Homework 2 Solution

Statistical Inference, Spring 1400

1- True/False Problems. Circle either True or False (and explain/correct if False):

a. If a fair coin is tossed many times and the last eight tosses are all heads, then the chance that the next toss will be heads is somewhat less than 50%.

False. These are independent trials

b. If events X and Y are independent, then they are also mutually exclusive.

False. Independence and mutual exclusivity are separate concepts.

c. The General Addition Rules states that the probability of either event X or event Y occurring is the sum of P(X) with P(Y).

False. P(X or Y) = P(X) + P(Y) – P(X and Y).

d. P(A | B) = P(B | A).

False. The events cannot be reversed around the conditional probability.

e. A probability tree is one way to represent a sample space.

True

f. If P(A and B) = 0, then either P(A) or P(B) must also be equal to zero.

False. P(A and B) = 0 means that the two events are mutually exclusive.

g. The mean of a Poisson distribution is not always equal to its variance.

False. always equal

h. If the binomial probability of success is near 0.5, then the distribution is approximately symmetric.

True

i. The probability of a student randomly guessing answers to a true/false exam is best modeled with a binomial distribution.

True

2- In France Ministry of Health reports daily number of covid-19 cases. The 2020 France final report indicates that 14.6% of French people are infected to covid-19, 20.7% speak a language other than French at home, and 4.2% fall into both categories.

Considering P as the event of being infected to covid-19, and let N as the event of speaking a language other than French at home.

a. Are Rate of infected people to covid-19 and speaking a language other than French at home disjoint?

$P(P)=0.146$; $P(N)=0.207$; $P(P\cap N)=0.042$

No, because 4.2% of the population is infected to covid-19 and speaks a language other than French at home.

b. What percent of France infected to covid-19 and only speak French at home?
$P(P\cap NC)=P(P)-P(P\cap N)=0.146-0.042=0.104$
So, the answer is 10.4%

c. What percent of France infected to covid-19 or speak a language other than French at home?

$P(P\cup N)=P(P)+P(N)-P(P\cap N)=0.146+0.207-0.042=0.311$ or 31.1%

d. What percent of France not infected to covid-19 and only speak French at home?

Let PC be the complement of event P

$P(PC\cap NC)=P(NC)-P(P\cap NC)=(1-0.207)-0.104=0.689$

So, the answer is 68.9%.

e. Is the event that someone get covid-19 independent of the event that the person speaks a language other than French at home?

If P and N are independent, $P(P\cap N)=P(P)\times P(N)$. $P(P\cap N)=0.042$.
$P(P)\times P(N)=0.146\times 0.207=0.030$. $P(P\cap N)\neq P(P)\times P(N)$ so P and N are not independent.

3- There are 64 teams in a football league. These teams compete in pairs in the first round and the winners go to the second round and again they compete in pairs. In the same way, these competitions will continue as long as there is only one team left. Given that 64 teams in the league the games of this tournament last 6 rounds. Someone has taken part in a challenge to predict the outcome of this league. He must present his prediction of all the games at the beginning of the season and at the end of the season earns points based on the predictions he made at the beginning of the season. The law of challenge is such that for every correct prediction in the first round gets 1 point. In each subsequent round of the tournament, the score is doubled from the previous round (if you correctly predict a match in the second you get 2 points and so on) This person has no information about these teams, so he decides to predict all the games by tossing a fair coin. Calculate expected value, E(X) for his predictions.

راه اول:

اگر فرض کنیم که 2^n تیم در این لیگ وجود دارد در آن صورت n دور مسابقات در این لیگ داریم. در دور اول 2^{n-1} بازی انجام می شود. در دور دوم 2^{n-2} بازی تا دور n ام که یک بازی خواهیم داشت. پس تعداد کل بازی ها $2^{n-1} + 2^{n-2} + \cdots + 2 + 1$ است. حالال یک بازی خاص را مانند g تصور کنید. I_g متغیری است که مشخص می کند فرد امتیاز این بازی را می گیرد یا خیر. اگر این بازی در دور r=r(g) ام انجام شود میزان امتیازی که این فرد کسب خواهد کرد برابر با 2^{r-1} است . پس به طور متوسط امتیز کسب شده از این بازی برابر است با:

Excepted points g=$E[2^{r-1}.\ I_g] = 2^{r-1}E[I_g] = 2^{r-1}P_g$

که p_g احتمال تشخیص برنده این بازی است. تشخیص برنده این بازی به منزله ی تشخیص درست نتیجه تمام بازی های قبلی آن تیم برنده در دور r ام است. پس $p_g = 2^{-r}$ و

