Statistical inference Homework 2

Abbas Nosrat

November 29, 2021

1-

Assuming there are two seats, there are two possible states.

- The first one sits on the correct seat and the last one sits on the correct seat.
- The first one does not sit correctly and the last one does not as well.

Which means that the probability of the last one to sit correctly is 0.5.

Lets try with three seats.

- The first one sitts at the correct seat and the others sit correctly too.
- The first one sits at the second seat and the second one sits at the first seat.
- The first one sits at the second seat and the second one sits at the third seat.
- The first one sits at the third seat, the second one sits at the second seat and the last one sits at the first.

This leaves four states in two of which, the last passenger sits correctly resulting in the probability of the last passenger sitting correctly to be 0.5.

This can be tested with more seats but the answer will be equal to 0.5

if all the passengers sat randomly, the answer would be $\frac{(n-1)!}{n!} = \frac{1}{n}$ but this is not the case and the answer is 0.5

2-

$$p(A) = 365 \times 364 \times \ldots \times (365 - n + 1) = \frac{365!}{(365 - n)!}$$

$p(A)$ is the probability of no two birthdays being in the same day. $\tilde{p}(A) = 1 - p(A)$ is the probability of at least two people sharing the same birthday. about 65 people makes $\tilde{p}(A)$ around 100 percent.

3-

a)

The time interval is fixed and is equal to one hour, the average number of events which is selling pizza is 20.($\lambda = 20$) Therefore the distribution is Poisson.

b)

$$\sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^{2n}}{(2n)!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \frac{1+(-1)^n}{2} = \frac{e^{-\lambda}}{2} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} + \frac{e^{-\lambda}}{2} \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} = \frac{1+e^{-2\lambda}}{2}$$

lambda is 20 hence $p(even) = \frac{1+e^{-2\lambda}}{2} = \frac{1+e^{-40}}{2} \approx 0.5$

c)

using a for loop from 0 to 10000 with increments of 2, the probability was evaluated using dpois function and summed. Figure 1 is the incremental probability.

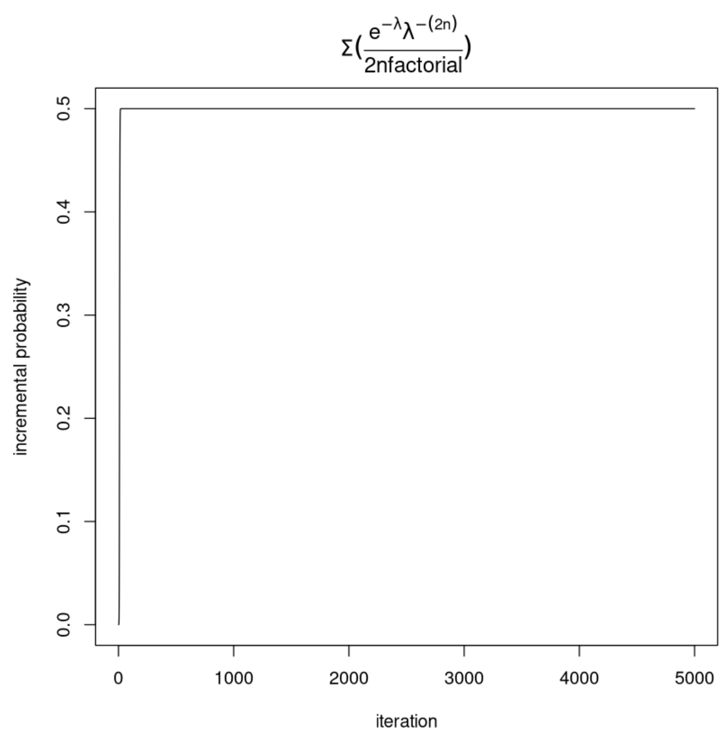

Figure 1: The incremental probability.

