The Catholic University of America
Washington, D. C.

Department of Mathematics
STATISTICAL LABORATORY
C.U. - Biom. - 2
December 1961

ON ESTIMATIONS WHEN CERTAIN TRUNCATED CONTINGENCY TABLES ARE POOLED

by

Chooichiro Asano

Prepared Under Contracts:
NSF-9968, National Science Foundation
Nonr-2249(05), and
Nr 301-579, U.S. Office of Naval Research

ON ESTIMATIONS WHEN CERTAIN TRUNCATED CONTINGENCY TABLES ARE POOLED*

by

Chooichiro Asano**

1. Introduction and summary.

In certain statistical analyses of contingency tables on genetic problems, it often happens that we have to estimate probabilities by appealing to combinations of observations. In such a situation, however, false associations among some observations due to the pooling of heterogeneous groups of observations have caused some comment by Batschelet [3] and Li [5].

To simplify the statement of our problem, we may restrict ourselves to a special subject studied by Batschelet in [3]. There, a problem is proposed where the question arises: Are the ages of onset for two siblings stochastically dependent or independent for special atopic diseases? To explain the situation, he illustrates a spurious relationship between the two lots of siblings by a very obvious model. Here we are confronted essentially with an estimation problem of probabilities on the basis of combining sibs of different registration.

The purpose of this paper is, first, to formulate the basic ideas and to work out the combined estimates and some properties, and, second to generalize these pooled problems related to various contingency tables. To conclude, the author justifies the validity of the above criticism.

---

*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.

**This study was aided 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.

2. Formulation of the problems and solutions.

Suppose that the first individuals of a population can be described as belonging to one of r_1 categories with respect to an attribute A and to one of s_1 categories with respect to an attribute B, and the second individuals can be truncatedly observed in one of r_2 in r_1 categories for A and in one of s_2 in s_1 categories for B. And let $n_{ij}^{(t)}$ and p_{ij} be the independent frequency of t-th observations and the probability in the i-th for A and j-th for B, t=1,2.

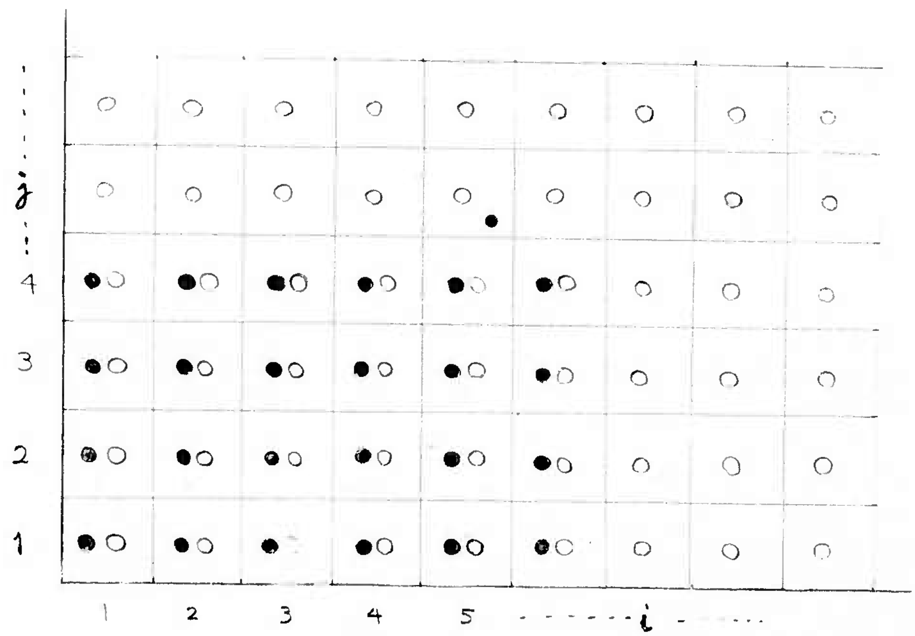

Fig. 1 (i,j)-cell has a probability p_{ij}.
Observations, $\{ \circ : n_{ij}^{(1)}, \bullet : n_{ij}^{(2)} \}$.

Moreover, to simplify the case in general, let Ω_1 and Ω_2 be the first observations space and the second denoted by $r_1 \times s_1$ and $r_2 \times s_2$ respectively and let $n_i^{(t)}$ and p_i be the independent frequency of t-th observations, t=1,2, and the probability defined by i which belongs to Ω_1 and/or Ω_2, where $\Omega_1 \supset \Omega_2$ and $\sum_{\Omega_1 \ni i} p_i = 1$.

Then the likelihood function L is given by

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

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

Under these circumstances, we obtain the following theorems.

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

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

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

Proof. The likelihood estimates given by (2.2) and (2.3) are directly obtained by solving the normal equations and using the invariant property of the maximum likelihood estimation.

