Ann. Inst. Statist. Math.
36 (1984), Part A, 475-479

MOMENTS OF A STATISTIC CAUSED BY RANDOM COMBINATIONS OR RANDOM MATINGS

SHINJI AZUMA, KENJI HAYASHI AND AKIO KUDÔ

(Received Aug. 31, 1983; revised Nov. 14, 1983)

Summary

Given two sets of size k, $\{\alpha_1,\cdots,\alpha_k\}$ and $\{\beta_1,\cdots,\beta_k\}$, there are $k!$ possible combinations of these two $\{(\alpha_1,\beta_{l_1}),\cdots,(\alpha_k,\beta_{l_k})\}$, and suppose there is apriori given a number corresponding to the partnership (α_i,β_j). The average of the numbers corresponding to $\{(\alpha_1,\beta_{l_1}),\cdots,(\alpha_k,\beta_{l_k})\}$ is a random variable, and this paper presents the first five moments of the average, and an application in the study of an isolated human population is demonstrated.

1. Introduction

This paper is motivated by a terminology "random mating" in population genetics. Let $\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$ and $\{\beta_1,\beta_2,\cdots,\beta_k\}$ be two sets. By "random combination" we mean that each α_i finds partner from one of the $\{\beta_1,\beta_2,\cdots,\beta_k\}$ randomly, and for each potential partnership (α_i,β_j) there is given apriori a number A_{ij} characterizing the partnership. The mathematical formulation is the following.

Let A_{ij} be a $k\times k$ matrix. For a permutation $l=(l_1,l_2,\cdots,l_k)$ of $(1,2,\cdots,k)$, we define

$$A(l)=k^{-1}\sum A_{il_i}\,.$$

Let $L=(L_1,L_2,\cdots,L_k)$ be a random permutation taking all possible $k!$ values with probability $1/k!$. In this paper, we are concerned with the moments of the random variable $A(L)$.

In the particular case $A_{ij}=ij$, we have $A(l)=k^{-1}\sum il_i$ so that the $A(L)$ is a generalized Spearman rank correlation coefficient.

Consider a complete block design. Suppose A_{ij} be the yield of the ith variety when planted at the jth plot, $A(l)$ is the average yield in this block for a particular arrangement, and when randomization is

Key words: Random combination, Gram-Chalier expansion, isolate human population, random mating.

carried out $A(L)$ is the average yield. Ogawa [6] considers the randomization in designs of experiments in this manner and reviewed the papers on the effects of randomization.

In Section 2 we list the first five moments around the mean. They do not come out in nice forms, and it is not easy to check if they are correct at all. In order to check its correctness, we generated $5\times 5=25$ random numbers and we computed all possible $5!=120$ values of $A(l)$ and computed the moments. They agree with the theoretical values. In Section 3, we illustrate an example related to the study of isolated population in human genetics.

2. Moments

The higher moments can be computed by painstaking calculations, and the method is similar to the calculation of the variance of the sample mean obtained by sampling without replacement from a finite population.

Let

$$\bar{\bar{A}}=k^{-2}\sum_\nu\sum_\mu A_{\nu\mu}\,,\qquad \bar{A}_{\nu\cdot}=k^{-1}\sum_\mu A_{\nu\mu}\,,\qquad \bar{A}_{\cdot\mu}=k^{-1}\sum_\nu A_{\nu\mu}$$

$$D_{\nu\mu}=A_{\nu\mu}-\bar{\bar{A}}\,,\qquad D_{\nu\cdot}=\bar{A}_{\nu\cdot}-\bar{\bar{A}}\,,\qquad D_{\cdot\mu}=\bar{A}_{\cdot\mu}-\bar{\bar{A}}\,,$$

then we have

(1) $\quad \mathrm{E}\,(A(L))=\bar{\bar{A}}$

(2) $$(k-1)V(A(L))=(k-1)\,\mathrm{E}\,[(A(L)-\bar{\bar{A}})^2]$$
$$=k^{-2}\sum_\nu\sum_\mu D^2_{\nu\mu}-k^{-1}\sum_\nu D^2_{\nu\cdot}-k^{-1}\sum_\mu D^2_{\cdot\mu}$$
$$=k^{-2}\sum_\nu\sum_\mu(D_{\nu\mu}-D_{\nu\cdot}-D_{\cdot\mu})^2$$

