School of Engineering and Applied Science (SEAS), Ahmedabad University

B.Tech(ICT) Semester IV: Probability and Random Processes (MAT 202)

- Group No : **S_B1**
- Name (Roll No) : **Nancy Radadia(AU1841070), Suhanee Patel(AU1841113), Yash Patel(AU1841125)**
- Project Title: **Prediction of the Probabilities of the Transmission of Genetic Traits within Bayesian Logical Inference**

1 Introduction

1.1 Background

- Have you ever noticed that some of us might resemble our parents in some way? Like we might have matching eye colour, or similar hair textures etc. We got these characteristics from our parents, and they got them from their parents. So eventually we can say that these biological informations are passed from one generation to another. These units of inheritance are known as traits. Here we aim to predict the traits in future generations provided the data of past generations. By using some probabilistic modelling we can predict the probability of transmission of genetic traits.
- There are many models designed to study the phenotypic prediction of traits at individual levels such as eye color [1], height [2] ,hair color, skin color [3] etc. Prediction of traits is used for many purposes like quantifying the risks of diseases for individuals [4] [5] [6] [7] . Diseases can be of different types. Among which one in autosomal recessive disease. Such diseases are transmitted when an infected recessive gene is crossed with another infected recessive gene . When a mutated gene from both parents is passed to their child then that child is said to be affected. We can model different and complex scenarios using appropriate probabilistic models to find the required likelihoods and estimate the risks accordingly [8].
- Prediction of an autosomal recessive disease named Cystic Fibrosis(CF) [9] is demonstrated in our base article. Cystic Fibrosis(CF) is a single-gene disorder caused by mutations in CF transmembrane conductance regulator (CFTR) genes found in cells that line the lungs, digestive tract, sweat glands, and genitourinary system. CF patients are expected to die within the first years of life. Their life expectancy has lengthened with advances in diagnosis and treatment and is currently about 38 years [10].

1.2 Motivation

- The main vision behind this study is to be able to predict the genetic traits in future generations from the data of the prior generation. This prediction can serve extremely helpful for understanding the risks of any diseases which could be passed by inheritance. Once predicted, suitable precautionary measures can be taken. This can increase the recovery chances of individuals and improve the overall life expectancy. So if we are able to know any information related to genetic traits of our future generations, we have a chance to take prior actions.

1.3 Problem Statement/ Case Study

Our base article aimed at calculating the probability of a heterozygous offspring based on the data and information provided by the previous generations and by taking some prior assumptions. We have derived a general formula for finding the probability for a mono hybrid genetic cross and have generated a matlab code to verify our results with those given in the article. We have used Bayesian logical inferences along with Markov chain to model the uncertainty. The final aim of the article was to predict autosomal recessive diseases for evaluating its performance on real-world data under Bayesian framework. Also in order to demonstrate and evaluate the flexibility of the method, we have tried to evaluate different examples of pedigree diagrams and punnett squares which contain different genetic crosses, such as monohybrid, dihybrid, trihybrid and multi hybrid genetic crosses and have came to a general formula using which we can find the probability upto 'n' number of generations and analyzing 'm' number of traits.

2 Data Acquisition

- The example which we worked on dont require any data set, as we are taking some assumptions for the past generations and trying to model it to derive a general formula for different scenarios for the nth generation.

3 Probabilistic Model Used/ PRP Concept Used

3.1 Methodolgy/Appraoch

- The very first goal of the probabilistic model here is to analyse inheritance of traits in human and animal populations. It then determines the mode of inheritance such as **dominant** or **recessive**. This can be achieved by the use of **pedigree analysis** which describes the characteristics of all the generations in a family.

  The next goal is to calculate the probability of an affected offspring for a given genetic cross. To achieve this there are various methods but here in this model we use **Bayesian logical inferences**.

  Also, the outcome of genetic crossing of traits depends only on that of the previous one which can

be known from pedigree analysis and is used to create transmission probability vector of traits for a given generation. A special kind of stochastic process called **Markov chain** uses this vector to predict posterior probability of transmission of genetic traits under some assumptions. Hence we calculate the probability of an affected offspring among generations within **Bayesian framework with Markov chain**. Also, using this probabilistic model we can even predict transmission of hereditary diseases among generations such as Cystic fibrosis (CF).

3.2 Bayesian logical inferences

- The task of Bayesian logical inferences here is to infer the probability for the hypothesis **H**, given some data **D** from experiment and capturing all relevant information **I**. This can be done within the setting of Bayes' theorem which states

$$Pr(H|D,I) = \frac{Pr(D|H,I) \cdot Pr(H|I)}{Pr(D|I)}$$

Here, $Pr(D|I)$ is is a global likelihood for entire class of **H** or evidence given some information **I**. The quantity $Pr(D|H,I)$ is called likelihood of H, which measures or determines a probability of observations D, or the statistic under the hypothesis being tested. The quantity $Pr(H|I)$ is known as a prior probability distribution function (PDF) of H in the absence of D and the quantity $Pr(H|D,I)$ is a posterior PDF of H.

