http://www.stat.ucla.edu/~dinov/courses_students.dir/10/Winter/STAT13.1.dir/

STAT 13 Homework 3 Solutions

Homework 3 solution

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:

   Before going on let's define the notation. We will let

   - M: a male child born / ¬ M: a female born (could have been "F". No intention of gender discrimination.)
   - D: Have the disease / ¬ D: Do not have disease

   Than, from the problem we can infer

   - P(M) = 0.51 / P(¬ M) = 1-P(M) = 1 – 0.51 = 0.49
   - P(D|M) = 0.5 / P(D| ¬ M) =0

   A) What is the probability that both siblings are affected, if there are one male and one female?

   Since we are told that 2 children are a male and a female

$$\begin{aligned} \text{P(a male affected} \cap \text{a female affected)} &= \text{P(M} \cap \text{D)} \cdot \text{P}(\neg M \cap D) \\ &= 0.5 \cdot 0 = 0 \end{aligned}$$

B) What is the probability that exactly one sibling is affected?

For this problem, because we are not given the gender of the 2 children we have to consider 3 cases; male and male, male and female, and female and female. We can get each of the probability using binomial distribution with n=2 and p=0.51 which gives us

- $\mathrm{P(MM)} = \binom{2}{2} 0.51^2 (1-0.51)^0 = 0.2601$
- $\mathrm{P(MF)} = \binom{2}{1} 0.51^1 (1-0.51)^1 = 0.4998$
- $\mathrm{P(\neg M \neg M)} = \binom{2}{0} 0.51^0 (1-0.51)^2 = 0.2401$

Also note that

- $\mathrm{P(D\neg D \mid \neg M \neg M)} = 0$
- $\mathrm{P(D\neg D \mid M \neg M)} = \binom{1}{1} 0.5^1 (1-0.5)^0 = 0.5$
- $\mathrm{P(D\neg D \mid MM)} = \binom{2}{1} 0.5^1 (1-0.5)^1 = 0.5$

Finally we have to combine the above result

$$\begin{aligned}
\mathrm{P}(D\neg\mathrm{D}) &= \mathrm{P}(D\neg\mathrm{D} \cap \mathrm{MM}) + \mathrm{P}(D\neg\mathrm{D} \cap \mathrm{M}\neg\mathrm{M}) + \mathrm{P}(D\neg\mathrm{D} \cap \neg\mathrm{M}\neg\mathrm{M}) \\
&= \mathrm{P}(D\neg\mathrm{D} \mid \mathrm{MM})\mathrm{P(MM)} + \mathrm{P}(D\neg\mathrm{D} \mid \mathrm{M}\neg\mathrm{M})\mathrm{P(M\neg M)} + \mathrm{P}(D\neg\mathrm{D} \cap \neg\mathrm{M}\neg\mathrm{M})\mathrm{P}(\neg\mathrm{M}\neg\mathrm{M}) \\
&= 0.5 \cdot 0.2601 + 0.5 \cdot 0.4998 + 0 \cdot 0.2401 = 0.37995
\end{aligned}$$

C) What is the probability that neither sibling is affected?

This follows a similar argument as previous problem the difference is

- $\mathrm{P(\neg D \neg D \mid \neg M \neg M)} = 1$
- $\mathrm{P(\neg D \neg D \mid M \neg M)} = \binom{1}{0} 0.5^0 (1-0.5)^1 = 0.5$

- $P(\neg D \neg D \mid MM) = \binom{2}{0} 0.5^0 (1-0.5)^2 = 0.25$

Than we get

$$\begin{aligned} P(\neg D \neg D) &= P(\neg D \neg D \cap MM) + P(\neg D \neg D \cap M \neg M) + P(\neg D \neg D \cap \neg M \neg M) \\ &= P(\neg D \neg D \mid MM)P(MM) + P(\neg D \neg D \mid M \neg M)P(M \neg M) + P(\neg D \neg D \cap \neg M \neg M)P(\neg M \neg M) \\ &= 0.25 \cdot 0.2601 + 0.5 \cdot 0.4998 + 1 \cdot 0.2401 = 0.555025 \end{aligned}$$

2. Suppose that a medical test has a 88% chance of detecting a disease if the person has it (i.e. 88% sensitivity) and a 92% chance of correctly indicating that the disease is absent if the person really does not have the disease (i.e. 92% specificity). Suppose that 15% of the population has the disease.