(3) $$(k-1)(k-2)\,\mathrm{E}\,[(A(L)-\bar{\bar{A}})^3]$$
$$=k^{-2}\sum_\nu\sum_\mu D^3_{\nu\mu}+2k^{-1}\sum_\nu D^3_{\nu\cdot}+2k^{-1}\sum_\mu D^3_{\cdot\mu}$$
$$+6k^{-2}\sum_\nu\sum_\mu D_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu}-3k^{-2}\sum_\nu\sum_\mu D^2_{\nu\mu}(D_{\nu\cdot}+D_{\cdot\mu})$$
$$=k^{-2}\sum_\nu\sum_\mu(D_{\nu\mu}-D_{\nu\cdot}-D_{\cdot\mu})^3$$

(4) $$(k-1)(k-2)(k-3)\,\mathrm{E}\,[(A(L)-\bar{\bar{A}})^4]$$
$$=-6k^{-1}(\textstyle\sum_\nu D^4_{\nu\cdot}+\sum_\mu D^4_{\cdot\mu})-24k^{-2}(\sum_\nu\sum_\mu D_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu}(D_{\nu\cdot}+D_{\cdot\mu}))$$
$$+12k^{-2}\textstyle\sum_\nu\sum_\mu D^2_{\nu\mu}(D^2_{\nu\cdot}+D^2_{\cdot\mu})-12k^{-3}\sum_\nu(\sum_\mu D_{\nu\mu}D_{\cdot\mu})^2$$
$$-12k^{-3}\textstyle\sum_\mu(\sum_\nu D_{\nu\mu}D_{\nu\cdot})^2+(k+1)k^{-3}\sum_\nu\sum_\mu D^4_{\nu\mu}$$
$$+12(k+1)k^{-3}\textstyle\sum_\nu\sum_\mu D^2_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu}-4(k+1)k^{-3}\sum_\nu\sum_\mu D^3_{\nu\mu}(D_{\nu\cdot}+D_{\cdot\mu})$$
$$+3k^{-1}(\textstyle\sum_\nu D^2_{\nu\cdot}+\sum_\mu D^2_{\cdot\mu})^2-6(k-2)k^{-3}(\sum_\nu\sum_\mu D^2_{\nu\mu}(\sum_\nu D^2_{\nu\cdot}+\sum_\mu D^2_{\cdot\mu}))$$
$$+12(k-1)k^{-4}(\textstyle\sum_\nu(\sum_\mu D^2_{\nu\mu})(\sum_\mu D_{\nu\mu}D_{\cdot\mu}))$$
$$+12(k-1)k^{-4}(\textstyle\sum_\mu(\sum_\nu D^2_{\nu\mu})(\sum_\nu D_{\nu\mu}D_{\nu\cdot}))$$

$$
\begin{aligned}
&-3(k-1)k^{-4}\big(\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2)^2+\sum_\mu(\sum_\nu D_{\nu\mu}^2)^2\big)\\
&+3(k^2-3k+1)k^{-5}(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2)^2+6k^{-5}\sum_\mu\sum_\kappa(\sum_\nu D_{\nu\mu}D_{\nu\kappa})^2
\end{aligned}
$$