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

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

$$V\{\hat{p}_i\} = \frac{p_i(1-p_i)}{N_1} + \frac{N_2}{N_1} p_i \left\{ p_i \sum_{\Omega_2 \ni j} p_j - 1 \right\} \sum_{n_1=0}^{N_1-1} \frac{\binom{N_1-1}{n_1}}{N_1+N_2-n_1} \left\{ \sum_{\Omega_1 \cap \Omega_2^c} p_j \right\}^{n_1} \left\{ \sum_{\Omega_2} p_j \right\}^{N_1-1-n_1}, \tag{2.5}$$

$$\text{for } i \in \Omega_2,$$

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

$$(2.7)\qquad \mathrm{Cov}\{\hat{p}_i \hat{p}_j\} = -\frac{p_i p_j}{N_1}\left[1 - N_2\left\{\sum_{\Omega_2} p_k\right\} \sum_{n_1=0}^{N_1-1} \frac{\binom{N_1-1}{n_1}}{N_1+N_2-n_1}\left\{\sum_{\Omega_1\cap\Omega_2^c} p_k\right\}^{n_1}\left\{\sum_{\Omega_2} p_k\right\}^{N_1-1-n_1}\right],$$

$$\text{for } i \neq j,\ i,j \in \Omega_2 ,$$

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

where we put $n_1 = \sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}$. Moreover, if we may assume that the size N_2 is sufficiently larger than the size of N_1, then, from (2.5) and (2.7), we obtain the following asymptotic relations.

$$(2.5)'\qquad V\{\hat{p}_i\} \sim p_i\left\{(1-p_i)^2 + p_i \sum_{\substack{\Omega_2,\\ i\neq j}} p_j\right\} \Big/ N_1 \qquad \text{for } i \in \Omega_2 ,$$

$$(2.7)'\qquad \mathrm{Cov}\{\hat{p}_i \hat{p}_j\} \sim -p_i p_j\left\{1 - \sum_{\Omega_2} p_k\right\} \Big/ N_1 \qquad \text{for } i \neq j,\ i,j \in \Omega_2 .$$

Proof

(i) Unbiasedness. Unbiasedness of $\hat{p}_i$ defined by (2.2) is obvious.

To show that (2.3) defines an unbiased estimate we study the likelihood function (2.1). It is based on the product of two separate multinomial distributions, that is, one ordinary multinomial distribution and a conditional multinomial distribution, and is essentially expressed as a general term of expansion of the numerator of $\left\{\sum_{\Omega_1} p_i\right\}^{N_1} \cdot \left\{\sum_{\Omega_2} p_i\right\}^{N_2} \Big/ \left\{\sum_{\Omega_2} p_i\right\}^{N_2}$. Now, from this viewpoint, we may rewrite the likelihood function as follows:

$$(2.9) \qquad L_0 = \frac{N_1!}{\left\{\sum_{\Omega_1 \cap \Omega_2^c} n_i^{(1)}\right\}! \left\{\sum_{\Omega_2} n_i^{(1)}\right\}!} \frac{\left(N_2 + \sum_{\Omega_2} n_i^{(1)}\right)!}{\left\{\sum_{\Omega_1 \cap \Omega_2^c} (n_i^{(1)} + n_i^{(2)})\right\}!} \frac{\left\{\sum_{\Omega_1 \cap \Omega_2^c} p_i\right\}^{\sum_{\Omega_1 \cap \Omega_2^c} n_i^{(1)}}}{\left\{\sum_{\Omega_2} p_i\right\}^{N_2}} \prod_{\Omega_1 \cap \Omega_2^c} p_i^{(n_i^{(1)} + n_i^{(2)})},$$

where naturally

$$(2.10) \qquad \sum_{\left(\sum_{\Omega_1 \cap \Omega_2^c} n_i^{(1)}, \sum_{\Omega_2} n_i^{(1)}\right)} \sum_{\left(\cdots, \sum_{t=1}^{2} n_i^{(t)}, \cdots\right)} L_0 = 1.$$

$$\sum_{\Omega_1} n_i^{(1)} = N_1, \sum_{\Omega_2} \left(\sum_t n_i^{(t)}\right) = N_2 + \sum_{\Omega_2} n_i^{(1)}.$$

Hence the expectation of $\hat{p}_i$ given by (2.3) is expressed by

$$(2.11)\qquad E\{\hat{p}_i\} = \sum\sum \frac{\sum_{\Omega_2} n_i^{(1)} \cdot \{\sum_{t=1}^{2} n_i^{(t)}\}}{N_1 (N_2 + \sum_{\Omega_2} n_i^{(1)})} L_0 \ ,$$

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

$$(2.12)\qquad E\{\hat{p}_i\} = p_i \sum\sum L_0 = p_i \ .$$

(ii) Variances and Covariances.

