NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A COMBINATORIAL FORMULA AND ITS APPLICATION TO THE THEORY OF PROBABILITY OF ARBITRARY EVENTS[1]

By Kai-Lai Chung and Lietz C. Hsu

National Southwest Associated University, Kunming, China

An important principle, known as a proposition in formal logic or the method of cross-classification can be stated as follows.[1]

Let F and f be any two functions of combinations out of $(\nu) = (1, 2, \cdots, n)$. Then the two formulas

$$F((\alpha)) = \sum_{(\beta) \in (\nu)-(\alpha)} f((\alpha) + (\beta)) \tag{1.1}$$

$$f((\alpha)) = \sum_{(\beta) \in (\nu)-(\alpha)} (-1)^b F((\alpha) + (\beta)) \tag{2.1}$$

are equivalent.

As an immediate application to the theory of probability of arbitrary events, we have the set of inversion formulas[2]

$$p((\alpha)) = \sum_{(\beta) \in (\nu)-(\alpha)} p[(\alpha) + (\beta)] \tag{3.1}$$

$$p[(\alpha)] = \sum_{(\beta) \in (\nu)-(\alpha)} (-1)^b p((\alpha) + (\beta)) \tag{4.1}$$

where $p((\alpha))$ is the probability of the occurrence of at least $E_{\alpha_1}, E_{\alpha_2}, \cdots, E_{\alpha_a}$ out of n arbitrary events $E_1, E_2, \cdots, E_n$ and $p[(\alpha)]$ is the probability of the occurrence of $E\alpha_1, E\alpha_2, \cdots, E\alpha_a$ and no others among the n events, $(\alpha_1, \alpha_2, \cdots, \alpha_a)$ denoting a combination of the integers $(1, 2, \cdots, n)$. They can be made to play a central rôle in the theory, since they supply a method for converting the fundamental systems of probabilities, $p[(\alpha)]$ and $p((\alpha))$, one into the other.

We may further generalize (1.1) and (2.1) by considering combinations with repetitions. Let such a combination be written as

$$(\alpha) = \overline{(\alpha^r)} = (\alpha_1^{r_1}\alpha_2^{r_2} \cdots \alpha_a^{r_a})$$

---

[1] For the notations and definitions see K. L. Chung, "On fundamental systems of probabilities of a finite number of events," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 123–133.

[2] Cf. Fréchet, *Les probabilités associées à un système d'événements compatibles et dépendants*, Hermann, Paris (1939), formulas (55) and (58).

where r_i ($r_i \geq 1$) denotes the number of repetitions of the number α_i, $i = 1, 2, \cdots, a$. Correspondingly we write

$$(\alpha)' = (\alpha_1\alpha_2 \cdots \alpha_a)$$

and call it the reduced combination corresponding to (α).

If there are n distinct elements $(1, 2, \cdots, n)$ in question, we may write every combination in the form

$$(1^{r_1}2^{r_2} \cdots n^{r_n})$$

where each r_i is zero or a positive integer. We say that $(1^{s_1}2^{s_2} \cdots n^{s_n})$ belongs to $(1^{r_1}2^{r_2} \cdots n^{r_n})$ and write

$$(1^{s_1}2^{s_2} \cdots n^{s_n}) \in (1^{r_1}2^{r_2} \cdots n^{r_n})$$

if and only if for each i, $i = 1, 2, \cdots, n$, we have $s_i \leq r_i$. We write

$$(1^{r_1}2^{r_2} \cdots n^{r_n}) + (1^{s_1}2^{s_2} \cdots n^{s_n}) = (1^{r_1+s_1}2^{r_2+s_2} \cdots n^{r_n+s_n});$$

and if $(1^{s_1}2^{s_2} \cdots n^{s_n}) \in (1^{r_1}2^{r_2} \cdots n^{r_n})$,

$$(1^{r_1}2^{r_2} \cdots n^{r_n}) - (1^{s_1}2^{s_2} \cdots n^{s_n}) = (1^{r_1-s_1}2^{r_2-s_2} \cdots n^{r_n-s_n}).$$

