STAT 13, section 1, Winter 2012, UCLA Statistics
HW 3; Problem Solution

- (**HW3.1**) Suppose that a disease is inherited via a sex-linked mode of inheritance, so that a male offspring has a 50% chance of inheriting the disease whereas the female offspring have no chance of getting the disease. Assume that 51% of births are male. In a family with 2 children:

P(D=1 | Child = Male) = 0.5 ➔ P(D=0 | Child = Male) = 0.5

P(D=1 | Child = Female) = 0 ➔ P(D=0 | Child = Female) = 1

P(Child = Male) = .51 ➔ P(Child = Female) = .49

- What is the probability that both sibling are affected, if there are one male and one female?

ANS)

P{Both Siblings Aff | One Male and One Female} = P(D_1 = 1 | C_1 = M) * P(D_2 = 1 | C_2 = F) =

= P(D_1 = 1 | C_1 = M) * 0 = 0

- What is the probability that exactly one sibling is affected?

ANS)

Sample Space with 2 "outcomes" of pairs { (D_1 = 1, D_2 = 0) , (D_1 = 0 , D_2 = 1) } representing the event that exactly one sibling is affected. However the two events are symmetric once we take the probabilities, ie

P((D_1 = 1, D_2 = 0)) = P((D_1 = 0 , D_2 = 1))

P (Exactly one sibling is aff) = P((D_1 = 1 , D_2 = 0) U (D_1 = 0 , D_2 = 1)) = note: inc/excl rule

= P((D_1 = 1 , D_2 = 0)) + P((D_1 = 0 , D_2 = 1)) = 2 * P((D_1 = 1 , D_2 = 0)) = note: symmetric

= 2 * [P(D_1 = 1) * P(D_2 = 0)]= note: Independence of Child 1 and Child 2

= 2 * [P(D_1 = 1 | C_1 = M) P(C_1 = M) + P(D_1 = 1 | C_1 = F)P(C_1 = F)] *

* [P(D_2 = 0 | C_2 = M) P(C_2 = M) + P(D_2 = 0 | C_2 = F)P(C_2 = F)] note: condition on M/F

= 2 * [(0.5)(0.51) + (0)(0.49)] * [(0.5)(0.51) + (1)(0.49)] = 0.37995

- What is the probability that neither sibling is affected?

ANS)

Neither siblings affected { (D_1 = 0, D_2 = 0) }

P (D_1 = 0 and D_2 = 0) = P(D_1=0) * P(D_2=0) = P(D_1=0) ^2 = note: independence

= P(D_1=0) = [P(D_1=0|C_1=M)*P(C_1=M) + P(D_1=0|C_1=F)*P(C_1=F)] ^2 = cond. on M/F

= [(0.5)(0.51) + (1)(0.49)] ^2 = 0.5550525.

- (**HW3.2**) Suppose that a medical test has a 82% chance of detecting a disease if the person has it (i.e., 82% sensitivity) and a 91% chance of correctly indicating that the disease is absent if the person really does not have the disease (i.e., 91% specificity). Suppose that 14% of the population has the disease.

Note: MAKE A TREE

P(Per = +) = 0.14 ➔ P(Per = -) = .86

P(T = + | Per = +) = 0.82 ➔ P(T = - | Per = +) = .18

P(T = - | Per = -) = 0.91 ➔ P(T = + | Per = -) = 0.09

  - What is the probability that a randomly chosen person will test positive?

ANS)

P(Test = +) = P(Test = + |Per = +)*P(Per = +) + P(Test = + | Per = -) * P(Per = -)

= (.82)(.14) + (.09)(.86) = 0.1922

  - Suppose that a randomly chosen person does test positive. What is the probability that the person does have the disease?

ANS)

“Invert the conditioning” then use “Bayes Theorem” x / (x+y) to compute denom.