3.3 Pedigree Analysis and Modelling for monohybrid crosses

- To study Pedigree Analysis for Monohybrid crossess, consider an example of hamster and we suppose that a single gene controls the hair length of a hamster. Then, a short hair is governed by a dominant gene and represented by $\boldsymbol{L}$ while a long hair is governed by a recessive gene and represented by $\boldsymbol{l}$. Therefore, a hamster will be phenotypically short haired unless its genotype is (ll). Consider the given figure which represents the pedigree analysis of monohybrid genetic crosses. Also, in G_1 both parents are known to be phenotypically short haired. Crossing of *(Ll)* and *(LL)* is given in figure 2-

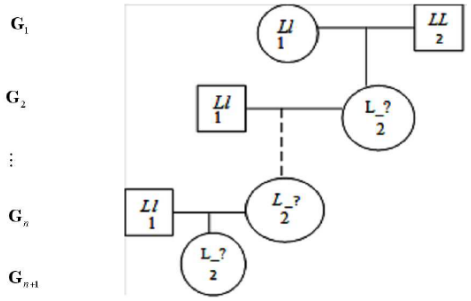


Figure 1: Pedigree diagram for monohybrid crosses of $G(n+1)_2$

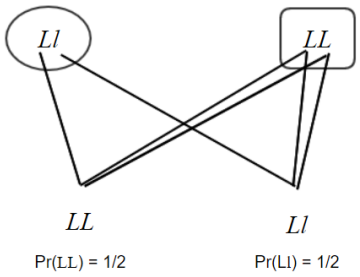


Figure 2: Genetic crossing of G_1 generation

- Hence, the probability of *(LL)* = *(Ll)* = 1/2 and both the outcomes ultimately turn out to be phenotypically short hair, where one is heterozygous short hair *(Ll)* whereas another is homozygous short haired *(LL)*. Given that G_{i1}'s are carriers and $G_{(n+1)2}$ has a short hair, we would like to find the probability of $G_{(n+1)2}$ being genotypically heterozygous *Ll*.

  To find the probability of the affected offspring we use bayesian method as discussed before. From the pedigree diagram we can say that our information ***I*** represents that the genotypes of parent G_{11} and G_{12} are initially known to be heterozygous short haired *(Ll)* and homozygous short haired *(LL)* respectively and our evidence ***E*** is that G_{i2} is short hair. Let us define H = $(h_i)_{i=1}^{k}$ = {*Ll*,*Ll*,*ll*} where k is the number of hypothesis. Now since we are finding the probability of h2 i.e. *Ll*, the bayseian equation takes the following form:

$$Pr(h_2|E,I) = \frac{Pr(E|h_2,I)\cdot Pr(h_2|I)}{Pr(E|I)}$$

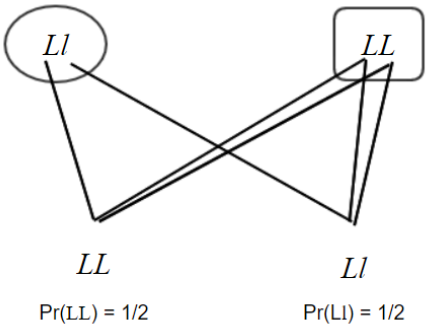
$$= \frac{Pr(E|h_2,I)\cdot Pr(h_2|I)}{\sum_{i=1}^{m} Pr(E|h_i,I)\cdot Pr(h_i|I)} \qquad (1).$$

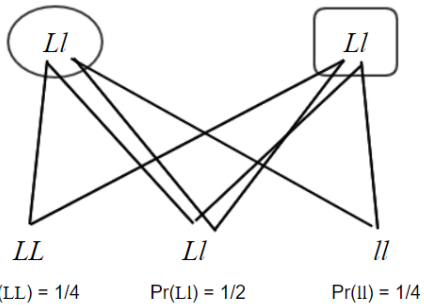
  The total probability in the denominator is calculated using Markov chain.

4 Pseudo Code/ Algorithm

- To calculate the total probability in Eq.(1) we use matrices whose components indicate probabilities of genetic crosses of traits. In this context, a relationship between the probabilities of the offspring's genotypes and that of its parents can be modeled using transition matrices used in the Markov chains. Therefore, the probability values of outcomes of genotypes by crossing G_{11} with *LL*, *Ll*, and *ll* is shown in the figure below and then using this we form transition matrix A which is shown in table I.

  Hence, from the figures obtained from genetic crossing of traits we get transition matrix A, shown in table I.





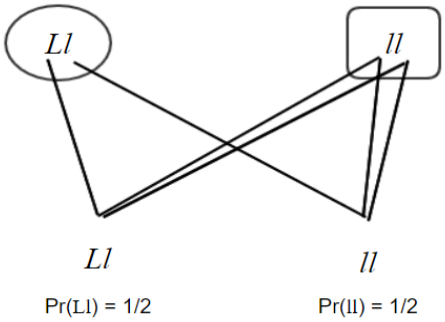


TABLE I

Probabilities of genotypes obtained by crossing G_{11} with LL, Ll and ll respectively.