$$
\begin{aligned}
(5)\qquad &(k-1)(k-2)(k-3)(k-4)\,\mathrm{E}\,[(A(L)-\bar{\bar{A}})^5]\\
&=4k^{-1}[6(\textstyle\sum_\nu D_{\nu\cdot}^5+\sum_\mu D_{\cdot\mu}^5)-5(\sum_\nu D_{\nu\cdot}^3+\sum_\mu D_{\cdot\mu}^3)(\sum_\nu D_{\nu\cdot}^2+\sum_\mu D_{\cdot\mu}^2)]\\
&\quad+60k^{-2}[2\textstyle\sum_\nu\sum_\mu D_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu}(D_{\nu\cdot}^2+D_{\nu\cdot}D_{\cdot\mu}+D_{\cdot\mu}^2)\\
&\qquad-(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu})(\sum_\nu D_{\nu\cdot}^2+\sum_\mu D_{\cdot\mu}^2)]\\
&\quad+120k^{-3}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}D_{\cdot\mu}^2\sum_\mu D_{\nu\mu}D_{\cdot\mu})+\sum_\mu(\sum_\nu D_{\nu\mu}D_{\nu\cdot}^2\sum_\nu D_{\nu\mu}D_{\nu\cdot})\\
&\qquad+\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}D_{\cdot\mu})^2D_{\nu\cdot}+\sum_\mu(\sum_\nu D_{\nu\mu}D_{\nu\cdot})^2D_{\cdot\mu}]\\
&\quad+20(k-3)k^{-3}(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2)(\sum_\nu D_{\nu\cdot}^3+\sum_\mu D_{\cdot\mu}^3)\\
&\qquad-60k^{-2}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2(D_{\nu\cdot}^3+D_{\cdot\mu}^3)\\
&\quad+30(k-2)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2(D_{\nu\cdot}+D_{\cdot\mu})(\sum_\nu D_{\nu\cdot}^2+\sum_\mu D_{\cdot\mu}^2)\\
&\quad-60(k+2)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2D_{\nu\cdot}D_{\cdot\mu}(D_{\nu\cdot}+D_{\cdot\mu})\\
&\qquad+20(k+2)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^3(D_{\nu\cdot}^2+D_{\cdot\mu}^2)\\
&\quad-10(k-2)^2k^{-4}(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^3)(\sum_\nu D_{\nu\cdot}^2+\sum_\mu D_{\cdot\mu}^2)\\
&\qquad+20(k+5)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^3D_{\nu\cdot}D_{\cdot\mu}\\
&\quad+60(k-3)k^{-4}(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2)(\sum_\nu\sum_\mu D_{\nu\mu}D_{\nu\cdot}D_{\cdot\mu})\\
&\quad-60k^{-3}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2D_{\cdot\mu}\sum_\mu D_{\nu\mu}D_{\cdot\mu})+\sum_\mu(\sum_\nu D_{\nu\mu}^2D_{\nu\cdot}\sum_\nu D_{\nu\mu}D_{\nu\cdot})]\\
&\quad-60(k-2)k^{-4}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2\sum_\kappa D_{\nu\kappa}D_{\cdot\kappa}^2)+\sum_\mu(\sum_\nu D_{\nu\mu}^2\sum_\nu D_{\nu\mu}D_{\nu\cdot}^2)]\\
&\quad-120(k-1)k^{-4}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2\sum_\mu D_{\nu\mu}D_{\cdot\mu})D_{\nu\cdot}\\
&\qquad+\textstyle\sum_\mu(\sum_\nu D_{\nu\mu}^2\sum_\nu D_{\nu\mu}D_{\nu\cdot})D_{\cdot\mu}]\\
&\quad+120k^{-4}\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}D_{\kappa\cdot}\sum_\kappa D_{\kappa\mu}D_{\kappa\nu}D_{\cdot\nu})\\
&\qquad-5(k+5)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^4(D_{\nu\cdot}+D_{\cdot\mu})\\
&\quad+20(k-1)k^{-4}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^3\sum_\mu D_{\nu\mu}D_{\cdot\mu})+\sum_\mu(\sum_\nu D_{\nu\mu}^3\sum_\nu D_{\nu\mu}D_{\nu\cdot})]\\
&\quad-30(k^2-4k+2)k^{-5}(\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^2)\sum_\nu\sum_\mu D_{\nu\mu}^2(D_{\nu\cdot}+D_{\cdot\mu})\\
&\quad+30(k-1)k^{-4}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2)^2D_{\nu\cdot}+\sum_\mu(\sum_\nu D_{\nu\mu}^2)^2D_{\cdot\mu}\\
&\qquad+\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^2\sum_\mu D_{\nu\mu}^2D_{\cdot\mu})+\sum_\mu(\sum_\nu D_{\nu\mu}^2\sum_\nu D_{\nu\mu}^2D_{\nu\cdot})]\\
&\quad-60(k-2)k^{-5}[\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}^2D_{\kappa\nu}\sum_\kappa D_{\kappa\nu}D_{\kappa\cdot})\\
&\qquad+\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}^2\sum_\kappa D_{\kappa\mu}D_{\kappa\nu}D_{\cdot\nu})]\\
&\quad-120k^{-5}[\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}D_{\kappa\nu}\sum_\kappa D_{\kappa\mu}D_{\kappa\nu}D_{\kappa\cdot})+\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}D_{\kappa\nu})^2D_{\cdot\mu}]\\
&\quad+(k+5)k^{-3}\textstyle\sum_\nu\sum_\mu D_{\nu\mu}^5+10(k^2-5k+5)k^{-5}\sum_\nu\sum_\mu D_{\nu\mu}^3\sum_\nu\sum_\mu D_{\nu\mu}^2\\
&\quad-10(k-1)k^{-4}[\textstyle\sum_\nu(\sum_\mu D_{\nu\mu}^3\sum_\mu D_{\nu\mu}^2)+\sum_\mu(\sum_\nu D_{\nu\mu}^3\sum_\nu D_{\nu\mu}^2)]\\
&\quad+30(k-3)k^{-5}\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}^2\sum_\kappa D_{\kappa\mu}D_{\kappa\nu}^2)\\
&\qquad+60k^{-5}\textstyle\sum_\nu\sum_\mu(\sum_\kappa D_{\kappa\mu}^2D_{\kappa\nu}\sum_\kappa D_{\kappa\mu}D_{\kappa\nu})\ .
\end{aligned}
$$

3. An application

A fairly extensive survey was conducted on an isolated population of Japan (I. Nishigaki [5]). In this survey all the ancestors of the existing population are traced back as far as possible: at maximum ten generations back. On the other hand those ancestors migrated from outside could not be traced back further, and we assumed they were unrelated in themselves and with those in the population.

