

# The World is Seldom Ideal

Léon Siklóssy

*Abstract: Based on the Hardy-Weinberg Equation, we derive a recurrence relation for the probability of the extinction of the recessive gene in a population after  $m$  generations given a random radial deviation of  $q$  from the ideal in any one generation.*

## Introduction

The Hardy-Weinberg Theorem states that in a reproducing population, with pure dominant, pure recessive, and hybrid genotypes, the percent of each genotype after the first generation will not change in subsequent generations. Given a starting percentage of dominant gene of  $x$ , the ratio in the second

generation would be  $\{x^2, 2 \cdot x \cdot (1 - x), (1 - x)^2\}$  in the  $\{p,q,r\}$  plane, or  $\{x, 2 \cdot x \cdot (1 - x)\}$  in the  $x$ - $y$  plane, where the **x-axis** is the percent of dominant gene in total population regardless of genotype and the **y-axis** is the percent of hybrid genotype in total population. (Note: The  $x$ -axis does not represent the pure dominant genotype, but rather, the percentage of dominant genes in the population. The actual percentage of pure dominant genotype in the total population can be computed as  $x - (y / 2)$ .)

This assumes that there is no bias between genotypes in life span, choice of mates, ability to reproduce, etc.. If such a bias were to exist for whatever reason, instead of the Hardy-Weinberg Theorem applying exactly, the population move slightly from the ideal value given by the Hardy-Weinberg equation in the  $x$ - $y$  plane. In this presentation, the bias will be arbitrarily represented by a circle of radius  $q$  in the  $x$ - $y$  plane around the ideal value gotten from the Hardy-Weinberg equation. The probability of extinction of the recessive gene in a generation  $m$  can be denoted  $P[x,q,m]$ . This can be computed by the following.

## Analysis

### Theorem 1

Given a reproducing population with dominant and recessive genes, with  $x$  being the percentage of dominant gene in the total population.  $P(x,q,m)$ , the probability that the recessive gene will die out in  $m$  generations given a radial deviation of  $q$  in any one generation, can be computed by the following formulae:

$$P(x, q, m) = \begin{cases} 0 & \text{for } x \leq 0 \text{ or } x \leq 1 - m \cdot q \\ 1 & \text{for } x \geq 1 \\ M(x, q, m, \lceil (1 - x) / q \rceil) & \text{otherwise} \end{cases}$$

$$M(x, q, m, h) = \frac{1}{\pi} \int_0^{\pi} P(x + q \cdot \cos(\Psi), q, m - 1) d\Psi$$

The variable  $x$  represents the percentage of the dominant gene in the total population, regardless of genotype, in generation 0. The variable  $m$  represents the generation number of the descendants of our original population with percentage  $x$ . The variable  $\Psi$  is the angle at which the actual value  $x$  prime deviates from the ideal in any given generation.  $\Psi$  is a local variable. It is also the variable of integration in the formulae. The variable  $q$  represents the radial deviation from the ideal value given by the Hardy-Weinberg equation whose angle of deviation is  $\Psi$  ( $q = 0$  represents the special case of the Hardy-Weinberg equation). The variable  $h$  represents the number of deviations  $q$  away from 1 that our number  $x$  is located within, rounded up to the next integer. The variable  $h$  can be viewed as a number defining a deviation bracket of values with the same genotype formulae for a given generation  $m$ .

The probability  $P(x, q, m)$  is the probability of the recessive gene becoming extinct based solely on comparing  $x$  to 0, 1 or the genotype deviation of  $m$  times  $q$ . The probability  $M(x, q, m, h)$  tries to compute the probability  $P(x, q, m)$  by integration based on the  $x$  percentage's deviation bracket,  $h$  prime. The designation  $x$  prime is the reference to the distribution of values of  $x$  from generation  $(m-1)$  to generation  $m$ . The designation  $h$  prime is the reference to the  $h$  value of a set of  $x$  prime values.

## Proof of Theorem 1

Cases 1 - 4. This problem can be looked at in four distinct cases. Note that these cases need not be mutually exclusive, but any overlap will result in the same value.

1.  $x \leq 0$
2.  $x \geq 1$
3. The recessive gene cannot die out within  $m$  steps
4. The recessive gene can die out within  $m$  steps

If  $x \leq 0$  or  $x \geq 1$  then  $P(x, q, m)$  is 0 and 1 respectively.

The third case, where  $x$  cannot die out,  $P(x, q, m)$  would also be 0. In order to use this case, one has to define exactly where the  $x$  value can and cannot die out. To die out in the first generation, the  $x$  value must be within one deviation of the value  $x = 1$ , allowing some of the rightmost arc of the circle to have a value  $x \geq 1$ . This logic can be extended to  $x$  having to be within  $m$  times  $q$  of 1 in order for the  $x$  value

to be caught in  $m$  generations. If the  $x$  value is not within this distance, it will have no "genocide" probability, or  $P(x,q,m) = 0$ .