	$Ll \times LL$	$Ll \times Ll$	$Ll \times ll$
$p(LL)$	1/2	1/4	0
$p(Ll)$	1/2	1/2	1/2
$p(ll)$	0	1/4	1/2

- On the other hand, let x represent the prior probabilities of hypothesis $(h_i)_{i=1}^{k}$ = of genotypes {*Ll*,*Ll*,*ll*} so that initially, $x_0 = [1, 0, 0]^T$ because G_{12} is given as LL genotype in Figure 1. After one generation,

$$x_1 = Ax_0 = [0.5, 0.5, 0]^T$$

Similarly, after two generations,

$$x_2 = Ax_1 = A^2x_0$$

Hence after n generation where $(n \in Z^+)$ we have

$$x_n = A^n x_0$$

Here it is required to calculate higher power of matrix A, hence we can use one of the linear algerbra

property i.e. **Diagonalization**

$$x_n = A^n x_0 = P \triangle^n P^{-1} x_0$$

where $\triangle$ is a diagonal matrix whose diagonal entries are eigenvalues of A and P is a matrix whose columns are linearly independent eigenvectors of A corresponding to its eigenvalues. To obtain the value of $Pr(E|I)$ we multiply x_n ny normalisation constant d which is $[1,1,0]^T$ as $p(E|h1,I) = p(E|h2,I) = 1$ and $p(E|h3,I) = 0$, respectively. Thus, going back to Eq.(1) we calculated $Pr(E|I)$, also the value of $Pr(h_2|I) = 1/2$ and $Pr(E|h_2,I) = 1$. Hence, substituting all these values in Eq(1) we can calculate probablity of hetrozygous short hair of any given n^{th} generation.

- To derive general expression, we first calculate the probability of first few generations, After **first generation**, we calculate probability as follows

$$x_1 = Ax_0 = [0.5, 0.5, 0]^T$$
$$x_1 \cdot d = [1,1,0] \cdot [0.5, 0.5, 0]^T = 1 = p(E|I)$$

Substituting, the above value in equation (1), the probability for hetrozygous short hair after generation 1 will be

$$Pr(h_2|E,I) = \frac{Pr(E|h_2,I) \cdot Pr(h_2|I)}{Pr(E|I)}$$

$$= \frac{1/2 \cdot 1}{1}$$
$$Pr(h_2|E,I) = \frac{1}{2}$$

Similarly, After **second generation**, we calculate as follows

$$x_2 = A^2 x_0 = [3/8, 1/2, 1/8]^T$$
$$x_2 \cdot d = [1,1,0] \cdot [3/8, 1/2, 1/8]^T = 7/8 = p(E|I)$$

Substituting, the above value in equation (1), the probability for hetrozygous short hair after generation 2 will be

$$Pr(h_2|E,I) = \frac{Pr(E|h_2,I) \cdot Pr(h_2|I)}{Pr(E|I)}$$

$$= \frac{1/2 \cdot 1}{7/8}$$
$$Pr(h_2|E,I) = \frac{4}{7}$$

After, **Third generation**,

$$x_3 \cdot d = [1, 1, 0] \cdot [5/16, 1/2, 3/16]^T = 13/16 = p(E|I)$$

$$Pr(h_2|E, I) = \frac{Pr(E|h_2, I) \cdot Pr(h_2|I)}{Pr(E|I)}$$

$$Pr(h_2|E, I) = \frac{8}{13}$$

Observing the above trend in values for generation, we arrive at general form of equation -

$$Pr(h_2|E, I) = \frac{1/2 \cdot 1}{\left[1 + \left(-\frac{1}{8}\right) + \left(-\frac{1}{16}\right) + ... + \left(-\frac{1}{2^n}\right)\right]}$$

for after n ($n \in Z^+$) generation.

Solving the above equation further,

$$Pr(h_2|E, I) = \frac{1/2 \cdot 1}{1 - \left(\frac{1}{8} + \frac{1}{16} + \frac{1}{32} + ... + \frac{1}{2^n}\right)}$$

Solvoing the above denominator by using Geometric Progression, where a (first term) = 1/8 and r (difference) is 1/2. Formula for solving G.P. goes like this-

$$Sum = \frac{a \cdot \left(r^n - 1\right)}{r - 1}$$

$$= \frac{1/8 \cdot \left(\frac{1}{2^n} - 1\right)}{\frac{1}{2} - 1}$$

Substituting in sum in the equation we get,

$$Pr(h_2|E, I) = \frac{\frac{1}{2} \cdot 1}{1 - \left(\frac{\frac{1}{8} \cdot \left(\frac{1}{2^n} - 1\right)}{-\frac{1}{2}}\right)}$$

$$= \frac{1/2}{\frac{3}{4} + \frac{1}{2^{n+1}}}$$

Multipling by 2^{n+1} on both numerator and denominator we get the final equation as-

$$\boxed{Pr(h_2|E,I) = \frac{2^n}{3 \cdot 2^{n-1} + 1}}$$

The above equation is the posterior PDF of h_2 i.e hetrozygous short hair *(Ll)* after n generations for **Monohybrid Crosses**

5 Coding and Simulation

5.1 Simulation Framework

We try to simulate the above derived equation for monohybrid crosses in matlab. The x-axis represent the generations and the the Y axis represents the probability of hetrozygous short hair in the correspoding generation of X-axis. The transition matrix A, prior probability distribution i.e. $Pr(h_2|I)$ and $Pr(E|h_2, I)$ are predefined and we calculate the total probability in the denominator and find the PDF for each generation. The no of generations will also be predefined and the no of times we calculate the total probability or the number of tumes we run Markov chain simulations depends on n (generation). We use for loop running from 0 to n which calculates simualtion of Total probability and calculates the PDF.

