EQUATION FROM HENRY HARPENDING

HENRI STEYER

Abstract. In the article from Frank Salter 'Misunderstandings of Kin Selection' in Mankind Quaterly Volume XLVIII 3, Spring 2008 is an appendix from Henry Harpending 'The Apportionment of Variation Within and Among Families' with equations that are discussed here.

Introduction

If in a population an allele for a gene loci has the frequency p its variance is qp. If the population is partitioned in groups, the mean of the variance of the allele within the groups plus the variance of the frequency of the allele among the group is always equal to qp. If the partition is randomly, depending on the size of the population and the size of the groups, their is expected values for the two term of the sum. In the model the randomly repartition is assumed binomial, wich is a good aproximation for a big population.

Main

Definition of the mean $E(X)$ and the variance $varX$ of a random variable X

$$\sum_i P(X = x_i) = \sum_i p_i = 1$$

$$E(X) = \sum_i p_i x_i$$

$$varX = \sum_i p_i(x_i - E(X))^2$$

N is the number of gene and N_A the number of the allele A.
$p = N_A/N \quad q = 1 - p$

$$E(X) = q * 0 + p * 1 = p$$

$$varX = q(0-p)^2 + p(1-p)^2 = qp^2 + pq^2 = qp(q+p) = qp$$

n2010-01-23, 2009-11-01
n for last date of correction.

If the population is divided in groups of n elements there is N/n subsets. i is the number of allele A in the subset and p_i is the percent of subset with i allele A. p_i must meet the two condition

$$\sum_{0\leq i\leq n} p_i = 1$$

$$\sum_{0\leq i\leq n} \frac{N}{n} p_i i = N_A$$

It holds

$$\sum_{0\leq i\leq n} \frac{N}{n} p_i i = N_A \Leftrightarrow \sum_{0\leq i\leq n} p_i i/n = p$$

The variation among groups is:

$$\begin{aligned} varX_a = \sum_{0\leq i\leq n} p_i(i/n - p)^2 &= \sum_{0\leq i\leq n} p_i(i/n)^2 - \sum_{0\leq i\leq n} p_i 2(i/n)p + \sum_{0\leq i\leq n} p_i p^2 \\ &= \sum_{0\leq i\leq n} p_i(i/n)^2 - p^2 = E(X^2) - E(X)^2 \end{aligned}$$

The mean of the variation within groups is:

$$\begin{aligned} E(X_w) = \sum_{0\leq i\leq n} p_i(\frac{i}{n}(\frac{n-i}{n})) &= \sum_{0\leq i\leq n} p_i(i/n) - \sum_{0\leq i\leq n} p_i(i/n)^2 \\ &= E(X) - E(X^2) \end{aligned}$$

As asserted it holds:

$$varX + E(X) = E(X) - E(X)^2 = p - p^2 = pq$$

For a binomial $Bi(n,p)$ and $X = Bi(n,p)/n$

$$(q+p)^n = \sum_{0\leq i\leq n} q^{n-i} p^i \binom{n}{i} = 1$$

$$E(X) = \sum_{1\leq i\leq n} q^{n-i} p^i \binom{n}{i} i/n$$

$$\binom{n}{i} i = \frac{n!}{i!(n-i)!} i = \frac{n!}{(i-1)!(n-i)!} = n\binom{n-1}{i-1}$$

$$E(X) = p \sum_{1\leq i\leq n} \binom{n-1}{i-1} p^{i-1} q^{n-i}$$

$$i - 1 = k$$

$$E(X) = p \sum_{0\leq k\leq n-1} \binom{n-1}{k} p^k q^{(n-1)-k} = p$$

E(X) meets the condition $E(X) = p$ so the variance among the groups is

$$varX = \sum_{0 \leq i \leq n} p^i q^{n-i} \binom{n}{i} (i/n)^2 - p^2$$

The mean of the variance within the groups is

$$E(X) = p - \sum_{0 \leq i \leq n} p^i q^{n-i} \binom{n}{i} (i/n)^2$$

References

[1] Salter, Frank: *Misunderstandings of Kin Selection*, Mankind Quarterly, Volume XLVIII Number 3, (Spring 2008), 337-338.

E-mail address: henri.steyer@freenet.com
URL: http://www.themathematicsofthelife.com