   Let the above events be denoted as following

   - Po: Test turn out positive / ¬ Po: Do not test positive
   - D: Has disease / ¬ D: Do not have disease

   Then, from the problem we can infer that

   - $P(Po \mid D) = 0.88$
   - $P(\neg Po \mid \neg D) = 0.92$ / $P(Po \mid \neg D) = 1 - P(\neg Po \mid \neg D) = 0.08$
   - $P(D) = 0.15$ / $P(\neg D) = 1 - P(D) = 0.85$

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

Since you can get positive result when you are ill and you can get positive result when you are not ill you have to consider both which translates to:

$$
\begin{aligned}
\mathrm{P(Po)} &= \mathrm{P(Po \mid D)P(D) + P(Po \mid \neg D)P(\neg D)} \\
&= 0.88 \cdot 0.15 + 0.08 \cdot 0.85 = 0.2
\end{aligned}
$$

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

This problem is reverse probability problem which requires straight application of Bayes Theorem

$$
\begin{aligned}
\mathrm{P(D \mid Po)} &= \frac{\mathrm{P(Po \mid D)P(D)}}{\mathrm{P(Po)}} \\
&= \frac{\mathrm{P(Po \mid D)P(D)}}{\mathrm{P(Po \mid D)P(D) + P(Po \mid \neg D)P(\neg D)}} \\
&= \frac{0.88 \cdot 0.15}{0.88 \cdot 0.15 + 0.08 \cdot 0.85} = 0.66
\end{aligned}
$$

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*

A) P(Y=4)

$$P(Y = 4) = \frac{1400}{5000} = \frac{7}{25}$$

B) P(Y>=8)

$$P(Y \geq 8) = P(Y=8) + P(Y=9) + P(Y=10)$$
$$= \frac{26}{5000} + \frac{3}{5000} + \frac{1}{5000} = \frac{3}{5}$$

C) *P(2<=Y<8)*

$$P(2 <= Y < 8) = P(Y=2) + P(Y=3) + P(Y=4) + P(Y=5) + P(Y=6) + P(Y=7)$$
$$= \frac{230}{5000} + \frac{610}{5000} + \frac{1400}{5000} + \frac{1760}{5000} + \frac{750}{5000} + \frac{130}{5000} = \frac{4880}{5000} = 0.976$$

D) Calculate the mean of the random variable Y.

$$\overline{Y} = \sum_{1}^{10} P(Y_i) \cdot Y_i = 1 \cdot \frac{90}{5000} + 2 \cdot \frac{230}{5000} + \cdots + 10 \frac{1}{5000} = 4.487$$

Broad Size	Brood Number	Probability	Size*Probability
1	90	0.0180	0.0180
2	230	0.0460	0.0920
3	610	0.1220	0.3660
4	1400	0.2800	1.1200
5	1760	0.3520	1.7600
6	750	0.1500	0.9000
7	130	0.0260	0.1820
8	26	0.0052	0.0416

9	3	0.0006	0.0054
10	1	0.0002	0.0020
Total	5,000	1	4.487

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:*

   A) Find Pr {Y ≥ 1}.

$$\begin{aligned} P(Y \geq 1) &= P(Y=1) + P(Y=2) + P(Y=3) \\ &= 0.441 + 0.189 + 0.027 = 0.657 \end{aligned}$$

