Twin Research and Human Genetics (2025), **28**, 466–467
doi:10.1017/thg.2025.10035

CAMBRIDGE UNIVERSITY PRESS

Article

Urn Models have a Place in Genetics

Alan Stark
School of Mathematics and Statistics, The University of Sydney

Abstract

A simple application of urn models is useful in spelling out the way in which, in abstraction, if the parents are distributed in Hardy-Weinberg form, that form is reproduced in offspring with nonrandom mating. A measure of divergence from random mating is proposed and illustrated by numerical examples. The scope of urn models of stochastic processes in population genetics theory is set out.

Keywords: Mendelian segregation; Hardy-Weinberg law; Pseudo-random mating; Urn models

(Received 26 September 2025; revise received 26 September 2025; accepted 12 November 2025; First Published online 16 January 2026)

Johnson and Kotz (1977) use urn models to derive results in probability theory. Here, urns are used to illustrate a facet of the Hardy-Weinberg (HW) model as applied to a diploid population in respect of a single locus with two alleles A and B. While the theory can be explained algebraically, there is explanatory value in the numerical approach used here.

The article gives a simple numerical procedure by which anyone can illustrate how, by imposing a constraint on the table of mating frequencies, HW frequencies are reproduced in the offspring with pseudo-random mating. The constraint is a property of random mating. The completeness and universality of the process is achieved by a simple application of urn models. A measure of divergence from random mating is proposed and illustrated. Some examples of urn models applied to stochastic processes in genetics are outlined.

In the illustration to follow, the population is in HW form in both sexes with arbitrary genotype numbers in respective genotypes AA, BB, AB, numbered respectively 1, 2, 3. The main point of interest is to follow the outcome, that is, the numbers of offspring from suggested couplings. It is shown by example that HW frequencies are reproduced if matings follow a particular constraint, which is also obeyed by random mating.

Taking random mating as the main point of focus, the simple measure of divergence from random mating is proposed and exemplified. While the demonstration is not rigorous, it can be tried easily with other sets of numbers. The method can be extended to a locus with more than two alleles.

Illustration

Table 1 gives the array of numbers of mating couples under random mating. The population has $M_1 = 100$ males of type AA, $M_2 = 225$ males of type BB, and $M_3 = 300$ males of type AB. The population has an identical array of females with numbers F_1, F_2, and F_3. The total number of males is $N = 625$, as is the number of females. C_{ij} is the number of couples in urn U_{ij}, that is, of type i in male and type j in female. $C_{ij} = C_{ji}$. N is the total number of couples. Gene A frequency is $q = 0.4$. Gene B frequency is $p = .6$. The marginal frequencies are Nq^2, Np^2, and $2Npq$.

The marginal arrays are in HW form corresponding to gene frequencies q and $1 - q$. The numbers in the body of Table 1 conform to 'random mating', being Nq^4, and so forth.

The offspring distribution is calculated from strictly Mendelian prescription with one child of each sex per couple. The number of male/female children of type AA is $C_{11} + C_{13}/2 + C_{31}/2 + C_{33}/4 = 100$, that is, equal to the number of parents of each sex. The other genotypes are reproduced.

The main point to notice is that $C_{33} = 4C_{12}$, so that C_{33} is a multiple of 4 for the illustration. A useful additional feature for the illustration is that C_{13} and C_{23} are even numbers.

The main point of this note is to illustrate that, starting from the same parental distribution, there is an infinite number of couplings that reproduce the parental distribution. This is done by the artifice of moving balls between urns. The urns are identified by the symbols U_{ij}, which hold the number of balls C_{ij}.

The process can be done in various ways, but a starting point is to vary C_{33} and move balls between urns in the same row or column keeping the marginal number constant. The number of degrees of freedom in the setup is one: varying one of the C_{ij} fixes the others automatically.

The proposed measure of divergence from random mating of a variant from the original table is $\Sigma_{ij}|C_{ij} - H_{ij}|/N$, where H refers to the table of randomly mating couples. Figure 1 displays the measure of divergence for an arbitrarily chosen group of couplings with the marginal array of parents with the same parental numbers as those in Table 1. There is a discernible pattern around the zero arising from random mating.

Table 2 is an example of mating whose measure of divergence is 0.5632. Table 3 is an extreme example of mating choice whose measure of divergence is 0.9984. There is no limit to the number of numerical examples starting from an array of parents in HW form.