5.2 Reproduced Figures

- Here, we plot generation vs Posterior PDF of hetrozygous short hait *(Ll)*, where we take n (no. of generation) = 20, hence we run for loop for 20 times and calculate PDF as shown below.

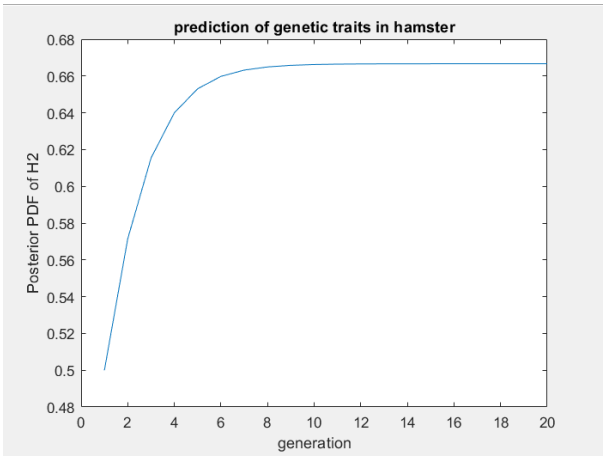


Figure 3: Posterior pdf for h2 *(Ll)*

$$p(h_2|E,I) = \frac{2^n}{3 \times 2^{n-1} + 1}. \quad (6)$$

It can be seen that $p(h_2|E,I)$ approaches to 2/3 as $n \to \infty$ and imply that under assumptions about parents and all crossings with heterozygous Ll an individual born after n-th generation is likely to be heterozygous shortly haired with a probability of two-thirds. In Eq. (5), it is

Figure 4: Results of PDF for h2 in base article

Here, figure 4 is as per the article which states that as $(n \to \infty), Pr(h2|E,I)$ approaches 2/3 i.e. 0.66 which matches with our plot given in Figure 3.

5.3 New Work Done

5.3.1 New Analysis

We derived a general expression and implemented mathametical code that calculates posterior probability of hypothesis of genetic crosses for monohybrid. Now we try to extened the study by finding probabililty of hetrozygous traits in Dihybrid and Tridybrid using different method called **Punnett Square** and we arrive deriving the general expression of finding probability in multihybrid crosses.

- **Dihybrid Crosses**

  In genetics, a dihybrid cross is known as cross between two hybrid genes of parents that differ in two traits of particular interest. The genes are located on separate chromosomes, so the traits themselves are **unrelated**. Let there be two traits in hamster, one which controls fur color and second which controls hair style (curly or straight). As, we discussed previously, for each trait, there are dominant and recessive genes. The dominant fur color is black and represented by $\boldsymbol{B}$, while the recessive fur color is white and represented by $\boldsymbol{b}$, similarly in second trait the dominant hair style is curly represented by $\boldsymbol{C}$ and recessive hairstyle is represented by $\boldsymbol{c}$. Consider the pedigree diagram of diybrid:

  According, to figure 5, under the assumption that both Generation II-2 and III-2 are phenotypically

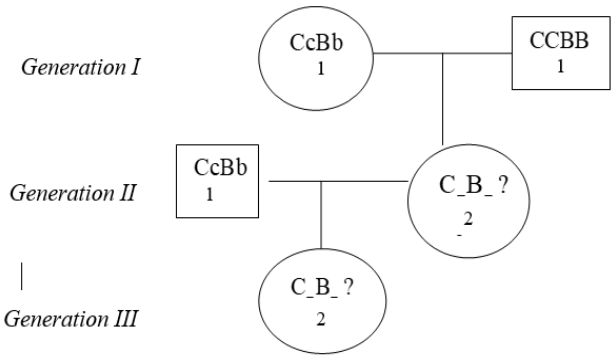


Figure 5: Genetic Pedigree for Dihybrid Cross

curly haired and black fur, we try to find the probability of the Generation III-2 being genotypically heterozygous curly haired and black fur (CcBb). Genetic crossing of dhybrid is complexed as compared to monohybrid and hence we use a method called **Punnett square** for crossing.

Since parents are heterozygous and homogeneous in a dihybrid cross, the possible gene combinations from each parent listed on the Punnett square are *Bb*, *BB*, *Cc* and *CC*. Given below are the punnet sqaure for both genertion II and generation III in table II and III respectively.