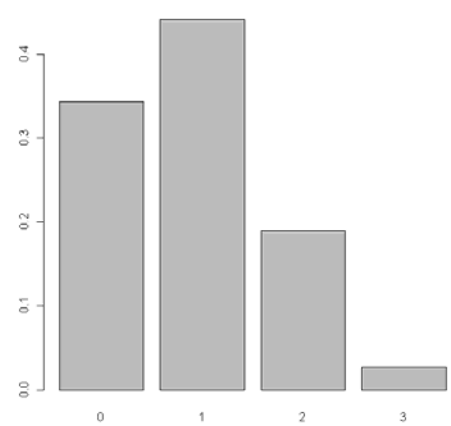
   B) Find Pr {Y < 3}.

$$\begin{aligned} P(Y < 1) &= P(Y=0) + P(Y=1) + P(Y=2) \\ &= 0.343 + 0.441 + 0.189 = 0.973 \end{aligned}$$

   C) Calculate the mean of Y.

If you notice that this is a binomial distribution with n=3 and p=0.3 you could use the formula. Otherwise you can calculate it manually.

$$E(Y) = np = 3 \cdot 0.3 = 0.9$$

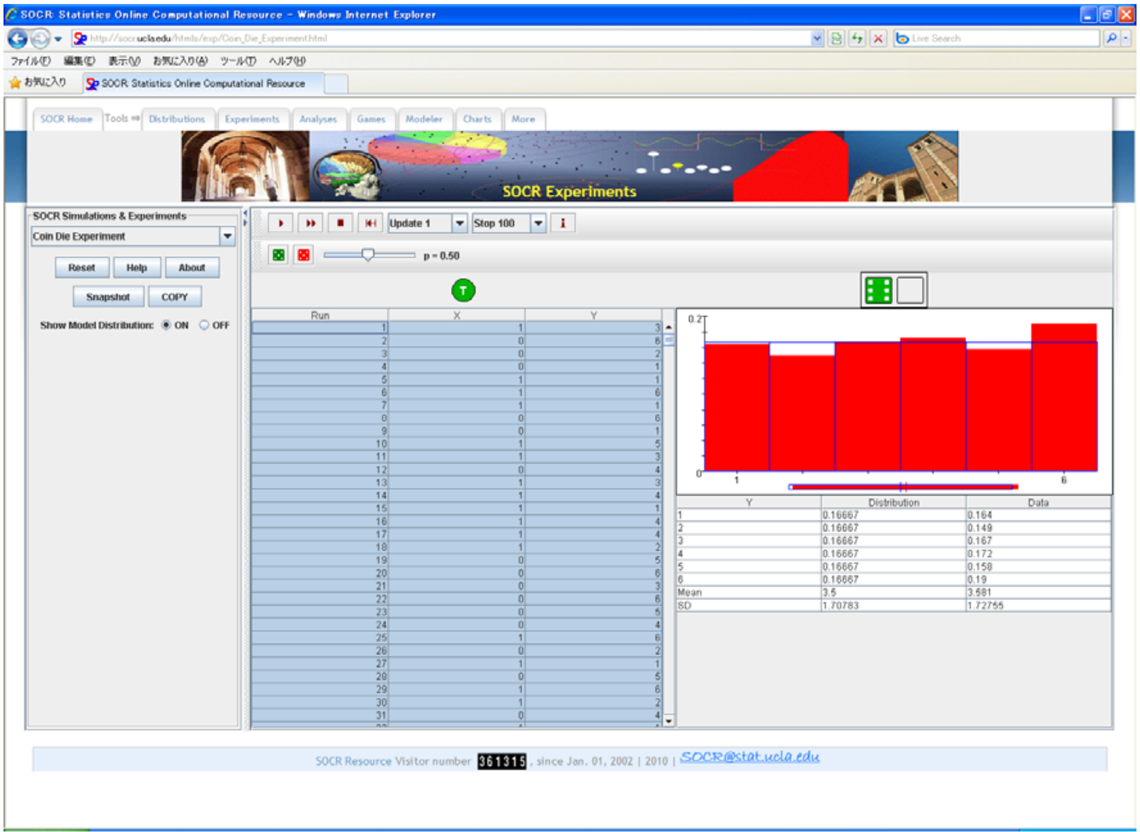


Y (No. Black)	Probability
0	0.343
1	0.441
2	0.189
3	0.027
Total	**1.00**

5. *Go to the [Coin Die Experiment](#).*

   A) *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(AB) = P(A) P(B), approximately.*