To prove (2.4) and (2.6) we have to calculate the following expectations:

$$(2.13)\qquad E\{\hat{p}_i^2\} = p_i^2 - \frac{p_i^2}{N_1} + \frac{p_i}{N_1} \quad \text{for } i \in \Omega_1 \cap \Omega_2^c \ ,$$

$$(2.14)\qquad E\{\hat{p}_i \hat{p}_j\} = \frac{N_1 - 1}{N_1} p_i p_j \quad \text{for } i \neq j,\ i, j \in \Omega_1 \cap \Omega_2^c \ ,$$

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 apply the likelihood function (2.9).

$$(2.15)\qquad E\{\hat{p}_i^2\} = \sum\sum \frac{\left(\sum_{\Omega_2} n_i^{(1)}\right)^2}{N_1^2\left(N_2+\sum_{\Omega_2} n_i^{(1)}\right)^2}\left\{\left(\sum_{t=1}^{2} n_i^{(t)}\right)\left(\sum_{t=1}^{2} n_i^{(t)}-1\right)+\left(\sum_{t=1}^{2} n_i^{(t)}\right)\right\} L_0$$

$$= \frac{N_1-1}{N_1} p_i^2 + \frac{N_2}{N_1} p_i^2 \left\{\sum_{\Omega_2} p_i\right\} \sum_{n_1=0}^{N_1-1} \frac{\binom{N_1-1}{n_1}}{N_1+N_2-n_1}\left\{\sum_{\Omega_1\cap\Omega_2^c} p_i\right\}^{n_1}\left\{\sum_{\Omega_2} p_i\right\}^{N_1-1-n_1}$$

$$+ \frac{p_i}{N_1} - \frac{N_2}{N_1} p_i \sum_{n_1} \frac{\binom{N_1-1}{n_1}}{N_1+N_2-n_1}\left\{\sum_{\Omega_1\cap\Omega_2^c} p_i\right\}^{n_1}\left\{\sum_{\Omega_2} p_i\right\}^{N_1-1-n_1},$$

for $i \in \Omega_2$.

$$(2.16)\qquad E\{\hat{p}_i\hat{p}_j\} = \sum\sum \frac{\left(\sum_{\Omega_2} n_i^{(1)}\right)^2\left(\sum_{t=1}^{2} n_i^{(t)}\right)\left(\sum_{t=1}^{2} n_j^{(t)}\right)}{N_1^2\left(N_2+\sum_{\Omega_2} n_i^{(1)}\right)^2} L_0$$

$$= \frac{N_1-1}{N_1} p_i p_j + \frac{N_2}{N_1} p_i p_j \left\{\sum_{\Omega_2} p_i\right\} \sum \frac{\binom{N_1-1}{n_1}}{N_1+N_2-n_1}\left\{\sum_{\Omega_1\cap\Omega_2^c} p_i\right\}^{n_1}\left\{\sum_{\Omega_2} p_i\right\}^{N_1-1-n_1},$$

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 formulae (2.1) and (2.9) in the following way:

$$(2.17)\qquad E\{\hat{p}_i\hat{p}_j\} = \sum_{\Omega_1,\Omega_2} \frac{n_j^{(1)}\left(n_i^{(1)}+n_j^{(2)}\right)\left(\sum_{\Omega_2} n_i^{(1)}\right)}{N_1^2\left(N_2+\sum_{\Omega_2} n_i^{(1)}\right)} L_0$$

$$= p_j \sum\sum \frac{\left(n_i^{(1)}+n_j^{(2)}\right)\left(\sum_{\Omega_2} n_i^{(1)}\right)}{N_1\left(N_2+\sum_{\Omega_2} n_i^{(1)}\right)} L_0',$$

where L'_0 denotes a formula where we replace N_1 with N_1-1 and $\sum n_i^{(1)}$ with $\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}-1$ at L_0, respectively, and then

$$(2.18)\qquad E\{\hat{p}_i\hat{p}_j\} = \frac{N_1-1}{N_1}p_ip_j\sum\sum L''_0 = \frac{N_1-1}{N_1}p_ip_j \quad \text{for } i\in\Omega_2,\ j\in\Omega_1\cap\Omega_2^c,$$

where L''_0 denotes a formula in which N_1-2, $\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}-1$ and $\sum_{\Omega_2} n_i^{(1)}-1$ substitute
for N_1, $\sum_{\Omega_1\cap\Omega_2^c} n_i^{(1)}$ and $\sum_{\Omega_2} n_i^{(1)}$ respectively. Thus we obtain (2.8) from $\mathrm{Cov}\{\hat{p}_i\hat{p}_j\} = E\{\hat{p}_i\hat{p}_j\} - p_ip_j$. This covariance is obviously independent of the second observations.

(iii) Consistency and Sufficiency. Consistency is evident since the estimates are unbiased and the variances and covariances tend to zero if $N_1\to\infty$. Finally, the estimates are sufficient because the likelihood function can be expressed itself as a function of parameters and statistics.

