IMPOSSIBLE DEVELOPMENTS II

Two menus M_1 and M_2 with respectively n_1 and n_2 letters on each, of which C are common to both, are run in the same wheel order giving r_1 and r_2 stecker stops respectively. M is req^d to find an estimate for the expected No. of pairs of stops of M_1 and M_2 respectively which give consistent stecker. If it is not possible to find an accurate measure of this number and our aim is $\therefore$ to find a convenient overestimate. We shall assume that punched cards will be produced by the ten Mammoths running the menus giving the stecker of every letter of the alphabet as far as they are determined at each stop, and the two packs of cards turned out by the two machines are sorted against one another, and two cards are paired off if they agree for every letter of the alphabet on which they both have stecker punched.

Notation: A Latin capital S will denote a set of letters. [S] will denote the no. of letters in the set. ST the set of letters common to S and T. S + T the set of letters belonging to either S or T. (i.e. the usual set identical notation with the addition of [S])

The principle in work is as follows. If we have a stop on M_1 we have the stecker given to us for the n_1 letters of the menu. Some of these stecker (in number $\geq n_1$) will be different from those in the menu. The we call the "stop letters", the sets being denoted respectively by S_1 and L_1. For a stop on M_2 we have correspondingly S_2 and L_2. C denotes the set L_1L_2. Suppose $S_1 + L_1$ and $S_2 + L_2$ have K letters in common. This gives at least K/2 letters from different stecker pairs of M_1, their stecker are determined by the stop. The odds against these letters having the same stecker for our stop on M_2 are then $26^{-K/2}$. $\therefore$ the chances of the two stops being consistent are less than or equal to $26^{-K/2}$. If p is the probability that two stops are consistent, the expected no. of consistent stops is pr_1r_2. Our aim is to find an overestimate for this number, which we do by drawing an overestimate for K/2. This overestimate (for K/2) I shall call λ or $\lambda(n_1, n_2, c)$ to emphasise that it is a function of n_1, n_2 and c.

Consider first the set S_1 (i.e. the stop letters for a stop on M_1). Then $[S_1] \geq n_1$, and is actually equal to n_1 if the stop gives no self stecker or confirmation which it does in the great majority of cases. It seems reasonable then to assume that ¾ n_1 will actually be an overestimate for the average no. of stop letters. We shall then work out the probability of two stops agreeing taking $[S_1] = \frac{3n_1}{4}$ $[S_2] = \frac{3n_2}{4}$

We wish to find now the expected no. of letters common to $S_1 + L_1$ and $S_2 + L_2$. To do this in appeal to the following theorem. If from a set of N different objects we take two samples of m and n objects respectively the expected no. of objects common to the two samples is mn/n . Now S_1 and L_1 are mutually exclusive and so on S_2 and L_2. $\therefore [(S_1 + L_1)(S_2 + L_2) = [S_1S_2] + [L_1S_2] + [L_2S_1] + {}^3[L_1L_2]$ $\{\propto\}$ now $[L_1L_2] = C$ by definition.

$[L_1S_2] = [(L_1 - C)S_2]$ for $CS_2CL_2S_2 = 0$

now L_1 - C and S_1 are both picked from 26 $-n_2$ letters not appearing in L_2. $\therefore$ the expected no. of common letters is

$$\frac{(n_1 - C)¾n_2}{26 - n_2}$$

i.e. $[L_1S_2] = \frac{3}{4} \frac{n_2(n_1-C)}{26 - n_2}$ and similarly $[L_2S_1] \frac{n_1(n_2-C)}{26 - n_1}$

$[S_1S_2]$ is harder to calculate as S_1 is selected from A - L_1 (when A denotes the whole alphabet) and S2 for A - L_2. However, all common letters of S_1 and S_2 are in A - (L_1 +L_2). S_1 is picked from A - L_1; $\therefore$ it will be an average when