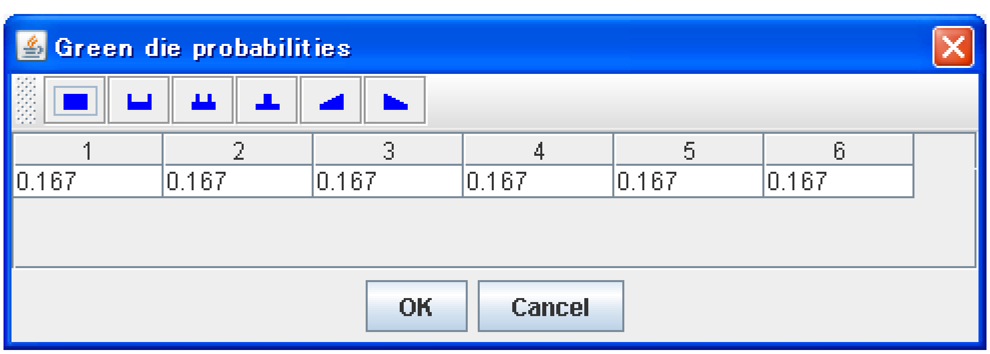
We know that P(A)=1/2 and P(B)=1/6 so P(A)P(B) should roughly equal 1/2*1/6=1/12. I ran SOCR for 1000 runs and got 86 cases of Head and 3. That is 86/1000=0.086. Which is pretty close to 0.083333.



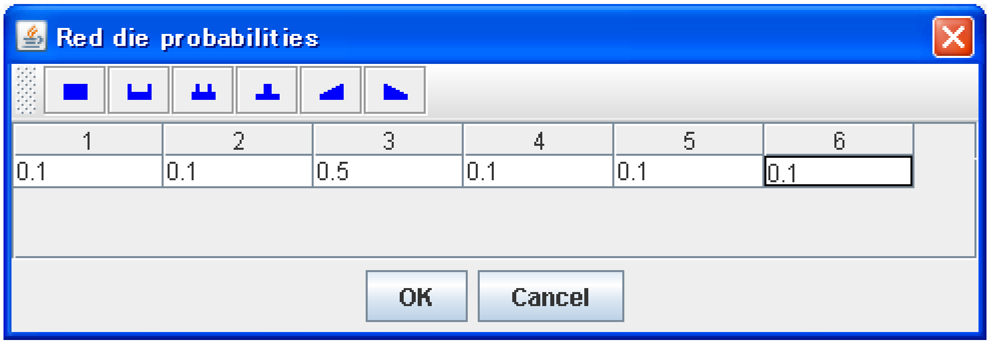
B) Now make the probability distributions of the two dice different (by clicking on the dice and manually changing the die probabilities). Show empirically the dependence of the probabilities, A={Coin=Head} and B={Die=3}.

So I set p=0.1, probability of Green die and Red die as following.

Green die



Red Die



Which makes my theoretical probabilities to be:

- P(H)=0.1
- P(3)=0.1*0.5+(1-0.1)*0.167=0.2003
- P(H and 3) =P(H)P(3|H)=0.1*0.5=0.05

Then I ran the experiment for 1000 times and out of 1000 tries, I got 105 heads, 177 3s, and 41 head and 3. This, in fact is very close to the theoretical probability.

So you see that 0.1*0.2003= 0.02003 is not equal to 0.05 and this is true for the empirical result also.