4-

a)

There are two distributions and both of them are binomial. The first one is being able to vote or not and the second is voting for x or y. Let the first distribution be $p(X)$ and the second $p(Y)$. Consequently, the total distribution is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(Y = y|X = x) = \binom{x}{y} q^y (1-q)^{x-y}$$

$$P(Y = y) = \binom{n}{y} (pq)^y (1-pq)^{n-y}$$

In this case, $P = 0.5$ and $q = P$ then

$$P(X_A = x) = \binom{N}{x} (0.5p)^x (1-0.5p)^{N-x},$$

$$P(X_B = x) = \binom{N}{x} 0.5(1-p)^x (1-0.5(1-p))^{N-x}$$

Now that the probability of stheuccess is found, the mean value can be estimated as:

$$E(X_A) = \frac{Np}{2}, E(X_B) = \frac{N(1-p)}{2}$$

and the expected value of variance of X_A is $\frac{Np(1-p)}{4}$ and the standard deviation is $\sqrt{\frac{Np(1-p)}{4}}$. The exact same applies for X_B since substitution of p and $1-p$ has no effect due to the product between p and $1-p$.

b)

As N reaches infinity, $\frac{N}{2}$ also becomes infinity. Then

$$E[X] = p \times \lim_{N \to \infty} \frac{N}{2} = Np$$

This is a number between $[0, N]$ to find the fraction of students, Np has to be devided by N which will be equal to p.

c)

In section A the expected value of two binomial distributions which are conditional was prove to be nqp the second binomial is fixed but the first binomial has changed.

The first binomial is the probability of voting if busy and supporting A. its success chance is $0.25q_a$. Another scenario is voting if free and supporting A which has a success of $0.75(1-q_A)$. These two probabilities have or between them. if two distributions are ored, their distributions are summed and expectation of sums is sum of expectations. Hence:

$$E(X_A) = 0.25q_ApN + 0.75(1-q_A)pN$$

By the same logic,

$$E(X_B) = 0.25q_ApN + 0.75(1-q_A)pN$$

5-

a)

Two disjointed events cannot happen at the same time but two independent events can happen although the outcome of one gives no information about outcome of the other

b)

If two events (A,B) are disjointed such that P(A)=1 and P(B)=0 then:

$$P(A \cap B) = 0$$

and

$$P(B|A) = \frac{P(A \cap B) = 0}{P(A)} = 0 = P(B)$$

c)

If N is greater than 2(for example N=3):

IF A is true, then there can be 2 boys one girl or 2 girls and one boy. For the first case, B is false and for the second case B is true. Hence Truth of A gives no information about truth of B

IF A is false, then either the children are all boys or girls. The former results in B being false since there are 3 boys and the latter makes B true since there are zero boys. Therefore, falsehood of A gives no information about B either.

6-

The sample space is [bb,bg,gb,gg]

a)

The mother's response removes "gg" from the sample space. Therefore, the probability of "bb" is $\frac{1}{3}$

b)

$p(Ali|BG) = p(ali|GB) = \alpha, P(ali|BB) = \alpha(1-\alpha) + (1-\alpha)\alpha + \alpha^2 = 2\alpha - \alpha^2$
using Bayes rule $P(BB|ali)$can be found:

$$p(BB|ali) = \frac{p(ali|BB)p(BB)}{P(ali)} = \frac{0.25(2\alpha - \alpha^2)}{0.25(2\alpha - \alpha^2) \times 0.5\alpha} = \frac{2-\alpha}{4-\alpha} \approx 0.5$$

c)

In both parts the sample space was reduced to BB BG and GB but in part B states did not have equal probability. The added information about name means that a family with two boys is more likely to name one of their children Ali than a family with one boy and one girl.