Table II. Punnett Square for *Generation II-2*

Property II / Property I	BB(1/2)	Bb(1/2)
CC(1/2)	CCBB(1/4)	CCBb(1/4)
Cc(1/2)	CcBB(1/4)	CcBb(1/4)

From the above punnet square of Generation II, we can say that -

$$Pr(h_2, I) = Pr(CcBb, I) = 1/4$$

Table III. Punnett Square for Dihybrid Cross of *Generation III*-1

CcBb x CCBB		
Property II / Property I	BB(1/2)	Bb(1/2)
CC(1/2)	CCBB(1/4)	CCBb(1/4)
Cc(1/2)	CcBB(1/4)	CcBb(1/4)

CcBb x CcBb			
Property II / Property I	BB(1/4)	Bb(1/2)	bb(1/4)
CC(1/2)	CCBB(1/8)	CCBb(1/4)	CCbb(1/8)
Cc(1/2)	CcBB(1/8)	CcBb(1/4)	Ccbb(1/8)

CcBb x CcBB		
Property II / Property I	BB(1/2)	Bb(1/2)
CC(1/4)	CCBB(1/8)	CCBb(1/8)
Cc(1/2)	CcBB(1/4)	CcBb(1/4)
cc(1/4)	ccBB(1/8)	ccBb(1/8)

CcBb x CcBb			
Property II / Property I	BB(1/4)	Bb(1/2)	bb(1/4)
CC(1/4)	CCBB(1/16)	CCBb(1/8)	CCbb(1/16)
Cc(1/2)	CcBB(1/8)	CcBb(1/4)	Ccbb(1/8)
cc(1/4)	ccBB(1/16)	ccBb(1/8)	ccbb(1/16)

To find the total probability in denominator of Equation (1):

$Pr(E|I) = Pr(\text{curly hair,black fur}|G)$

$Pr(E|I) = 1$ - $Pr(\text{straight hair,black fur } \textbf{or} \text{ straight hair,white fur } \textbf{or} \text{ curly hair,while fur}|G)$

$$Pr(E|I) = 1 - Pr(\{ccBb\}, \{ccBB\}, \{Ccbb\}, \{CCbb\}, \{ccbb\}|G)$$

From the punnett square of generation III, we can find the above probability:

$$Pr(E|I) = 1 - \left(\left(\frac{1}{8} + \frac{1}{8}\right) \cdot \frac{1}{4} + \left(\frac{1}{8} + \frac{1}{8}\right) \cdot \frac{1}{4} + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16}\right) \cdot \frac{1}{4}\right)$$

$$Pr(E|I) = 1 - \left(\frac{15}{64}\right)$$

$$Pr(E|I) = \frac{49}{64}$$

Now substituting the values in bayesian Equation (1) we get the probablity of hetrozygous curly haired and black fur after Generation II (n=2) i.e. for Generation III

$$Pr(h2|E, I) = \frac{Pr(h_2|I) \cdot 1}{Pr(E|I)}$$

$$Pr(h2|E, I) = \frac{16}{49}$$

- **Trihybrid Crosses**

  Trihybrid crosses involves crosses of three organisms, in which the genes of three traits are examined. Let us add a new trait, called fur texture, to the previous example of Dihybrid. The same rules as before applied for hairstyle and fur color are also used for fur texture, where rough fur denoted by $\boldsymbol{R}$ is taken to be dominant over smooth fur denoted by $\boldsymbol{r}$. A pedigree related with these traits is shown below in figure 6. Again, our objective is to determine the probability that Generation III-2 is geno-

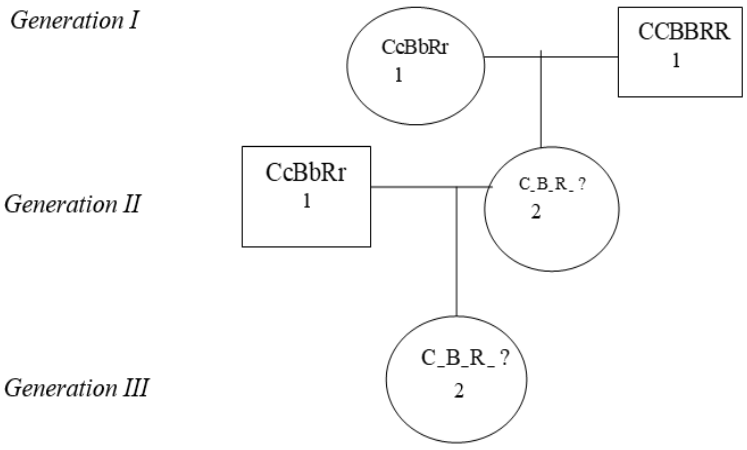


Figure 6: genetic pedigree for trihybrid Cross

typically heterozygous curly haired, black and rough fur *(CcBbRr)*. Punnett square of Generation II-2 is shown in Table 4 and for Generation III-2 is shown in table V. From table 4 we can know the value

Table IV Punnett Square for trihybrid crosses of *Generation II*-2

Gametes and probabilities	CBR	CBr	CbR	Cbr	cBR	cBr	cbR	cbr
CBR	CCBBRR (1/8)	CCBBrR (1/8)	CCbBRR (1/8)	CCbBrR (1/8)	cCBBRR (1/8)	cCBBrR (1/8)	cCbBRR (1/8)	cCbBrR (1/8)