Theorem 3. The information matrix of our estimates in the Fisher's sense is given by the following elements:

$$(2.19)\qquad \sigma^{ii} = N_1\left(\frac{1}{p_i}-\frac{1}{p_0}\right) \qquad \text{for } i\neq 0,\ i\in\Omega_1\cap\Omega_2^c,$$

$$(2.20)\qquad \sigma^{ii} = \frac{N_1+N_2}{p_i} - \frac{N_1}{p_0} - \frac{N_2}{\left(\sum_{\Omega_2} p_i\right)^2} \qquad \text{for } i\neq j,\ i,j\neq 0,\ i,j\in\Omega_1\cap\Omega_2^c,$$

$$(2.21)\qquad \sigma^{ij} = \frac{N_1}{p_0} \qquad \text{for } i\in\Omega_2,$$

$$(2.22)\qquad \sigma^{ij} = \frac{N_1}{p_0} - \frac{N_2}{\left(\sum_{\Omega_2} p_i\right)^2} \qquad \text{for } i\neq j,\ i,j\in\Omega_2,$$

$$(2.3) \qquad \sigma^{ij} = \sigma^{ji} = \frac{N_1}{p_0} \qquad \text{for } i \neq 0,\ i \in \Omega_1 \cap \Omega_2^c,\ j \in \Omega_2,$$

where p_0 belongs to $\Omega_1 \cap \Omega_2^c$ and is assumed to be a linearly dependent probability by reason of $\sum_{\Omega_1} p_i = 1$. And the estimates are inefficient.

Proof. The information matrix is obtained by the ordinary method, $\sigma^{ij} = -E\left\{\partial^2 \log L / \partial p_i \partial p_j\right\}$. And inefficiency is obvious, since the variances-covariances matrix given by Theorem 2 cannot attain the inverse information matrix.

3. Numerical example.

To illustrate the above formulation and results, let us suppose the following ficticious samples with regard to the heredity of atopic disease. There are 72 pairs of siblings with an atopic disease and they are classified into one of 12 cells each assigned to the interval of five years, in Table 1. Furthermore, we have investigated 99 pairs of atopic children which belong to the same family and which are nearly of the same age at their onset as a supplement of Table 1. They are classified, as in Table 2, into one of 4 cells of the same scale as Table 1.

Table 1

X_2 (i) \ X_1 (j)	(1) 0–5	(2) 5–10	(3) 10–15	(4) 15–20
(3) 10–15	15	10	7	3
(2) 5–10	12	6	3	2
(1) 0–5	9	1	4	1

Table 2

X_2 (i) \ X_1 (j)	(1) 0–5	(2) 5–10
(2) 5–10	53	18
(1) 0–5	17	11

X_1 = the age of onset for the elder sibling.
X_2 = the age of onset for the younger sibling.

Now let the probability in each cell be denoted by p_{ij} with suffix (i,j) in order of youth $i=1,2,3$ and $j=1,2,3.4$. Then our problem is to estimate each p_{ij} by combining the above tables.
Indeed, the estimates are calculated as follows:

$$\hat{p}_{11} = \frac{(7+17)\times 28}{73\times(28+99)} \qquad \hat{p}_{12} = \frac{(12+53)\times 28}{73\times(28+99)} \qquad \hat{p}_{13} = \frac{15}{73} \qquad \hat{p}_{21} = \frac{(1+17)\times 28}{73\times(28+99)}$$

$$\hat{p}_{22} = \frac{(6+53)\times 28}{73\times(28+99)} \qquad \hat{p}_{23} = \frac{10}{73} \qquad \hat{p}_{31} = \frac{4}{73} \qquad \hat{p}_{32} = \frac{3}{73}$$

$$\hat{p}_{33} = \frac{1}{73} \qquad \hat{p}_{41} = \frac{1}{73} \qquad \hat{p}_{42} = \frac{2}{73} \qquad \hat{p}_{43} = \frac{3}{73}$$

4. Generalization of the problem.

Let us consider the combined estimates in the most generalized case. Suppose that Ω_i, $p_j^{(i)}$ and $n_j^{(i)}$ denote the i-th sample space $i=1,2,\ldots,t$ a probability of j-th cell in Ω_i, and a mutually independent frequency of the i-th observations in the j-th cell, $j=1,2,\ldots,k_i$, respectively. And assume that essentially each observation space is at least partially associated to one or several common cells. Furthermore, we put $\Omega_i \neq \Omega_{i'}$ for any i and i', $i \neq i'$, $i, i' = 1, 2, \cdots, t$, without a loss of generality, since we may make afresh $\Omega_{i''}$ by means of combining both i-th and i'-th observations in case $\Omega_i = \Omega_{i'}$ for some $i \neq i'$.

Now let us make decompositions $\{\omega_u\}$ of $\bigcup_{i=1}^{t} \Omega_i$ such as

$$(4.1) \qquad \sum_{u=1}^{2^t-1} \omega_u = \sum_{\substack{r_i=0,1 \\ 1<\sum r_i \leq t \\ \Omega_i^1 \equiv \Omega_i,\ \Omega_i^0 \equiv \Omega_i^c}} \bigcap_{i=1}^{t} \Omega_i^{r_i} = \cup\, \Omega_i \ (\equiv \Omega),$$

where $\omega_u \cap \omega_{u'} = 0$ for $u \neq u'$, $u, u' = 1, 2, \cdots, 2^t-1$.