7-

a and b)

Suppose there are N players. When two players meet in round , they are effectively the chosen 2 from a pool of 2^k players who competed in the sub-bracket leading to that particular match. Let us call such a sub-bracket a k-sub-bracket. There are $M_k = 2^N/2^k$ k-sub-brackets. The probability that both players end up in a particular k-sub-bracket is

$$P_1(k) = \frac{2^k}{2^N} \cdot \frac{2^k - 1}{2^N - 1}.$$

The probability that two players from a k-sub-bracket meet in round k is

$$P_2(k) = 2 \cdot \frac{1}{2^k} \cdot \frac{1}{2^k - 1} = \frac{1}{2^{k-1}} \cdot \frac{1}{2^k - 1}.$$

Thus, the probability that the two players play each other is

$$P_N = \sum_{k=1}^{N} M_k \cdot P_1(k) \cdot P_2(k) = \sum_{k=1}^{N} \left(\frac{2^N}{2^k}\right)\left(\frac{2^k}{2^N}\right)\left(\frac{2^k - 1}{2^N - 1}\right)\left(\frac{1}{2^{k-1}}\right)\left(\frac{1}{2^k - 1}\right) = \frac{1}{2^N - 1}\sum_{k=1}^{N}\frac{1}{2^{k-1}} = \frac{1}{2^{N-1}}.$$

for $N = 4$

$$\frac{1}{8}$$

8

a)

There is a single outlier which ruins the plot at index 1429.

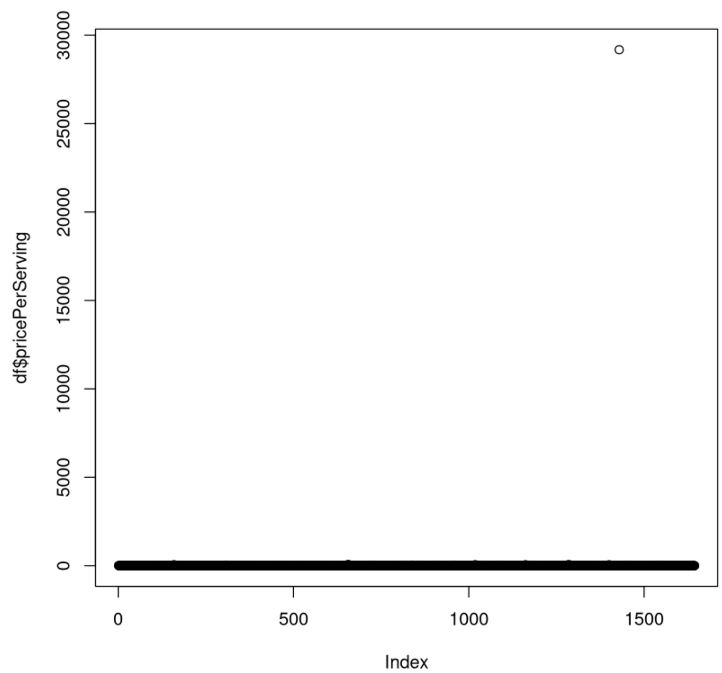

Figure 2: outlier demonstration

After removing the outlier the plot looks like

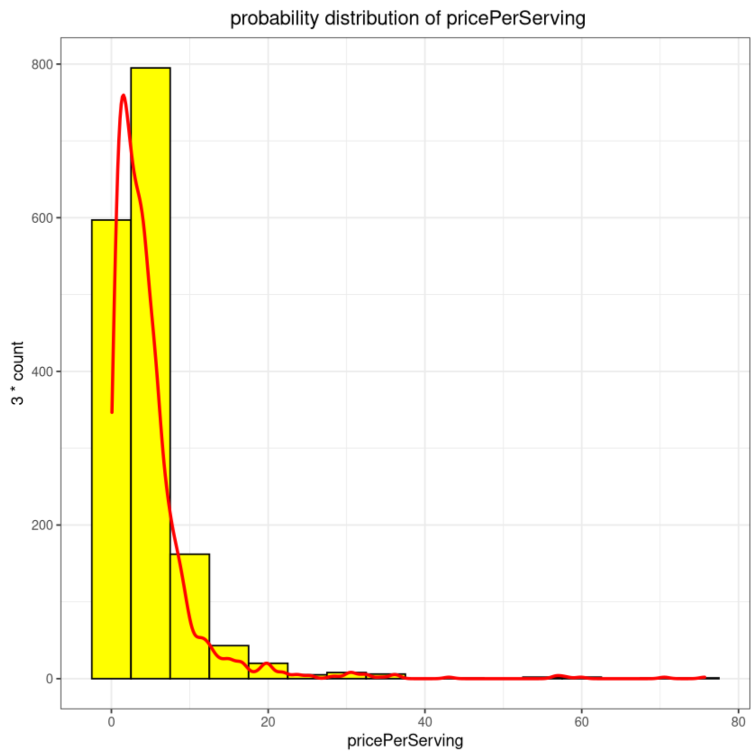

Figure 3: Histogram of price per serving with density line

b)