We stored the information including all the offspring-parents relationships in a computer file. We selected all those who were 20–29 years of age at the time of survey, and there were 52 families. For the $52\times 52=2704$ pairs of males and females we computed the coefficients consanguinity forming 52×52 matrix.

The coefficient of consanguinity is a measure of closedness of relationship between two individuals. This is defined as the probability that two homologous genes drawn at random, one from each of the two individuals, will be identical (see J. F. Crow and M. Kimura [1], p. 68). In case of cousins it is 1/16, and it is 1/8 for an uncle and a nephew. It is the inbreeding coefficient of potential offspring of them.

For the method of calculation of the coefficient, the readers are referred to J. F. Crow and M. Kimura [1] and R. Elandt Johnson [2].

The observed value of average coefficient of consanguinity for these 52 couples is $A(l)=0.01949363$ whereas the value of $\bar{\bar{A}}=0.007933126$. The higher moments are also computed. After normalizing the variable

$$Y=(A(L)-\bar{\bar{A}})/V(A(L))^{1/2}$$

we have

$$Y=(A(l)-\bar{\bar{A}})/V(A(L))^{1/2}=3.127237$$

$$\mathrm{E}(Y)=0\,,\qquad V(Y)=1\,,\qquad M_3=\mathrm{E}(Y^3)=0.8925\ \text{(skewness)}$$

$$M_4=\mathrm{E}(Y^4)=3.884\ \text{(kurtosis)}\,,\quad \text{and}\quad M_5=\mathrm{E}(Y^5)=9.683\,.$$

In order to test the difference between $A(l)$ and $\bar{\bar{A}}$, we used the Gram-Chalier expansion formula (6.33) in Kendall ([4], p. 169) to obtain

$$\begin{aligned}\mathrm{P}(Y\leqq y)&=\Phi(y)-\phi(y)[M_3(y^2-1)/6+(M_4-3)(y^3-y)/24\\&\quad+(M_5-10M_3)(y^4-6y^2+3)/120+\cdots]\\&=0.9991-0.0039-0.0023-0.0008=0.9921\,,\end{aligned}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions.

The outcome of the analysis indicates that the hypothesis of random mating in the sense of this paper is rejected at 1% level of significance.

4. Discussions

These days there is a powerful computer program called "REDUCE 2". We examined the manual (A. C. Hearn [3]) and concluded that the calculation of these moments in this paper is beyond its capability.

Apparently some approximation formula for the moments are needed. In the present application among $2704=52^2$ A_{ij}'s 2058 are zero and the minimum positive value is 2^{-9} and the maximum is 0.28125. To work out an adequate approximation formula is left open.

The result of our analysis well supports the description of Schull and Neel ([7], p. 17–20) about the cause and attitude toward consanguineous marriages in Japan. They argue that there were several factors which made Japanese people prefer consanguineous marriages.

Acknowledgements

The authors are grateful to those who draw the authors' attention to the computer program "REDUCE 2". Dr. S. Konishi and Dr. N. Niki kindly provided us with a lot of information including the manual. Finally their thanks are due to a referee who suggested improvements in presentation.

SCIENCE UNIVERSITY OF TOKYO
SCIENCE UNIVERSITY OF TOKYO
KYUSHU UNIVERSITY

REFERENCES

[1] Crow, J. F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory*, Hamper and Row, New York.
[2] Elandt Johnson, R. (1971). *Probability Models and Statistical Method in Genetics*, John Wiley and Sons, New York.
[3] Hearn, A. C. (1973). *REDUCE 2 User's Manual Second Edition*, University of Utah.
[4] Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, 4th ed., Vol. 1, Charles Griffin, London.
[5] Nishigaki, I. (1978). Studies on population genetics in the isolates, (VIII) Tomiyama, with special reference to the enzyme activity of polymorphic red cell enzymes, *Japanese Jour. of Constitutional Medicine*, **42**, 53-89.
[6] Ogawa, J. (1982). The Fisher experimental randomization, *Jour. Japan Statist. Soc.*, **12**, No. 1, 1-24.
[7] Schull, W. J. and Neel, J. V. (1965). *The Effect of Inbreeding on Japanese Children*, Harper and Row, New York.