We define a generalized Möbius function $\mu((\alpha))$ for combinations (with or without repetitions) as follows

$$\mu((\alpha)) = \begin{matrix} (-1)^a & \text{if} & (\alpha) = (\alpha)' \\ 0 & \text{if} & (\alpha) \neq (\alpha)'. \end{matrix}$$

This function has the property

$$\sum_{(\beta) \in (\alpha)} \mu((\beta)) = \begin{matrix} 1 & \text{if} & (\alpha) = (0) \\ 0 & \text{if} & (\alpha) \neq (0). \end{matrix}$$

For we have

$$\sum_{(\beta) \in (\alpha)} \mu((\beta)) = \sum_{(\beta) \in (\alpha)'} (-1)^b = \sum_{b=0}^{a'} (-1)^b \binom{a'}{b}$$

$$= \begin{matrix} 1 & \text{if} & a' = 0 \\ 0 & \text{if} & a' \neq 0 \end{matrix} = \begin{matrix} 1 & \text{if} & (\alpha) = (0) \\ 0 & \text{if} & (\alpha) \neq (0). \end{matrix}$$

Now we state and prove the following general theorem.

THEOREM. *Let* $(\alpha)_i = (\alpha_{i1}^{r_{i1}}\alpha_{i2}^{r_{i2}} \cdots \alpha_{ia_i}^{r_{ia_i}})$ *and* $(\nu)_i = (1^{\lambda_{i1}}2^{\lambda_{i2}} \cdots n_i^{\lambda_{in_i}})$ *where* λ_{ij} *and* n_i *are finite and* $1 \leq r_{ij} \leq \lambda_{ij}$, $1 \leq a_i \leq n_i$ *for* $i = 1, 2, \cdots, m$ *and* $j = 1, 2, \cdots, n_i$. *Then for any two functions of the m combinations (with repetitions),* $(\alpha)_1, (\alpha)_2, \cdots, (\alpha)_m$ *out of* $(\nu)_1, (\nu)_2, \cdots, (\nu)_m$, *the two sets of formulas:*

$$(1) \quad F((\alpha)_1, (\alpha)_2, \cdots (\alpha)_m) = \sum_{(\beta)_i \in (\nu)_i - (\alpha)_i} f((\alpha)_1 + (\beta)_1, (\alpha)_2 + (\beta)_2, \cdots, (\alpha)_m + (\beta)_m)$$

and

$$f((\alpha)_1, (\alpha)_2, \cdots, (\alpha)_m)$$

(2) $$= \sum_{(\beta)_i \in (\nu)_i - (\alpha)_i} \left[\prod_{i=1}^{m} \mu((\beta)_i) \right] F((\alpha)_1 + (\beta)_1, (\alpha)_2 + (\beta)_2, \cdots, (\alpha)_m + (\beta)_m)$$

are equivalent.

PROOF. To deduce (2) from (1)

$$\sum_{(\beta)_i \in (\nu)_i - (\alpha)_i} \left[\prod_{i=1}^{m} \mu((\beta)_i) \right] F((\alpha)_1 + (\beta)_1, \cdots, (\alpha)_m + (\beta)_m)$$

$$= \sum_{(\beta)_i \in (\nu)_i - (\alpha)_i} \left[\prod_{i=1}^{m} \mu((\beta)_i) \right] \sum_{(\gamma)_i \in (\nu)_i - (\alpha)_i - (\beta)_i} \cdot f((\alpha)_1 + (\beta)_1 + (\gamma)_1, \cdots, (\alpha)_m + (\beta)_m + (\gamma)_m)$$

$$= \sum_{(\delta)_i \in (\nu)_i - (\alpha)_i} f((\alpha)_1 + (\delta)_1, \cdots, (\alpha)_m + (\delta)_m) \sum_{(\gamma)_i \in (\delta)_i} \prod_{i=1}^{m} \mu((\delta)_i - (\gamma)_i).$$