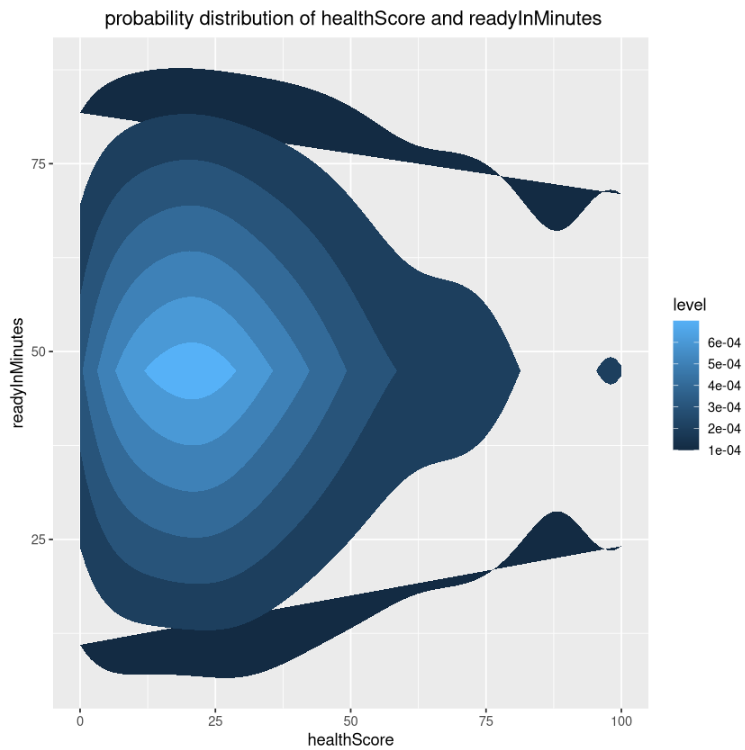

Figure 4: probability distribution of healthScore and readyInMinutes

c)

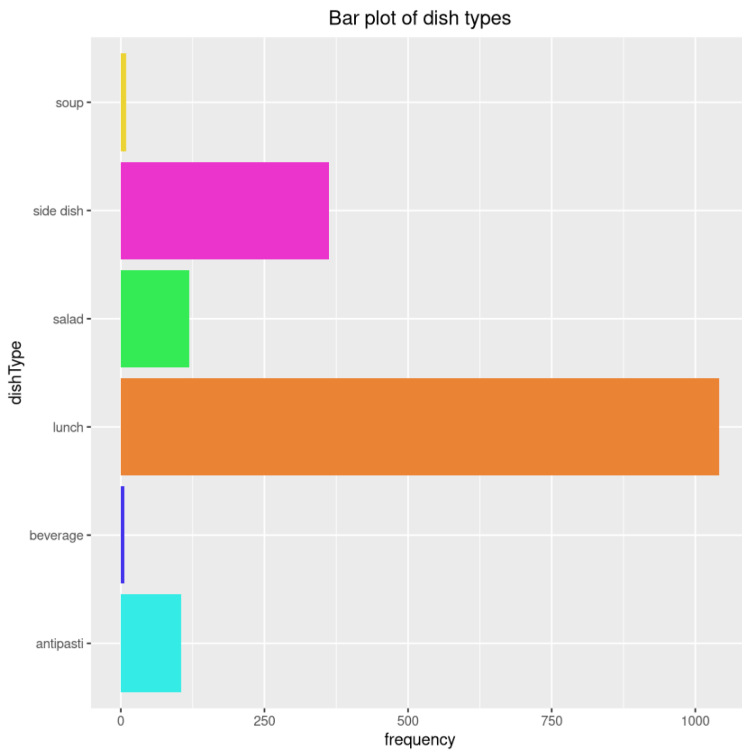

Figure 5: Bar plot of dish types

d)