Corresponding author: Alan Stark; Email: alans@exemail.com.au
Cite this article: Stark A. (2025) Urn Models have a Place in Genetics. *Twin Research and Human Genetics* **28**: 466–467, https://doi.org/10.1017/thg.2025.10035

© The Author(s), 2026. Published by Cambridge University Press on behalf of International Society for Twin Studies. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Check for updates

Table 1. An example of random mating

Male/female	AA	BB	AB	Total
AA	16	36	48	100
BB	36	81	108	225
AB	48	108	144	300
Total	100	225	300	625

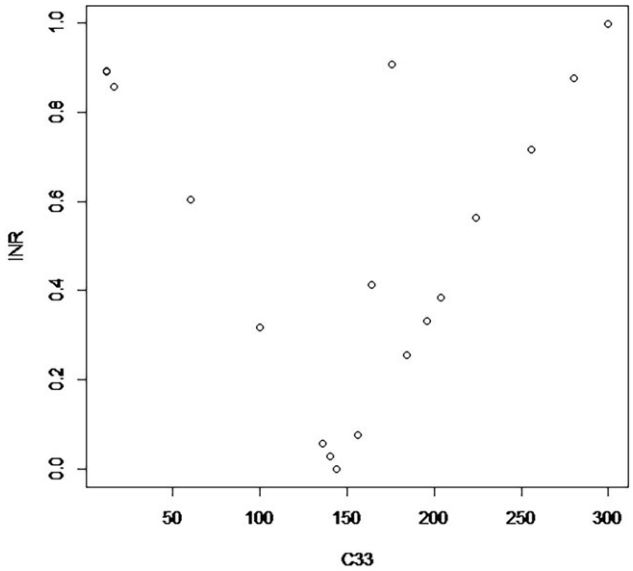


Figure 1. Display of calculated indices of divergence from random mating plotted against C_{33}. These relate to the original population of males and females in Table 1.

Table 2. An example of nonrandom mating that reproduces Hardy-Weinberg frequencies. The measure of divergence from random mating is 0.5632

Male/female	AA	BB	AB	Total
AA	0	56	44	100
BB	56	137	32	225
AB	44	32	224	300
Total	100	225	300	625

Table 3. An extreme example of nonrandom mating that reproduces Hardy-Weinberg frequencies. The measure of divergence from random mating is 0.9984

Male/female	AA	BB	AB	Total
AA	25	75	0	100
BB	75	150	0	225
AB	0	0	300	300
Total	100	225	300	625

For most practical purposes attention would be focused on situations of divergence close to zero, so Figure 1 is only an indication of its properties and is based on an arbitrary set of cases starting from the numbers given in the margins of Table 1. The values of divergence given with Tables 2 and 3 fit on a straight line radiating to the right from the zero value of Table 1. More aberrant values can be constructed, as illustrated in Figure 1.

The illustration is contradictory, being a concrete example of an abstraction. Applications of the Hardy-Weinberg model to real problems are usually accompanied by caveats. While no one expects the abstraction to be followed precisely in practice it is of interest that matings of low divergence can reproduce the Hardy-Weinberg distribution. The user can select the gene frequency and the size of the population to achieve a desired level of precision and convenience.

Stochastic Processes

Johnson and Kotz (1977) have a section on applications of urn models in genetics. This was based on a draft supplied by W. J. Ewens. The approach is more sophisticated dealing with random fluctuations in the genetic composition of populations and sampling of populations. Johnson and Kotz (p. 239) give the basic formula for the change in a population from one generation to the next:

$$P_{ij} = \binom{m}{j}\left(\frac{i}{m}\right)^{j}\left(1-\frac{i}{m}\right)^{m-j} \qquad (*)$$

Equation (*) gives the probability that a population of size m with i genes of the first allele changes in one generation to one having j copies of the allele. This is derived by drawing with replacement genes from one urn and putting them in a second urn. Johnson and Kotz use (*) to derive the probability that the population will eventually contain genes of the first type only as im^{-1}.

Johnson and Kotz (1977) consider the way in which, as sampling of the kind described is repeated, only one of the original two alleles becomes fixed. They state that no simple explicit formula has been found for the mean number of urn replacements taken to achieve uniformity. Johnson and Kotz turn their attention to the diploid case, considering the probability that, in the sequence of draws from urns, two balls will be of the same allelic type. Their calculation shows that the loss of variation is slow, with the comment 'this fact is of great importance in reconciling the Mendelian mechanism of inheritance with the Darwinian theory of natural selection, since it shows that genetic variation tends to be preserved for long time intervals and thus allows ample time for selective forces to act' (p. 241). Johnson and Kotz appear to believe that there is plan and purpose in Nature.

Other more complex cases, including provision for mutation, are dealt with by Johnson and Kotz (1977). Since that time much has been written on reconstructing the past history of a population based on a sample of individuals taken at the present time. Since much of this is done without knowledge of the characteristics of the ancestry of the sample it is not easy to see how one could apply urn models to mimic the scenario.

Acknowledgment. I thank a reviewer for critical comments and constructive suggestions.

Competing interests. None.

Reference

Johnson, N. L., & Kotz, S. (1977). *Urn models and their application*. John Wiley & Sons.

https://doi.org/10.1017/thg.2025.10035 Published online by Cambridge University Press