P (Per = +| T = +) = P (T = + | Per = +) * P(Per = +) / P (T = +) =

Denominator: P(T = +) = [P (T = + | Per = +)P(Per = +) + P (T = + | Per = -) P(Per = -)

P (Per = + | T = +) = (0.82)(0.14) / [(0.82)(0.14) + (0.09)(0.86)] = 0.5972945

- (**HW3.3**) In a certain population of the European starling, there are 5,000 nests with young. The distribution of brood size (number of young in a nest) is given in the accompanying table. Suppose one of the 5,000 broods is chosen at random and let *Y be the size of the brood*. Find

Broad Size	**Brood Number**
1	90
2	230
3	610
4	1300
5	1810
6	800
7	130
8	26
9	3
10	1
Total	**5,000**

  - P(Y=4) = 1300 / 5000 = 0.26
  - P(Y>=8) = [26 + 3 + 1] / 5000 = 0.006
  - P(2<=Y<8) = [230 + 610 + 1300 + 1810 + 800 + 130] / 5000 = 0.976
  - Calculate the mean of the random variable Y.

    [1*90 + 2*230 + … + 9*3 + 10*1] / 5000 = 4.517

- (**HW3.4**) Consider a population of the fruit fly Drosophila melanogaster in which 30% of the individuals are black because of a mutation, while 70% of the individuals have the normal gray color. Suppose three flies are chosen at random from the population; let Y denote the number of black flies out of the three. Then the probability distribution for Y is given by the following table:
  - Find Pr {Y ≥ 1}.

   Pr (Y >= 1) = .441 + .189 + .027 = 0.657 or

   1 – Pr (Y < 1) = 1 - .343 = 0.657

  - Find Pr {Y < 3}.

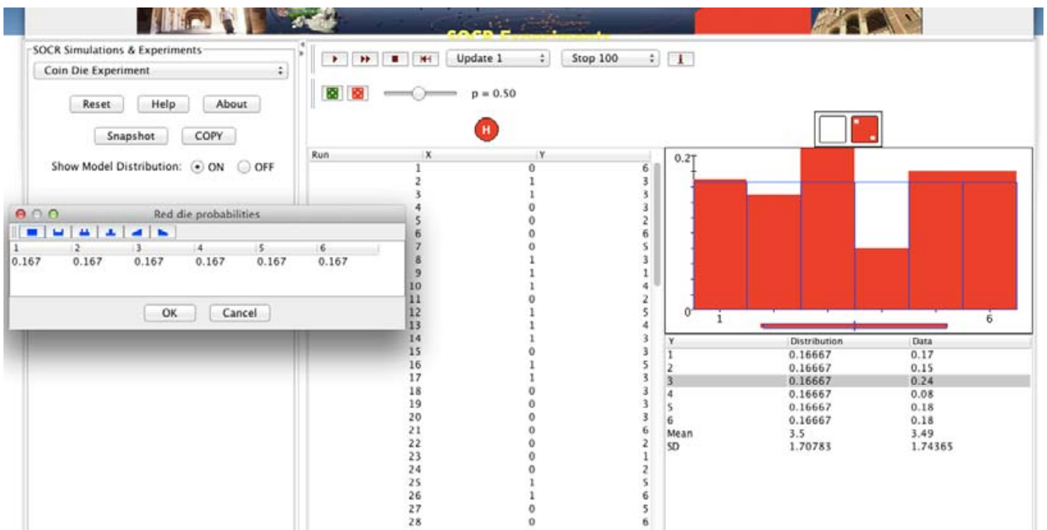
   Pr (Y <3) = 1 – Pr (Y >=3) = 1- 0.027 = .973

  - Calculate the mean of Y.

   0(.343) + 1(.441) + 2(.189) + 3(.027) = 0.9

Y (No. Black)	Probability
0	0.343
1	0.441
2	0.189
3	0.027
Total	**1.00**

- (**HW3.5**) Go to the Coin Die Experiment.
    - Simulate event independence between the outcome of the die (event B) and the outcome of the coin (event A), by setting the probabilities of both dice to be identical. Run 100 experiments and argue that the observed data implies independence between the events A={Coin=Head} and B={Die=3}, i.e., P(A∩B) = P(A) P(B), approximately.



Using the 1-6 flat with p=0.5, we have the above situation.

We have chosen the setting of X == A == Coin to be P(X=1) = 0.5 and P(X=0) = 0.5,

Further, we have chosen the setting of Y == B == Die to be P(Die = d) = 0.16667 for all d.

Since we have the empirical distribution of Y == B == Die, we see that the P(Die=3) = is appx 0.24, where the theoretical should have been 0.1667.