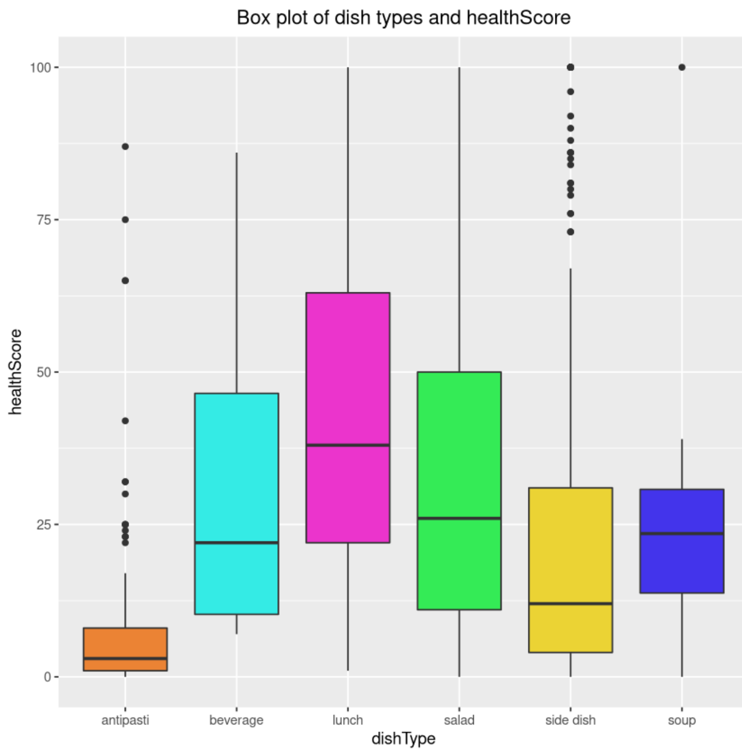

Figure 6: Box plot of dish types and healthScore

e)

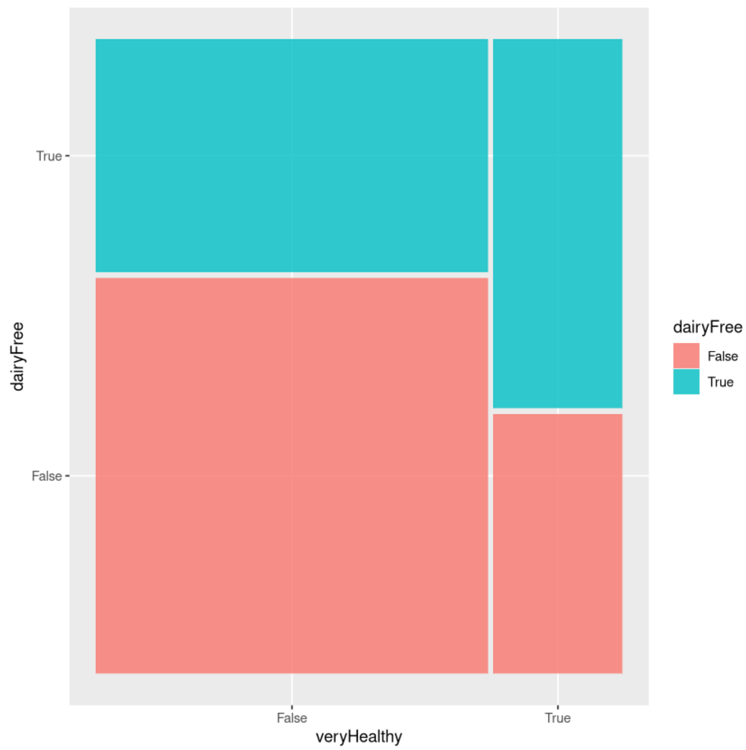

Figure 7: mosaic plot of dairyFree and veryHealthy