Table V Punnett Square for Trihybrid Cross of CcBbRr × CcBbRr

CcBbRr × CcBbRr								
Gametes and probabilitie	CBR (1/8)	CBr (1/8)	CbR (1/8)	Cbr (1/8)	cBR (1/8)	cBr (1/8)	cbR (1/8)	cbr (1/8)
CBR (1/8)	CCBBRR (1/64)	CCBBRr (1/64)	CCBbRR (1/64)	CCBbRr (1/64)	CcBBRR (1/64)	CcBBRr (1/64)	CcBbRR (1/64)	CcBbRr (1/64)
CBr (1/8)	CCBBrR (1/64)	CCBBrr (1/64)	CCBbrR (1/64)	CCBbrr (1/64)	CcBBrR (1/64)	CcBBrr (1/64)	CcBbrR (1/64)	CcBbrr (1/64)
CbR (1/8)	CCbBRR (1/64)	CCbBRr (1/64)	CCbbRR (1/64)	CCbbRr (1/64)	CcbBRR (1/64)	CcbBRr (1/64)	CcbbRR (1/64)	CcbbRr (1/64)
Cbr (1/8)	CCbBrR (1/64)	CCbBrr (1/64)	CCbbrR (1/64)	CCbbrr (1/64)	CcbBrR (1/64)	CcbBrr (1/64)	CcbbrR (1/64)	Ccbbrr (1/64)
cBR (1/8)	cCBBRR (1/64)	cCBBRr (1/64)	cCBbRR (1/64)	cCBbRr (1/64)	ccBBRR (1/64)	ccBBRr (1/64)	ccBbRR (1/64)	ccBbRr (1/64)
cBr (1/8)	cCBBrR (1/64)	cCBBrr (1/64)	cCBbrR (1/64)	cCBbrr (1/64)	ccBBrR (1/64)	ccBBrr (1/64)	ccBbrR (1/64)	ccBbrr (1/64)
cbR (1/8)	cCbBRR (1/64)	cCbBRr (1/64)	cCbbRR (1/64)	cCbbRr (1/64)	ccbBRR (1/64)	ccbBRr (1/64)	ccbbRR (1/64)	ccbbRr (1/64)
cbr (1/8)	cCbBrR (1/64)	cCbBrr (1/64)	cCbbrR (1/64)	cCbbrr (1/64)	ccbBrR (1/64)	ccbBrr (1/64)	ccbbrR (1/64)	ccbbrr (1/64)

of $Pr(h_2|I)$ and the total prior probabilities among all possible off-spring in Generation III, producing an individual of genotype, can be kown from table V which is found by summation of all red colored rectangles labeled probabilities. All of the yellow colored-rectangles in Table V within eight Punnet Square tables are used in order to obtain normalized constant $Pr(E|I)$.

$$Pr(h_2|I) = Pr(CcBbRr|I) = \frac{1}{8},$$

Now, we find the total probability,

$$\begin{aligned} Pr(E|I) &= Pr(A \ = \ curly \ \ hair, \ \ black \ \ and \ \ rough \ \ fur|I) \\ &= 1 - Pr(A^T|I) \\ &= \frac{343}{512} \end{aligned}$$

Now substituting, the above values in bayseian Equation (1), we get our required probability-

$$Pr(h2|E, I) = \frac{Pr(h_2|I) \cdot 1}{Pr(E|I)}$$

$$Pr(h2|E, I) = \frac{64}{343}$$

The above was the probability of hetrozygous curly haired and black and rough fur after Generation II (n=2) i.e for Generation III for trihybrid crosses.

- **Multihybrid**

  We found the probability of hetrozygous traits in hamster after Generation II (n=2) in dihybrid and trihybrid. If we observe the trend of praobability for n=2, in monohyrid, dihybrid and tribrid, we conclude the following-

  In monohybrid, for n=2, $Pr(h_2|E.I) = \frac{4}{7}$

  In dihybrid, for n=2, $Pr(h_2|E.I) = \frac{16}{49} = \left(\frac{4}{7}\right)^2$

  In trihybrid, for n=2, $Pr(h_2|E.I) = \frac{64}{343} = \left(\frac{4}{7}\right)^3$

  As we derived previously the genral exression for monohybrid cross to find the posterior PDF of h_2 for any generation n is as follows-

$$Pr(h_2|E, I) = \frac{2^n}{3 \cdot 2^{n-1} + 1}$$

Hence for, m different traits, we can say that-

$$Pr(h_2|E,I) = \left(\frac{2^n}{3\cdot 2^{n-1}+1}\right)^m$$

OR

$$Pr(h_2|E,I) = \left(\frac{2^n}{3}\right)^m \cdot \left(\frac{1}{2^{n-1}+3^{-1}}\right)^m$$

Here, we assumed the equation for posterior probability of h_2 for m different traits. We try prove the correctness of our assumption using Induction method. From equation (2), we can say that

$$\text{To Prove } \left(\frac{1/2\cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+.....+\left(-\frac{1}{2^n}\right)}\right)^m = \left(\frac{2^n}{3\cdot 2^{n-1}+1}\right)^m$$

$$\text{Let P(m): } \left(\frac{1/2\cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+.....+\left(-\frac{1}{2^n}\right)}\right)^m = \left(\frac{2^n}{3\cdot 2^{n-1}+1}\right)^m$$