Empirically: P(A∩B) = P(B | A) P(A) = (0.24) * (0.5) = 0.12 = P(B) * P(A)

Hence P(A∩B) = P(B) * P(A) as required, specifically, P(B|A) = P(B)

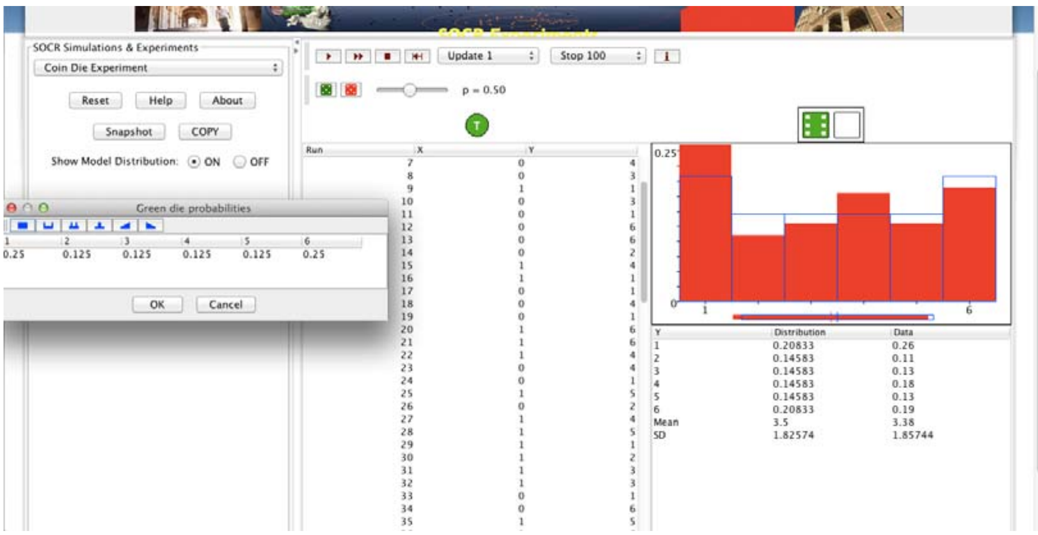
The statement P(B|A) = P(B) is true since knowing the outcome of A "The fair coin where p = 0.5" tells us nothing about B "The die roll" since we chose a 1-6 flat setting – i.e., all the outcomes {1,...,6} are still equally likely (discrete uniformly distributed)

Theoretically: $P(A \cap B) = P(B|A)\,P(A) = (0.16667) * (0.5) = 0.0833 = P(B) * P(A)$

$P(B=3) = P(B=3|A=1)P(A=1) + P(B=3|A=0)P(A=0) = 0.0833(0.5) + 0.0833(0.5) = 0.16667$

$P(A) = 0.5$

- Now make the probability distributions of the two dice different (by clicking on the dice and manually changing the die probabilities). Show empirically the dependence of the probabilities, A={Coin=Head} and B={Die=3}.



Using the 2-5 flat with p=0.5, we have the above situation.

We have chosen the setting of X == A == Coin to be P(X=1) = 0.5 and P(X=0) = 0.5,

Further, we have chosen the setting of Y == B == Die to be

P(Die) = 0.20833 if d=1,6 and 0.14583 if d=2,3,4,5

Although the COIN is still fair where P(X=1) = 0.5, (equally likely to get heads or tails),

We are using different DIE. One of the Die (RED) favors the two numbers {1,6} more than the other numbers {2,3,4,5}.

Check: $P(B \cap A) = P(B|A)\,P(A)$

We will demonstrate the computations using "Theoretical" values, since it is a more standard way of comparing the values. Using empirical values is easily done in the same manner.

$P(A=1) = 0.5 \Rightarrow P(A=0) \Rightarrow 0.5$

$P(B=3) = P(B=3|A=1)P(A=1) + P(B=3|A=0)P(A=0) = (0.14583)(0.5) + (1/6)P(0.5) = 0.1562$

$P(A) * P(B) = 0.5 * 0.1562 = 0.078$

But, $P(B \cap A) = P(B|A)\, P(A) = (0.14583) * (0.5) = 0.07215$

Hence $P(B \cap A) \neq P(B)*P(A)$