Evidently we have

$$\sum_{(\gamma)_i \in (\delta)_i} \prod_{i=1}^{m} \mu((\delta)_i - (\gamma)_i) = \prod_{i=1}^{m} \left\{ \sum_{(\gamma)_i \in (\delta)_i} \mu((\delta)_i - (\gamma)_i) \right\}$$

$$= \prod_{i=1}^{m} \left\{ \sum_{(\gamma)_i \in (\delta)_i} \mu((\gamma)_i) \right\} = \begin{matrix} 1 & \text{if } (\delta)_i = (0) \text{ for } i = 1, \cdots, m \\ 0 & \text{otherwise} \end{matrix}$$

by the property of the μ-function. Hence the preceding sum reduces to $f((\alpha)_1, \cdots, (\alpha)_m)$ in accord with (2).

(1) is deduced from (2) in a similar way.

Although the general case is not without importance in the treatment of several sets of events,[3] we shall for the sake of convenience restrict ourselves to the special case $m = 1$.

In order to apply these formulas we must first introduce combinations with repetitions into the theory of arbitrary events. This can be done in various ways. Firstly, we may consider the number of occurrences of each event in a given time-interval or in a series of trials not necessarily independent. Secondly, we may regard each event as possessing various degrees of intensity. If the event E_i occurs r_i times in a given time-interval or occurs with r_i degrees of intensity, we write it as $E_i^{r_i}$. Hereafter we shall make use of the first interpreta-

[3] Cf. FRÉCHET, Loc. Cit. pp. 50–52; also, K. L. CHUNG, "Generalization of Poincaré's formula in the theory of probability," *Annals of Math. Stat.*, Vol. 14 (1943). We may note that our general theorem may be used to give another proof of the generalized Poincaré's formula for several sets of events.

tion and we shall assume that the maximum number of occurrences of each event is finite:

$$0 \leqq r_i \leqq \lambda_i, \qquad i = 1, \cdots, n.$$

We define

$p[E_1^{r_1} \cdots E_n^{r_n}] = p[(\nu^r)] =$ The probability that E_i occurs exactly r_i times in the given time-interval.

$p(E_1^{r_1} \cdots E_n^{r_n}) = p((\nu^r)) =$ The probability that E_i occurs at least r_i times in the given time-interval.

These quantities play the same rôle as the $p[(\alpha)]$'s and $p((\alpha))$'s in the ordinary theory. Evidently the probability of every complex event in question can be expressed as the sum of certain $p[(\nu^r)]$'s. To prove that the $p((\nu^r))$'s also form a fundamental system of quantities we have only to express $p[(\nu^r)]$'s in terms of the $p((\nu^r))$'s. This is given immediately by an application of the general theorem with $m = 1$. For we have in an obvious way

$$p(E_1^{r_1} \cdots E_n^{r_n}) = \sum_{r_i \leqq t_i \leqq \lambda_i} p[E_1^{t_1} \cdots E_n^{t_n}]$$

or

$$(3) \qquad p((\nu^r)) = \sum_{(\nu^s) \in (\nu^\lambda) - (\nu^r)} p[(\nu^r) + (\nu^s)] = \sum_{(\nu^s) \in (\nu^{\lambda - r})} p[(\nu^{r+s})].$$

Hence we obtain the inversion

$$(4) \qquad p[(\nu^r)] = \sum_{(\nu^s) \in (\nu^\lambda) - (\nu^r)} \mu((\nu^s)) p((\nu^r) + (\nu^s)).$$

Let (α') denote a running combination without repetitions. Then since $\mu((\nu^s)) = 0$ unless (ν^s) is a (ν'),

$$(4') \quad p[(\nu^r)] = \sum_{(\alpha') \in (\nu^{\lambda - r})} \mu((\alpha')) p((\nu^r) + (\alpha')) = \sum_{(\alpha') \in (\nu^{\lambda - r})} (-1)^a p((\nu^r) + (\alpha'))^{\bullet}$$

The set of formulas (3) and (4) generalize (3.1) and (4.1).