Base case: For m = 1 (monohybrid),

$$\textbf{L.H.S.} = \left(\frac{1/2\cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+.....+\left(-\frac{1}{2^n}\right)}\right)^1$$

$$= \frac{1/2\cdot 1}{1-\left(\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+.....+\frac{1}{2^{n+1}}\right)}$$

$$= \frac{1/2}{\frac{3}{4}+\frac{1}{2^{n+1}}}$$

$$= \frac{2^n}{3\cdot 2^{n-1}+1}$$

$$= \textbf{R.H.S.}$$

Hence, the equation holds true for base case i.e. monohybrid crosses (as derived earlier).
Now assume that this equation is true for m = k. Hence, for m = k,

$$\left(\frac{1/2 \cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+.....+\left(-\frac{1}{2^n}\right)}\right)^k = \left(\frac{2^n}{3 \cdot 2^{n-1}+1}\right)^k$$

Now for m = k+1

L.H.S.

$$= \left(\frac{1/2 \cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+.....+\left(-\frac{1}{2^n}\right)}\right)^{k+1}$$

$$= \left(\frac{1/2 \cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+..+\left(-\frac{1}{2^n}\right)}\right)^k \left(\frac{1/2 \cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+..+\left(-\frac{1}{2^n}\right)}\right)$$

From above assumption for n=k, we substitute the value-

$$= \left(\frac{2^n}{3 \cdot 2^{n-1}+1}\right)^k \cdot \left(\frac{1/2 \cdot 1}{1+\left(-\frac{1}{8}\right)+\left(-\frac{1}{16}\right)+..+\left(-\frac{1}{2^n}\right)}\right)$$

$$= \left(\frac{2^n}{3 \cdot 2^{n-1}+1}\right)^k \cdot \left(\frac{1/2}{\frac{3}{4}+\frac{1}{2^{n+1}}}\right)$$

$$= \left(\frac{2^n}{3 \cdot 2^{n-1}+1}\right)^{k+1}$$

$$= \textbf{R.H.S.}$$

Hence, P(K+1) is true

$\therefore$P(1) is true and P(k) is true $\implies$ P(K+1)is true

$\therefore$ P(m) is true, $\forall m \in N$ by principal of Mathematical Induction. Hence, our above observation for calculating the probability for multihybrid crosses holds correct.

5.3.2 New Coding / Algorithm

We try to plot Generation vs. Posterior PDF of hetrozgous characteristics for n generation of Monhybrid, Dihybrid and Trihybrid crosses. To plot these graphs, we use the above derived equation for multihybrid by substituting the values as m = 1 (for monohybrid), m = 2 (for dihybrid) and m = 3 (for trihybrid).
From the plot given in figure 7 we observe that as $n \to \infty$, $Pr(h_2|E, I)$ approaches 0.666 in monohybrid, in dihybrid it approaches to 0.444 and in trihybrid it approahes to 0.2963

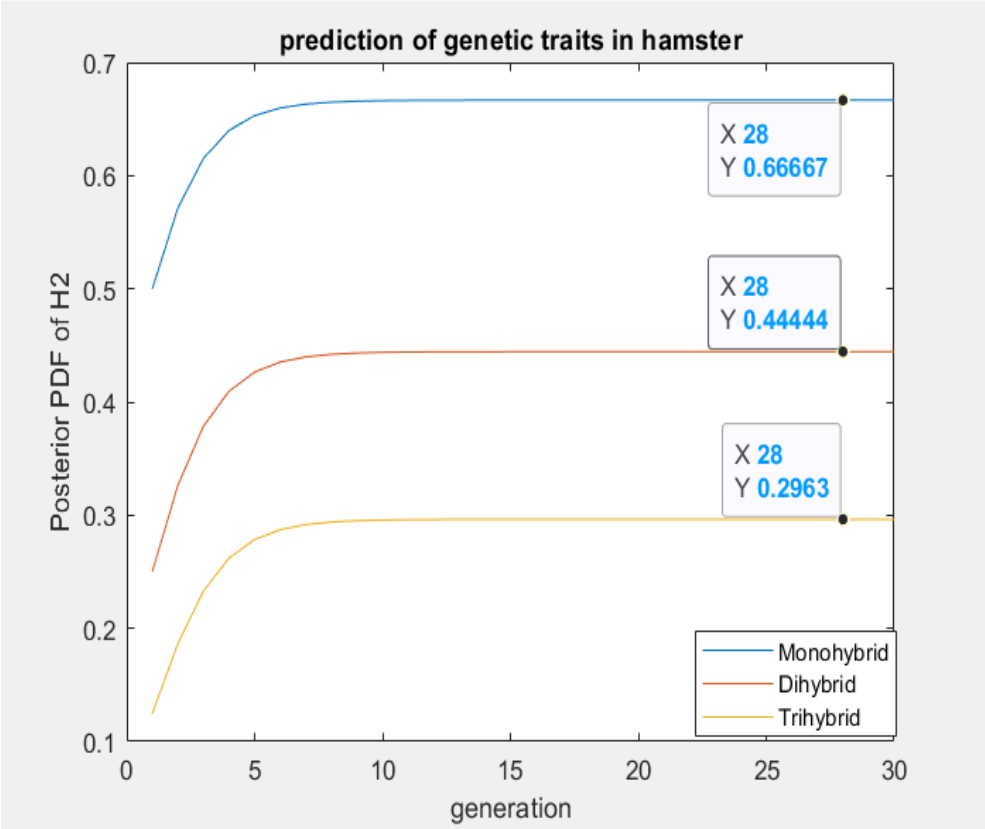


Figure 7: Probability of hetrozygous characteristics in monohybrid, dihybrid and trihybrid