Excepted points g= $E[2^{r-1}.\ I_g] = 2^{r-1}E[I_g] = 2^{r-1}P_g = 2^{r-1}2^{-r} = 1/2$

که مستقل از بازی است. امید ریاضی امتیاز کسب شده از هر بازی برابر است با ½ . با توجه به اینکه 2^{n-1} بازی در لیگ داریم امید ریاضی امتیاز فرد برابر است با $\frac{2^{n-1}}{2}$ خواهد بود. در این لیگ 64 تیم داریم یعنی n=6 پس:

Excepted points=$\frac{2^{6-1}}{2}$=31.5

راه دوم:

Each coin toss has a probability of 1/2

Possibilities	1/2	1/4	1/8	1/16	1/32	1/64
Competitions	32	16	8	4	2	1
points	1	2	4	8	16	32

$$E[X]=\sum_{i=1}^{6} x_i p_x(x_i) = (1*32*\tfrac{1}{2}) + (2*16*\tfrac{1}{4}) + \left(4*8*\frac{1}{8}\right) + \left(8*4*\frac{1}{16}\right) + \left(16*2*\frac{1}{32}\right) + (32*1*\tfrac{1}{64}) = 31.5$$

4- To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event LC), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^C) for j ∈ {1,2}. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that P(M_j|L) =P(M_j^C| LC) = 9/10.

Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

a. Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

By Bayes' rule,

$$\mathrm{P(L|M1)} = \frac{P(\mathrm{M1|L})P(L)}{P(\mathrm{M1})} = \frac{\frac{9}{10}*\frac{1}{10}}{\frac{9}{10}*\frac{1}{10}+\frac{1}{10}*\frac{9}{10}} = \frac{1}{2}$$

b. Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

By Bayes' rule,

$$\mathrm{P(L|M1,M2)} = \frac{P(\mathrm{M1} \text{ و } \mathrm{M2|L})P(L)}{P(\mathrm{M1} \text{ و } M2)} = \frac{(\frac{9}{10})^2*\frac{1}{10}}{(\frac{9}{10})^2*\frac{1}{10}+(\frac{1}{10})^2*\frac{9}{10}} = \frac{9}{10}$$

c. Bob runs the 1st program and M1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{p}(A) = P(A|M1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{p}(L|M2) = P(L|M1 \cap M2)$: is conditioning on $M1 \cap M2$ in one step equivalent to first conditioning on M1, then updating probabilities, and then conditioning on M2?

Yes, they are the same, since Bayes' rule is coherent. The probability of an event given various pieces of evidence does not depend on the order in which the pieces of evidence are incorporated into the updated probabilities.

$$\tilde{p}(L|\mathrm{M2}) = \frac{p(M2|L)\tilde{p}(L)}{p(M2|L)\tilde{p}(L) + p(M2|L^c)\tilde{p}(L^c)}$$

$$\tilde{p}(L) = \mathrm{p(L|M1)} = \frac{p(M1|L)p(L)}{p(M1)}$$

$$\tilde{p}(L^c) = \mathrm{p}(L^c|\mathrm{M1}) = \frac{p(M1|L^c)p(L^c)}{p(M1)}$$

$$\tilde{p}(L|\mathrm{M2}) = \frac{p(M2|L)\frac{p(M1|L)p(L)}{p(M1)}}{P(M2|L)\frac{P(M1|L)P(L)}{P(M1)}+P(M2|L^c)\frac{P(M1|L^c)P(L^c)}{P(M1)}}$$

$$= \frac{P(M2|L)P(M1|L)P(L)}{P(M2|L)P(M1|L)P(L)+P(M2|L^c)P(M1|L^c)P(L^c)} = \mathrm{p}(L|M1 \cap M2)$$

5- One pharmaceutical company makes Cancer pills in boxes of 100. If QC (quality control) says that 0.5% of the Cancer pills are damaged, then what percent of the boxes will:

As n is large and p, the P (defective Cancer pills), is small, use the Poisson approximation to the binomial probability distribution. If X= number of defective Cancer pills in a box, then:

a. no damage?

$X \sim P(\mu)$ where $\mu = n \times p = 100 \times 0.005 = 0.5$

$P(X=0) = \frac{e^{-0.5}(0.5)^0}{0!} = \frac{e^{-0.5}(1)}{1} = 0.6065 \approx 61\%$

b. 2 or more damaged?