Then the suffix u can be defined for arbitrary i as follows:

(4.2) $$u = h \qquad \text{for } i=1,$$

(4.3) $$u = \sum_{r=1}^{i-1} 2^{t-r} + h \qquad \text{for } i=2,3,\cdots,t,$$

(4.4) $$\text{and} \quad h = 1,2,3,\cdots,2^{t-i} \qquad \text{for both (4.2) and (4.3).}$$

Let p_{uv} and P_u be a probability of cells in ω_u, $v=1,2,\ldots,l_u$, and $P_u \equiv \sum_v p_{uv}$, where $\sum_{u=1}^{2^t-1} P_u = 1$. Hence, we have $\sum_{u=1}^{2^t-1} l_u - 1$ parameters to be estimated and we may and now shall assume P_1 is a linearly dependent parameter, since the maximum likelihood estimation is invariant.

Under these circumstances, we obtain the following theorem in a general case.

<u>Theorem 4</u> The combined estimates of the p_{uv}'s, $u = 1,2,\ldots,2^t-1$ and $v = 1,2,\ldots,l_u$, are obtained as follows:

(4.5) $$\hat{p}_{uv} = \frac{\sum_{\cup\omega_{u(i)}} n_{uv}}{\sum_{\cup\omega_{u(i)}} n_{u\cdot}} \hat{P}_u \qquad \text{for } u=1,2,3,\cdots,2^t-1,$$
$$\text{and } v=1,2,\cdots,l_u,$$

where

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

and $\hat{P}_2, \hat{P}_3, \ldots, \hat{P}_{2^t-1}$ are obtained as solutions of 2^t-2 simultaneous equations (4.7) .

$$(4.7)\qquad \sum_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}}\left[\frac{n_{u\cdot}}{P_u}-\frac{N_i}{\sum\limits_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}} P_u}\right]=0 \quad \text{for } u=1,2,\cdots,2^t-1,$$

and where we put

$$(4.8)\qquad \{n_{uv}\}\equiv\{n_j^{(i)}\,|\,\Omega_i\supset\omega_u\ni j\},\quad n_{u\cdot}\equiv\sum_{v=1}^{l_u} n_{uv},\quad N_i\equiv\sum_{j=1}^{k_i} n_j^{(i)},$$

$$(4.9)\qquad \{\omega_{u(i)}\}\equiv\left\{\omega_u\,\middle|\,\left(\Omega_i^1\cap\Big(\bigcap_{\substack{i'=1\\ \left[\substack{i'\neq i,\ r_{i'}=0,1\\ 0\le\sum r_{i'}<t}\right]}}^{t}\Omega_{i'}^{r_{i'}}\Big)\right)\supset\omega_u\right\},\quad \cup\,\omega_{u(i)}\equiv\bigcup_{\substack{i\\ \{i|\Omega_i\supset\omega_u\}}}\omega_{u(i)}.$$

Proof: The likelihood function L is obviously given by

$$(4.10)\qquad L=\frac{\prod_{i=1}^{t} N_i!}{\prod_{i=1}^{t}\prod_{\{j|\Omega_i\ni j\}} n_j^{(i)}!}\prod_{i=1}^{t}\prod_{\{j|\Omega_i\ni j\}}\left[\frac{p_j^{(i)}}{\sum\limits_{\{j|\Omega_i\ni j\}} p_j^{(i)}}\right]^{n_j^{(i)}},$$

and is also written by our notations, (4.8) and (4.9), as follows

$$(4.11)\qquad L=\text{const.}\prod_{i=1}^{t}\prod_{\{u|\Omega_i\supset\omega_u\}}\prod_{v=1}^{l_u}\left[\frac{p_{uv}}{\sum\limits_{\{u|\Omega_i\supset\omega_u\}} P_u}\right]^{n_{uv}},$$

and then we obtain

$$(4.12)\qquad \log L=\log(\text{const.})+\sum_{i=1}^{t}\left[\sum_{\{u|\Omega_i\supset\omega_u\}}\sum_{v=1}^{l_u} n_{uv}\left\{\log p_{uv}-\log\Big(\sum_{\{u|\Omega_i\supset\omega_u\}} P_u\Big)\right\}\right]$$

and

$$(4.13)\qquad \frac{\partial\log L}{\partial p_{uv}}=\sum_{\{i|\Omega_i\supset\omega_u\}}\left[\frac{n_{uv}}{p_{uv}}-\frac{N_i}{\sum\limits_{\{u|\Omega_i\supset\omega_u\}} P_u}\right]=0\,.$$

While the likelihood function is written for P_u's as follows:

$$(4.14) \qquad L_1 = \frac{\prod_{i=1}^{t} N_i!}{\prod_{i=1}^{t} \prod_{\{u|\Omega_i \supset \omega_u\}} n_{u\cdot}!} \prod_{i=1}^{t} \prod_{\{u|\Omega_i \supset \omega_u\}} \left[\frac{P_u}{\sum_{\{u|\Omega_i \supset \omega_u\}} P_u} \right]^{n_{u\cdot}},$$

and then we obtain

$$(4.15) \qquad \frac{\partial \log L_1}{\partial P_u} = \sum_{\{i|\Omega_i \supset \omega_u\}} \left[\frac{n_{u\cdot}}{P_u} - \frac{N_i}{\sum_{\{u|\Omega_i \supset \omega_u\}} P_u} \right] = 0 \quad \text{for } u = \ , 3, \cdots, 2^t - 1,$$

where $\sum_{\Omega} P_u = 1$ which is also (4.7) .