This leaves us only with the case when  $x$  is between 0 and 1, and does not die out within the first generation. In my Introduction, I have stated that my bias will be represented as the perimeter of a circle of radius  $q$  around my initial point  $x$ . This means for a given angle  $\Psi$ , the  $x$ - $y$  coordinates of the resulting point would be  $\{x + q \cos(\Psi), 2x(1-x) + q \cos(\Psi)\}$ . Since the  $y$  coordinate is not important in determining the value of the Hardy-Weinberg equation, it can be discarded. This results with our learning that  $x \text{ prime} = x + q \cos(\Psi)$ . If we start with  $x$  being not the percentage of dominant genes in generation 0, but in generation  $m$ ,  $x \text{ prime}$  would be the percentage of dominant genes in generation  $(m-1)$ . The radius  $q$  would be unchanged, and the variable of integration would be  $\Psi$ . Since the variable  $\Psi$  can range from 0 to  $2\pi$ , we would have to divide the equation by  $2\pi$ . But one can see from the following graph that the top half of the circle would give you corresponding  $x$  values for the bottom half of the circle. Therefore, we need only integrate from 0 to  $\pi$ , and divide by  $\pi$  in order to get all possible values for  $P(x,q,m-1)$ .

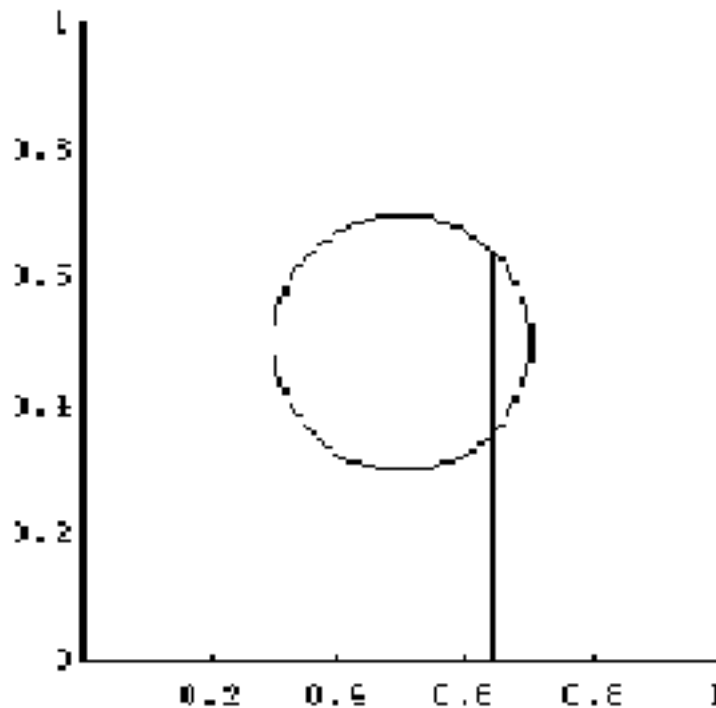


Figure 1: ???

EOP

## Conclusions from Theorem 1

The above formulae are good for a simplified general solution, but application of these formulae in such a reduced form tends to be a little hectic, and useless. In Theorem 2, I have expanded the equation  $M(x,q,m,h)$  with integration over sets of  $\Psi$ , based on the number of deviations  $q$  from 1.

## Theorem 2

Given a reproducing population with dominant and recessive genes, with  $x$  being the percentage of dominant gene in the total population.  $P[x, q, m]$ , the probability that the recessive gene will die out in  $m$  generations given a radial deviation of  $q$  in any one generation, can be computed by the following formulae:

$$P(x, q, m) = \begin{cases} 0 & \text{for } x \leq 0 \text{ or } x \leq 1 - m \cdot q \\ 1 & \text{for } x \geq 1 \\ M(x, q, m, \lceil (1 - x) / q \rceil) & \text{otherwise} \end{cases}$$

$$E_A(x, q, m, \varphi, h) = \begin{cases} 0 & \text{for } m \leq h + 1 \\ U_A(x, q, h) \\ \int_{\varphi=L_A(x, q)} M(x + q \cdot \cos(\varphi), q, m - 1, h + 1) & \text{otherwise} \end{cases}$$

$$E_B(x, q, m, \varphi, h) = \begin{cases} 0 & \text{for } m \leq h \\ U_B(x, q, h) \\ \int_{\varphi=L_B(x, q, h)} M(x + q \cdot \cos(\varphi), q, m - 1, h) & \text{otherwise} \end{cases}$$

$$E_C(x, q, m, \varphi, h) = \begin{cases} 0 & \text{for } x = 1 - h \cdot q \\ \arccos((1 - x) / q) & \text{for } h = 1 \\ U_C(x, q, h) \\ \int_{\varphi=L_C(x, q, h)} M(x + q \cdot \cos(\varphi), q, m - 1, h - 1) & \text{otherwise} \end{cases}$$

$$M(x, q, m, h) = \text{Block} \left[ \left\{ \varphi \right\}, (E_A(x, q, m, \varphi, h) + E_B(x, q, m, \varphi, h) + E_C(x, q, m, \varphi, h)) / \pi \right]$$

Note that limits have not been stated at this point. They are included under "Limits on Case 4, Subclasses 1-3".

