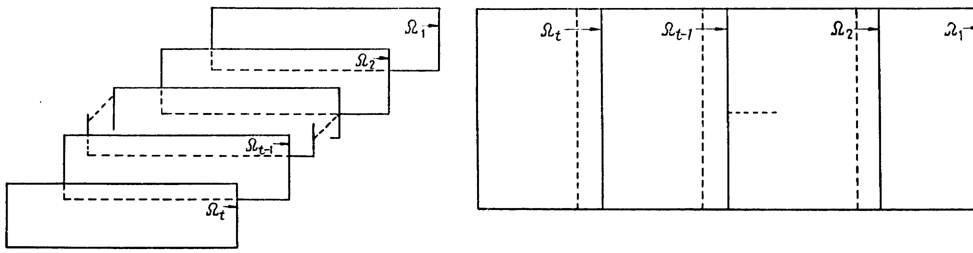

Fig. 2. The combination of contingency tables

Here we use double subscripts for the decomposition as follows:

$$\omega_{i1}\equiv\Omega_{i-1}\cap\Omega_i, \tag{4.8}$$
$$\omega_{i2}\equiv\Omega_i-\omega_{i1}-\omega_{i3},$$
$$\omega_{i3}\equiv\Omega_i\cap\Omega_{i+1}\,,\quad \text{for } i=1,\ 2,\ \cdots,\ t\,,$$

where $\Omega_0\equiv\Omega_{t+1}\equiv\phi$.

Furthermore, let P_{ij}'s be partial sums of p_{uv}'s on ω_{ij}, then P_u's are corresponding to P_{ij}'s so that

$$\sum_{u=1}^{2t-1}\sum_{v=1}^{l_u} p_{uv}\equiv\sum_{u=1}^{2t-1} P_u=\sum_{i=1}^{t}\sum_{j=1}^{3} P_{ij}=1\,, \tag{4.9}$$

where $P_{i,\,1}=P_{i-1,\,3}$ for $i=2,\ 3,\ \cdots,\ t$. Now we denote $n^*_{gh}\equiv\sum_{\omega_{gh}} n_j^{(i)}$. Then we obtain the following corollary 2 from theorem 4.

COROLLARY 2. *For all* $v\in\omega_u$, $u=1,\ 2,\ \cdots,\ 2t-1$, *the following explicit expression for the maximum likelihood estimate of* p_{uv} *is obtained from the combined samples:*

$$\hat{p}_{uv}=\frac{\displaystyle\sum_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} n_{uv}^{(i)}}{\displaystyle\sum_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} n_{u.}^{(i)}}\hat{P}_u \tag{4.10}$$

while

$$\hat{P}_1=\frac{n^*_{11}}{N_1}\,\frac{1}{1+\displaystyle\sum_{g=2}^{t}\frac{n^*_{12}}{N_1}\,\frac{\prod_{k=2}^{g-1} n^*_{k3}(n^*_{g2}+n^*_{g3})}{\prod_{k=2}^{g} n^*_{k1}}}\,, \tag{4.11}$$

$$\hat{P}_u = \hat{P}_{2i-3+j} = \frac{\prod_{k=2}^{i-1} n_{k3}^* n_{ij}^* \Big/ \prod_{k=2}^{i} n_k^*}{\dfrac{N_1}{n_{12}^*} + \sum_{g=2}^{t} \dfrac{\prod_{k=2}^{g-1} n_{k3}^*(n_{g2}^* + n_{g3}^*)}{\prod_{k=2}^{g} n_{k1}^*}} \tag{4.12}$$

for $u=2, 3, \cdots, 2t-1, \quad t \geqq i \geqq 2, \quad j=1, 2, 3.$

These estimates are also unbiased, consistent and jointly sufficient.

For example, in case $t=2$, we simply get

$$\hat{P}_1 = \frac{n_{11}^* n_{21}^*}{n_{12}^* n_{21}^* + n_{11}^* n_{21}^* + n_{12}^* n_{22}^*}, \quad \hat{P}_2 = \frac{n_{12}^* n_{21}^*}{n_{12}^* n_{21}^* + n_{11}^* n_{21}^* + n_{12}^* n_{22}^*} \tag{4.13}$$

$$\hat{P}_3 = 1 - \hat{P}_1 - \hat{P}_2 .$$

The proof of the latter half of this corollary may be omitted, because the principle is quite the same as before.

Acknowledgment

I wish to thank Professor E. Batschelet for the suggestion of this problem, and Mr. J. F. Gilroy, S. J., for his help in preparing the manuscript.

KOBE UNIVERSITY & SHIONOGI RESEARCH LABORATORY

REFERENCES

[1] E. Batschelet, " Über eine Kontingenztafel mit fehlenden Daten," *Biometr. Z.*, 2 (1960), 236–243.

[2] E. Batschelet, " Auslesefreie Verteilung des Manifestationsalters mit einer Anwendung auf die Respirationsatopien," *Biometr. Z.*, 2 (1960), 244–256.

[3] E. Batschelet, " Spurious correlation of the age of onset, with special reference to atopic diseases," *Biometr. Z.*, 4 (1962), 111–120.

[4] M. P. Geppert, " Erwartungstreue plausibelste Schätzer aus dreieckig gestutzten Kontingenztafeln," *Biometr. Z.*, 3 (1961), 55–67.

[5] C. C. Li, *Human Genetics, Principles and Methods,* McGraw Hill Book Co., 1961.

[6] M. A. Kastenbaum, " Estimation of relative frequencies of four sperm types in Drosophila Melanogester," *Biometrics*, 14 (1958), 223–228.

[7] G. S. Watson, " Missing and " mixed-up " frequencies in contingency tables," *Biometrics*, 12 (1956), 47–50.