P (X= 2 or more) =P (X= 2) +P (X= 3) +P(X= 4) +...but it is easier to consider:

$P(X \geq 2) = 1 - [P(X=0) + P(X=1)]$

$P(X=1) = \frac{e^{-0.5}(0.5)^1}{1!} = \frac{e^{-0.5}(0.5)}{1} = 0.3033$

i.e. $P(X \geq 2) = 1 - [0.6065 + 0.3033] = 0.0902 \approx 9\%$

6- The travel agency knows that over the long run, 90% of passengers who reserve seats will show up for their trip. On a particular trip with 300 seats, the travel agency accepts 324 reservations. Use then normal approximation to answer question:

a. What is the chance that the trip will be overbooked? Event of a passenger will show up is independent of one another.

The travel agency will be overbooked if more than 301 or more passengers show up. Since passengers show up independently of one another and each passenger has probability 0.9 of showing up, the probability of more than 301 (inclusive) passengers showing up is a binomial probability. Let S=the number of passengers who show up. Since n= 324, the number of trials is large, and we can use the normal approximation. As in the previous problem, we have

$P(S \geq 301) = P(S \geq 301 - 0.5)$

$$P(S \geq 301) = P\left(\frac{S-(324)(0.9)}{\sqrt{324(0.9)(0.1)}} \geq \frac{301-0.5-(324)(0.9)}{\sqrt{324(0.9)(0.1)}}\right)$$

$$\approx 1 - \Phi\left(\frac{301-0.5-(324)(0.9)}{\sqrt{324(0.9)(0.1)}}\right) = 0.0495$$

by evaluating Φ from the table in Appendix 5.

b. Regarding previous question now assuming that passengers always trip in couple “and trip only if both people in the couple show up”. compare that your answer to the previous question is consistent. The probability of overbooking will increase or decrease? Explain.

Suppose passengers travel in pairs and that the probability that the pair travels is $(0.9)^2$= 0.81, and the probability that pair does not travel is 0.19. We have 162 pairs, each with probability 0.81 of showing up, so μ= 131.22 and $\sigma=\sqrt{162(0.81)(0.19)}$ = 4.993. If the travel agency is to avoid being overbooked, 150 pairs or fewer must show up. We get

$$P(S\geq 151) = P(S\geq 151-0.5)$$

$$P(S\geq 151)=P\left(\frac{S-\mu}{\sigma} \geq \frac{151-0.5-\mu}{\sigma}\right)$$

$$\approx 1-\Phi\left(\frac{151-0.5-131.22}{4.993}\right)= 1-\Phi(3.86)\approx 0.0001$$

the probability of overbooking to decrease.

Since the increase in σ happens in square-root land, while the decrease in μ does not. Therefore, we expect the probability of overbooking to decrease.

8- N persons gave their raincoats and their umbrellas to the doorman at the entrance of the Marlinspike” mansion in order to attend to the party. At the end of the night, the doorman was completely plastered. So, he gave the leaving guests a random raincoat and a random umbrella, in a way that each person got a pair of raincoat and umbrella in a uniformly random manner.

a. What is the probability that nobody gets back his own raincoat and his umbrella?

$p_{umbrella}$=the probability that a specific person gets his own umbrella

$p_{raincoat}$=the probability that a specific person gets his own raincoat

p_{nobody}=the probability that nobody gets his/her own raincoat and his/her umbrella

$$p_{nobody} = \left((1-p_{umbrella})(1-p_{raincoat})\right)^N = \left(1-\frac{1}{N}\right)^{2N}$$

b. What is the probability that everybody gets at least his own raincoat or his own umbrella? Calculate this probability when N goes to infinity.

$P_{at\ least}$=the probability that everybody gets at least his own raincoat or his umbrella

$$p_{at\ least} = (p_{umbrella} + p_{raincoat} - p_{umbrella} \times p_{raincoat})^N = \left(\frac{2}{N}-\frac{1}{N^2}\right)^N$$

$$\lim_{N\to\infty}\left(\frac{2}{N}-\frac{1}{N^2}\right)^N = 0$$

c. If each guest can find his/her own raincoat with probability p=0.2 and find his/her umbrella with probability p=0.1, then what is the probability that N/2 persons neither get their raincoats nor their umbrella right, given that the other N/2 persons have already received their belongings correctly. (N is divisible by 2).