## Proof of Theorem 2

*Importance of the  $h$  number:* From the above explanation of the genocide probability, the distance the  $x$  value is from 1 plays a role in what  $P[x]$  value it receives. If one is outside the genocide deviation,  $m$  times  $q$ ,  $P[x]$  is zero. On the other side of the coin, the distance within the genocide deviation, and consequently the number of deviations from 1, should and does affect the solution formulae.

Seeing from above that any value of  $x$  in the first deviation would result in the same formula for  $P[x]$  in the first generation, one can hypothesize correctly that all values of  $x$  within a deviation bracket,  $h$ , will have the same genocide formulae for a given generation  $m$ . If one then defines one's integrals according to the  $h$  values instead of the  $x$  values, one can find integral formulae for ranges of  $x$  instead of single values. This is helpful when one is taking an integral of an integral of  $x$ .

In the evaluation of the fourth case, three Subclasses have to be looked at each comparing the present  $h$  value of  $x$  and the resulting  $h$  prime value of  $x$  prime:

1. The  $x$  prime value is in a deviation bracket with an  $h$  prime value of  $(h+1)$ .
2. The  $x$  prime value is in a deviation bracket with an  $h$  prime value equal to  $h$ .
3. The  $x$  prime value is in a deviation bracket with an  $h$  prime value of  $(h-1)$ .

The generation with a higher  $h$  class would actually have lower  $x$  values. This part of the equation would exist only when the generation number,  $m$ , exceeds the deviation bracket,  $h$ , by at least two. The generation with the same  $h$  class would have approximately the same  $x$  prime values as the original  $x$ . It would exist only when the generation number,  $m$ , exceeds the deviation bracket,  $h$ , by at least one. The generation with a lower  $h$  class would have a higher  $x$  prime value than the original  $x$ . It would always be evaluated, but does have a simplified equation at  $h=1$  and  $x=(1 - h q)$ .

**Limits on Case 4, Subclasses 1-3:** In each case, the evaluation limits must be no less than  $x=0$  and no more than  $x=1$ . The upper limit of Subclass 1 cannot be 1, since there must be a lower bound for Subclass 2 greater than 0 in order for there to be a Subclass 1 in the first place.

Note: As of yet, I have defined each of these cases in terms of  $m$ , (the generation number),  $h$  (the deviation bracket), and  $x$  (the original position value). Unfortunately, the limits of integration have to be based on  $¥$ . Cases will be defined in terms of  $m$ ,  $h$ , and  $x$ , but limits will be in terms of  $¥$ :

$$L_A(x, q) = \begin{cases} \arccos(x/q) & \text{for } 0 > x - q \\ \pi & \text{otherwise} \end{cases}$$

$$U_A(x, q, h) = \pi - \arccos\left(\frac{x - (1 - q \cdot h)}{q}\right)$$

$$L_B(x, q) = \begin{cases} \arccos\left(\frac{1 - x}{q}\right) & \text{for } 0 > 1 - q \cdot h \\ \pi - \arccos\left(\frac{x - (1 - q \cdot h)}{q}\right) & \text{otherwise} \end{cases}$$

$$U_B(x, q) = \begin{cases} \arccos\left(\frac{1 - x}{q}\right) & \text{for } 1 < 1 - q \cdot (h - 1) \\ \arccos\left(\frac{1 - q \cdot (h - 1) - x}{q}\right) & \text{otherwise} \end{cases}$$

$$L_C(x, q, h) = \begin{cases} \arccos(x/q) & \text{for } 0 > 1 - q \cdot (h - 1) \\ \arccos\left(\frac{1 - q \cdot (h - 1) - x}{q}\right) & \text{otherwise} \end{cases}$$

$$U_C(x, q, h) = \begin{cases} \arccos\left(\frac{1 - x}{q}\right) & \text{for } 1 < x + q \\ 0 & \text{otherwise} \end{cases}$$

Equations for class 4 Subclass 1-3 defined in are defined in Theorem. "Equation Subclass" has been abbreviated " $E_I$ ".

**EOP**

## Conclusions from Theorem 2

These formulae break down the integral of Theorem 1 into integrals with the exact same resulting escape formulae. Due to the fact that my final formulae are based on a recursive relationship, even though theoretically possible, you may not wish to go to such a length.

## Directions for Further Research

The only assumption that I made was that the deviation was on the circumference of a circle of radius  $q$ .

What other forms could this deviation take? How would this effect the genocide probability?