Thus, by combining both (4.13) and (4.15) , we obtain the following simple formula shown in Theorem 4.

$$(4.16) \qquad \hat{P}_{uv} = \frac{\sum n_{uv}}{\sum n_{u\cdot}} \hat{P}_u \qquad \text{for } u = 1, 2, 3, \cdots, 2^t - 1,$$

which may be understood intuitively as an expected relation. (Q.E.D.)

Generally, to obtain the solutions $\hat{P}_u$'s, u=2,3,...,2^t-1 , from (4.7) explicitly, it becomes more complicated as t increases. The estimates, however, are given by iterative solutions of maximum likelihood equations in the following way.

Let the solutions be $\hat{P}_2, \hat{P}_3, \ldots, \hat{P}_{2^t-1}$. Suppose that

$\tilde{P}_{21}, \tilde{P}_{21}, \ldots, \tilde{P}_{2^t-1,1}$ are approximations to $\hat{P}_2, \hat{P}_3, \ldots, \hat{P}_{2^t-1}$ obtained in any manner; the easiest procedure will be to apply values of a rough calculation neglecting any combination of observations at overlapping cells and using temporary values of the first observations.

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

(4.17) $\tilde{P}_{22} = \tilde{P}_{21} + \delta\tilde{P}_{21}$, $\tilde{P}_{32} = \tilde{P}_{31} + \delta\tilde{P}_{31}$, --- , $\tilde{P}_{2^t-1,2} = \tilde{P}_{2^t-1,1} + \delta\tilde{P}_{2^t-1,1}$,

where the increments $\delta\tilde{P}_{21}$, $\delta\tilde{P}_{31}$, ---, $\delta\tilde{P}_{2^t-1,1}$ are the solutions of

$$(4.18) \qquad \frac{\partial \log L}{\partial \tilde{P}_{u1}} + \delta\tilde{P}_{u1} \frac{\partial^2 \log L}{(\partial \tilde{P}_{u1})^2} + \sum_{\substack{j=2 \\ j \neq u}}^{2^t-1} \delta\tilde{P}_{j1} \frac{\partial^2 \log L}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = 0$$

for $u \neq j$, $u, j = 2, 3, \ldots, 2^t - 1$.

Addition of a suffix 1 to P_u, P_j indicates replacement by $\tilde{P}_{u1}$, $\tilde{P}_{j1}$ after differentiations, and further each term of (4.18) is given as follows:

$$(4.19) \qquad \frac{\partial \log L}{\partial \tilde{P}_{u1}} = \sum_{\{i|\Omega_i \ni \omega_u\}} \left[\frac{n_{u\cdot}}{\tilde{P}_{u1}} - \frac{N_i}{\sum \tilde{P}_{u1}} \right] ,$$

$$(4.20) \qquad \frac{\partial^2 \log L}{(\partial \tilde{P}_{u1})^2} = -\sum_{\{i|\Omega_i \ni \omega_u\}} \left[\frac{n_u}{\tilde{P}_{u1}^2} - \frac{N_i}{(\sum \tilde{P}_{u1})^2} \right] ,$$

$$(4.21) \qquad \frac{\partial^2 \log L}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = -\sum_{\{i|\Omega_i \ni u, j\}} \frac{N_i}{(\sum \tilde{P}_{u1})^2} \qquad \text{for } j \in (\Omega_i \ni u) ,$$

$$(4.22) \qquad \frac{\partial^2 \log L}{\partial \tilde{P}_{u1} \partial \tilde{P}_{j1}} = 0 \qquad \text{for } j \bar{\in} (\Omega_i \ni u) .$$

An iterative process may be based on 2^t-2 equations shown by (4.16), 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 $\hat{P}_u, \hat{P}_j$ is achieved. Thus we obtain the individual estimates of $\hat{p}_{uv}$ by applying (4.16).

Theorem 5 The necessary and sufficient condition for p_i to be estimable is that any observation space is at least partially associated to one or several common cells. In other words, it is the condition that each observation space is somewhere overlapping another.

Proof: The readers would feel that this theorem is natural. (Necessity) Suppose that t observation spaces are obtained independently 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 each of s connected observation 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 we cannot obtain uniquely the k parameters. From this we conclude that s should be 1.

(Sufficiency) By using the above notations, if $s=1$, then we are indeed able to obtain the estimates of all p_{uv}'s as shown in Theorem 4. (Q.E.D.)

5. Certain typical cases in generalization

Concerning certain special types of generalization, let us give some explicit estimates of p_{uv}'s as colloraries of Theorem 4.

5.1 Let us consider an estimation problem where the observation spaces form a sequence of nested spaces, such as $\Omega \equiv \Omega_1 \supset \Omega_2 \supset \cdots \supset \Omega_t$, and this is corresponding to a generalization of section 2, (cf. Fig.2).