For this question, one does not need to use Bayesian rule. If you carefully inspect the condition of the requested probability, you will see that this question is very close to part a. But, instead of N to be the number of the guests as part a, there are only N/2 guests. So:

$$p_{nobody\ for\ \frac{N}{2}} = \left((1 - p_{umbrella})(1 - p_{raincoat})\right)^{\frac{N}{2}} = (0.9 \times 0.8)^{\frac{N}{2}} = 0.84^{N}$$

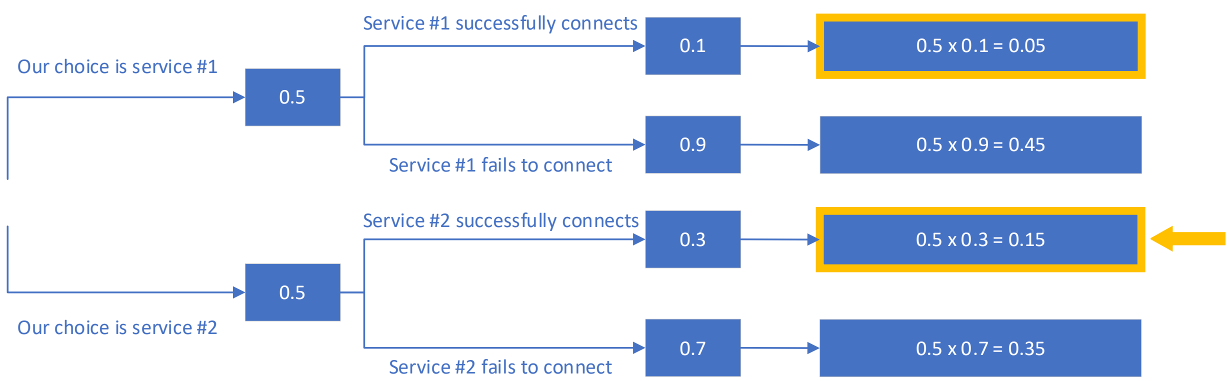
9- Suppose you desperately need to access Telegram and the VPN service you were buying subscriptions from, has been blocked entirely. You now have to choose from one the two free VPN services that are still working. But, as you may know, the free VPN services are very slow and does not necessarily connect whenever you request a connection. The first service connects successfully 10% of the time, and the other one connects 30% of the time. You want to use the service that has more chance of establishing a VPN connection. However, you don't know which one is better. So, initially you assume that both VPN services are equally likely to be the better one. Then:

a. You randomly try one of the services. Given that you managed to establish a connection, what is the probability that the VPN service you've tried was the better one?

We define two random variables whose values are complementary.

S_1: Our selection is VPN service #1, S_2: Our selection is VPN service #2

Initial belief: $p(S_1) = p(S_2) = 0.5$



According to the probability tree above, we want to calculate the probability that we had selected the better service, given that our selection led to a successful connection. In other word, we are looking for:

$$P(S_2 \mid the\ selected\ VPN\ service\ connected\ successfully) = \frac{0.15}{0.15 + 0.05} = 0.75$$

11- In one of the ancient tribes of Maya civilization, there was a popular game called "Buga-Uga". In this game, the player picks a knife and stands beside a special table designed for this game. The player then tosses the knife to the air in a way that the knife would equally likely land anywhere on the table but it never lands on the edges (for simplicity, you can assume that the table is very large). Whenever the knife lands on the table, it leaves a trace with the shape of a line of length K. The surface of the table is made of wooden strips of two types. Below, you can see a small portion of the table from the top. The width of each wooden strip is W (W > K).

If the player tosses the knife and it lands in such a way that the trace is left only in a single strip, the player wins. What is the probability that the player wins three successive rounds?

Let x be the distance from the center of the trace of the knife to the closest boundary of the wooden strips, and θ be the acute angle between the trace and a strip boundary. The uniform probability density function of x between 0 and W/2 is:

$$\begin{cases} \frac{2}{W}; & if\ 0 \le x \le \frac{W}{2} \\ 0; & otherwise \end{cases}$$

The uniform probability density function of θ between 0 and π/2 is:

$$\begin{cases} \frac{2}{\pi}; & if\ 0 \le \theta \le \frac{\pi}{2} \\ 0; & otherwise \end{cases}$$

So, the joint probability density function of the independent random variables x and θ is the product of their individual probability density functions:

$$\begin{cases} \frac{4}{W\pi}; & if\ 0 \le \theta \le \frac{\pi}{2}, 0 \le x \le \frac{W}{2} \\ 0; & otherwise \end{cases}$$

The knife trace will cross a boundary if $x \leq \frac{K}{2}\sin(\theta)$. So, by integrating the joint probability distribution over the interested area, we can compute the probability of whether the trace has crossed a boundary or not as follows:

$$p_{crossed} = \int_0^{\frac{\pi}{2}} \int_0^{\frac{K}{2}\sin(\theta)} \frac{4}{W\pi}\, dx\, d\theta = \frac{2K}{W\pi}$$

So, the probability for a player to win three succussive rounds equals:

$$p_{win} = (1 - p_{crossed})^3 = \left(1 - \frac{2K}{W\pi}\right)^3$$