Corresponding to the $p_{[a]}((\nu))$ for the ordinary events we define for $a + b + \cdots = n$ and $r, s, \cdots$ all distinct:

$p_{[a]^r, [b]^s, \cdots}(E_1^{\lambda_1} \cdots E_n^{\lambda_n}) =$ The probability that among n events $E_1, E_2, \cdots, E_n$ exactly a events occur r times, exactly b events occur s times and so on.

By (4) we easily obtain

$$(5) \qquad \begin{aligned} &p_{[a]^r, [b]^s, \cdots}((\nu^\lambda)) \\ &= \sum_S \sum_{(\nu^x) \in (\nu^\lambda) - ((\alpha)^r + (\beta)^s + \cdots)} \mu((\nu^x)) p((\nu^x) + (\alpha)^r + (\beta)^s + \cdots) \end{aligned}$$

where $(\alpha)^r = (E_{\alpha_1}^r \cdots E_{\alpha_a}^r)$, $(\beta)^s = (E_{\beta_1}^s \cdots E_{\beta_b}^s)$, $\cdots$ and the first summation is a symmetric sum which extends to all $n!/a!b! \cdots$ different combinations $(\alpha_1 \cdots \alpha_a)$, $(\beta_1 \cdots \beta_b)$, $\cdots$ out of $(\nu) = (1, 2 \cdots n)$.

The equality (5) is obviously a generalization of Poincaré's formula.

Similarly for the probabilities in the definition of which the word "exactly" is sometimes substituted for the words "at least." Of course we can express all of them in terms of the $p[(\nu^r)]$'s or of the $p((\nu^r))$'s. However elegant formulas such as in the ordinary theory seem to be lacking.

Finally, we may also consider conditions of existence for the $p[(\nu^r)]$'s and the $p((\nu^r))$'s. For the former system the conditions are that they be all non-negative and that their sum be 1. For the latter system, the conditions are given by (4′), viz. for every $(\nu^r) \epsilon (\nu^\lambda)$,

$$\sum_{(\alpha') \epsilon (\nu^{\lambda - r})} \mu((\alpha'))p((\nu^r) + (\alpha)) \geqq 0.$$

These conditions are necessary and sufficient since (3) and (4) are equivalent.

ON THE MECHANICS OF CLASSIFICATION

By Carl F. Kossack

University of Oregon

1. Introduction. Wald[1] has recently determined the distribution of the statistic U to be used in the classification of an observation, z_i $(i = 1, 2, \cdots, p)$, as coming from one of two populations. He also determined the critical region which is most powerful for such a classification. It is the purpose of this paper to show how such a classification statistic under the assumption of large sampling can be applied in an actual problem and to present a systematic approach to the necessary computations.

The data used in this demonstration are those which were obtained from the A.S.T.P. pre-engineering trainees assigned to the University of Oregon. The problem considered is that of classifying a trainee as to whether he will do unsatisfactory or satisfactory work[2] in the first term mathematics course (Intermediate Algebra). The variables used in the classification are: (1) A Mathematics Placement Test Score. This is the score obtained by the trainee on a fifty-minute elementary mathematics test (including elementary algebra). The test was given to each trainee on the day that he arrived on the campus. (2) A High School Mathematics Score. A trainee's high school mathematics record was made into a score by giving 1 point to students who had had no high school algebra, 2 points to students with an F in first-year, high-school algbra and no second-year algebra, 3 points for a D, $\cdots$, 10 points for an average grade of A in first- and second-year algebra. (3) The Army General Classification Test Score. An individual needed a score of 115 or better in order to be assigned to the A.S.T.P. These data were obtained for 305 trainees along with the actual

[1] Abraham Wald, "On a statistical problem arising in the classification of an individual into one of two groups," *Annals of Math. Stat.*, Vol. 15, (1944), No. 2.

[2] Unsatisfactory work was defined as a grade of F or D in the course (failure or the lowest passing grade).