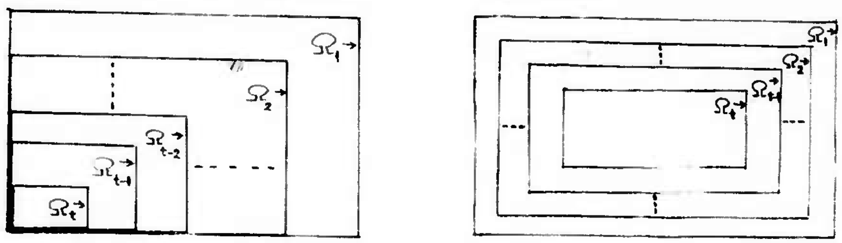

Fig. 2 Contingency tables

In this case, the decompositions $\{\omega_u\}$ of $\bigcup_{i=1}^{t} \Omega_i$ are defined as follows:

$$\sum_{u=1}^{t} \omega_u = \bigcup_{i=1}^{t} (\Omega_i \cap \Omega_{i+1}^c) = \Omega \quad , \tag{5.1}$$

where we put $\Omega_{t+1}^c \equiv \Omega$ and values of u equal to those of i . And when p_{uv}'s are denoted probabilities of cells in ω_u, $v=1,2,\ldots,l_u$, we obtain the following Collorary 1.

Collorary 1 The combined estimates of the p_{uv}'s, $u=1,2,\ldots,t$ and $v=1,2,\ldots,lu$, are obtained explicitly by the maximum likelihood estimation as follows:

$$\hat{p}_{uv} = \frac{n_{uv}}{\displaystyle\sum_{j=1}^{t} \frac{n_v^{(j)}}{\left\{1 - \sum_{r=0}^{j-1} \sum_{\omega_r} \hat{p}_{rv}\right\}}} \tag{5.2}$$

for all $v \in \omega_u$, $u=1,2,\ldots,t$, where we define $\omega_0 \equiv \phi$ and $\hat{p}_{0v} \equiv 0$. Furthermore, these estimates are also unbiased, consistent and sufficient as in Theorem 2.

Indeed, we can obtain the explicit estimates of p_{uv}'s by representing successively the above formula (5.2) through $u=1$.

For example, in case $t=3$, we obtain as follows:

$$(5.3) \qquad \hat{p}_{1v_1} = \frac{n_{1v_1}}{N_1} \qquad \text{for} \quad v_1 \in \omega_1 ,$$

$$(5.4) \qquad \hat{p}_{2v_2} = \frac{n_{2v_2}}{N_1\left\{1 + \dfrac{n^{(2)}_{v_2}}{N_1\left\{1 - \dfrac{\sum_{\omega_1} n^{(1)}_{v_1}}{N_1}\right\}}\right\}} \qquad \text{for} \quad v_2 \in \omega_2 ,$$

and

$$(5.5) \qquad \hat{p}_{3v_3} = \frac{n_{3v_3}}{N_1\left\{1 + \dfrac{n^{(2)}_{v_3}}{N_1\left\{1 - \sum_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1}\right\}} + \dfrac{n^{(3)}_{v_3}}{N_1\left\{1 - \sum_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1} - \sum_{\omega_2}\dfrac{n_{2v_2}}{N_1\left\{1 + \dfrac{n^{(2)}_{v_2}}{N_1\left\{1 - \sum_{\omega_1}\dfrac{n^{(1)}_{v_1}}{N_1}\right\}}\right\}}\right\}}\right\}}$$

$$\text{for} \quad v_3 \in \omega_3 .$$

The latter half of this corollary is proved by quite the same manner as in Theorem 2 excepting the following likelihood functions are applied in place of L and L_0 given by (2.1) and (2.9) in the previous theorem respectively.

$$(5.6) \qquad L = \prod_{j=1}^{t}\left[\frac{N_j!}{\prod_{\Omega_j \ni i} n^{(j)}_i!}\prod_{\Omega_j \ni i}\left(\frac{p_i}{\sum_{\Omega_j} p_i}\right)^{n^{(j)}_i}\right] ,$$

$$(5.7)\qquad L_0 = \prod_{j=1}^{t-1}\left[\frac{\left(N_j+\sum_{\Omega_{j-1}\cap\Omega_j^c} n_i^{(j-1)}\right)!}{\left(\sum_{\Omega_{j+1}} n_i^{(j)}\right)!\left(\sum_{\Omega_j\cap\Omega_{j+1}^c} n_i^{(j)}\right)!}\;\frac{\left(\sum_{\Omega_j\cap\Omega_{j+1}^c} p_i\right)^{\sum_{\Omega_{j+1}} n_i^{(j)}}}{\left(\sum_{\Omega_j\ni i} p_i\right)^{N_j}}\right]\frac{\left(N_t+\sum_{\Omega_{t-1}\cap\Omega_t^c} n_i^{(t-1)}\right)!}{\sum_{\Omega_t\ni i} p_i}\;\frac{\prod_{\Omega_t\ni i} p_i^{\sum_{j=1}^{t} n_i^{(j)}}}{\prod_{\Omega_t\ni i}\left(\sum_{j=1}^{t} n_i^{(j)}\right)!}\,.$$

<u>5.2</u> Let us consider an estimation problem in case when the observation spaces are linked like a chain, and this case is corresponding to a generalization in chapter 6 of Li [5], (c.f. Fig. 3).