6 Inference Analysis

- In this study, we predicted the probability of transmission of genetic traits from generation to generation within a Bayesian framework and developed a mathematical code that calculates posterior probability of hypothesis for genetic crossing of traits. We extended the study to calculate the probability for multihybrid species. We implemened this trivial experiment to find different characteristic of hamster.
- Further, this can model the trend of transmission of autosomal recessive disesases like Cystic Fibrosis (CF). Under assumption that any person being CF patient and all crossings with him/her and next generations after him/her are not affected by phenotypically CF disease. We can find the probability that his child, grandson and so on are being genetically either heterozygous or homozygous.

7 Contribution of team members

7.1 Technical contribution of all team members

Tasks	Nancy Radadia	Suhanee Patel	Yash Patel
Concept Maps	✓	✓	✓
Data Scraping	✓	✓	✓
Latex Coding	✓	✓	✓
Coding in Matlab	✓	✓	✓
References and Conclusions	✓	✓	✓

7.2 Non-Technical contribution of all team members

Tasks	Nancy Radadia	Suhanee Patel	Yash Patel
Report Writing	✓	✓	✓
Web Surfing	✓	✓	✓
Latex Coding	✓		
Reviewing Article		✓	
References and editing Article			✓

References

[1] K. L. Hart, S. L. Kimura, V. Mushailov, Z. M. Budimlija, M. Prinz, and E. Wurmbach, "Improved eye-and skin-color prediction based on 8 snps," *Croatian medical journal*, vol. 54, no. 3, pp. 248–256, 2013.

[2] L. Dubois, K. Ohm Kyvik, M. Girard, F. Tatone-Tokuda, D. Pérusse, J. Hjelmborg, A. Skytthe, F. Rasmussen, M. J. Wright, P. Lichtenstein *et al.*, "Genetic and environmental contributions to weight, height, and bmi from birth to 19 years of age: an international study of over 12,000 twin pairs," *PLOS one*, vol. 7, no. 2, p. e30153, 2012.

[3] F. Liu, M. Visser, D. L. Duffy, P. G. Hysi, L. C. Jacobs, O. Lao, K. Zhong, S. Walsh, L. Chaitanya, A. Wollstein *et al.*, "Genetics of skin color variation in europeans: genome-wide association studies with functional follow-up," *Human genetics*, vol. 134, no. 8, pp. 823–835, 2015.

[4] P. J. McLaren, J. L. Raisaro, M. Aouri, M. Rotger, E. Ayday, I. Bartha, M. B. Delgado, Y. Vallet, H. F. Günthard, M. Cavassini *et al.*, "Privacy-preserving genomic testing in the clinic: a model using hiv treatment," *Genetics in medicine*, vol. 18, no. 8, pp. 814–822, 2016.

[5] N. Yi and D. Zhi, "Bayesian analysis of rare variants in genetic association studies," *Genetic epidemiology*, vol. 35, no. 1, pp. 57–69, 2011.

[6] H. C. Wijeysundera, G. Tomlinson, C. M. Norris, W. A. Ghali, D. T. Ko, and M. D. Krahn, "Predicting eq-5d utility scores from the seattle angina questionnaire in coronary artery disease: a mapping algorithm using a bayesian framework," *Medical Decision Making*, vol. 31, no. 3, pp. 481–493, 2011.

[7] L. A. Farrer, L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, R. H. Myers, M. A. Pericak-Vance, N. Risch, and C. M. Van Duijn, "Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: a meta-analysis," *Jama*, vol. 278, no. 16, pp. 1349–1356, 1997.

[8] S. Ogino and R. B. Wilson, "Bayesian analysis and risk assessment in genetic counseling and testing," *The Journal of Molecular Diagnostics*, vol. 6, no. 1, pp. 1–9, 2004.

[9] M. L. Drumm, A. G. Ziady, and P. B. Davis, "Genetic variation and clinical heterogeneity in cystic fibrosis," *Annual Review of Pathology: Mechanisms of Disease*, vol. 7, pp. 267–282, 2012.

[10] A. D. Jackson, L. Daly, A. L. Jackson, C. Kelleher, B. C. Marshall, H. B. Quinton, G. Fletcher, M. Harrington, S. Zhou, E. F. McKone *et al.*, "Validation and use of a parametric model for projecting cystic fibrosis survivorship beyond observed data: a birth cohort analysis," *Thorax*, vol. 66, no. 8, pp. 674–679, 2011.