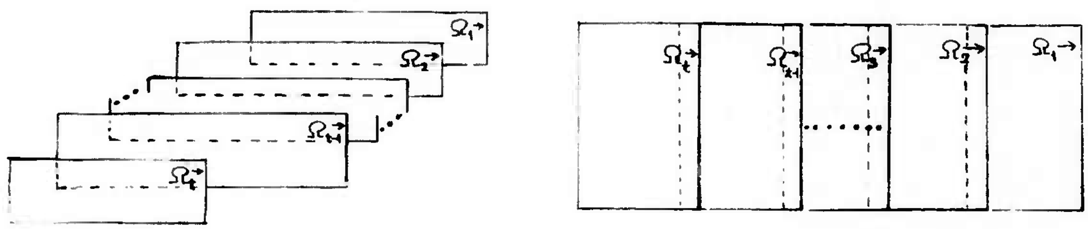

Fig. 3

In this case, the decompositions of $\bigcup_{i=1}^{t}\Omega_i$ are expressed by using conventional double suffix in place of $\{\omega_u\}$ as follows:

$$(5.8)\qquad \begin{array}{ll} \omega_{i1}\equiv\Omega_{i-1}\cap\Omega_i\,, & \omega_{i3}\equiv\Omega_i\cap\Omega_{i+1}\,,\\ \omega_{i2}\equiv\Omega_i-\omega_{i1}-\omega_{i3}\,, & \text{for } i=1,2,\cdots,t\,, \end{array}$$

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

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

$$(5.9)\qquad \sum_{u=1}^{2t-1}\sum_{v=1}^{l_u} p_{uv}\equiv\sum_{u=1}^{2t-1}P_u=\sum_{i=1}^{t}\sum_{j=1}^{3}P_{ij}=1\,,$$

where $P_{i,1} = P_{i-1,3}$ for $i=2,3,\ldots,t$. Then we obtain the following Corollary 2 by using Theorem 4.

<u>Corollary 2</u> The combined estimates of the p_{uv}'s, $u=1,2,\ldots,t$ and $v=1,2,\ldots,\ell_u$, are obtained explicitly as follows:

$$(5.10)\qquad \hat{p}_{uv} = \frac{\sum_{\cup \omega_{u(i)}} n_{uv}}{\sum_{\cup \omega_{u(i)}} n_{u\cdot}} \hat{P}_u \qquad \text{for } u=1,2,\ldots,2t-1 \text{ and } v=1,2,\ldots,\ell_u ,$$

and

$$(5.11)\qquad \hat{P}_1 = \frac{n_{11}^*}{N_1} \frac{1}{1+\sum_{g=2}^{t} \frac{n_{12}^*}{N_1} \frac{\prod_{k=2}^{g-1} n_{k3}^* (n_{g2}^*+n_{g3}^*)}{\prod_{k=2}^{g} n_{k1}^*}} ,$$

$$(5.12)\qquad \hat{P}_u = \hat{P}_{2i-3+j} = \frac{\dfrac{\prod_{k=2}^{i-1} n_{k3}^* \cdot n_{ij}^*}{\prod_{k=2}^{i} n_{k1}^*}}{\dfrac{N_1}{n_{12}^*} + \sum_{g=2}^{t} \dfrac{\prod_{k=2}^{g-1} n_{k3}^* (n_{g2}^*+n_{g3}^*)}{\prod_{k=2}^{g} n_{k1}^*}} \qquad \text{for } u=2,3,\ldots,2t-1,\ t \geq i \geq 2,\ j=1,2,3,$$

where $n_{gh}^* \equiv \sum_{\omega_{gh}} n_j^{(i)}$. These estimates are also unbiased, consistent and sufficient.

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

$$(5.13)\qquad \hat{P}_1 = \frac{n_{11}^* n_{21}^*}{n_{12}^* n_{21}^* + n_{11}^* n_{21}^* + n_{12}^* n_{22}^*} , \quad \hat{P}_2 = \frac{n_{12}^* n_{21}^*}{n_{12}^* n_{21}^* + n_{11}^* n_{21}^* + n_{12}^* n_{22}^*}$$

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

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

Acknowledgment

I wish to thank Professor E. Batschelet for the suggestion of this problem.

REFERENCES

[1] Batschelet, E. (1960). Auslesefreie Verteilung des Manifestationsalters mit einer Anwendung auf die Respirationsatopien, Biometr. Zeitschr. 2: p. 244-56.

[2] Batschelet, E. (1960). Über eine Kontingenztafel mit fehlenden Daten, ibid 2, p. 236-243.

[3] Batschelet, E. (1961). Spurious correlation of the age of onset, with special reference to atopic diseases, (to be published).

[4] Geppert, M. P. (1961). Erwartungstreue plausibelste Schätzer aus dreieckig gestutzten Kontingenztafeln, Biometr. Zeitschr. 3, p.55-67.

[5] Li, C. C. (1961). Human Genetics--Principles and Methods, McGraw-Hill Book Co., Chap. 6.

[6] Neel, J. V. and Schull, W. J. (1958). Human heredity, The University of Chicago Press, Chicago.