MTH 7241 Fall 2022: Prof. C. King

Notes 2: Conditioning

$\mathbb{P}(B \cap A) = \mathbb{P}(B|A)\mathbb{P}(A).$

Conditional probability

$$\mathbb{P}(B|A) = \text{conditional probability that } B \text{ is true given that } A \text{ is true}$$

This is computed using the formula

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}, \quad \mathbb{P}(A) \neq 0$$

It is important to note that $\mathbb{P}(B|A)$ is defined only if $\mathbb{P}(A) \neq 0$.

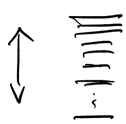
Example 1 *The probability of drawing an Ace from a standard deck of cards is*

$$\mathbb{P}(Ace) = \frac{4}{52} = \frac{1}{13}$$

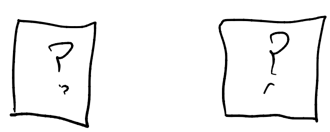
Draw two cards in sequence, and let A_1, A_2 be the events that the first, second cards are Aces respectively, then it is easy to see that

$$\mathbb{P}(A_1) = \frac{4}{52}, \quad \mathbb{P}(A_2|A_1) = \frac{3}{51}, \quad \mathbb{P}(A_2|A_1^c) = \frac{4}{51}$$

52 cards



Draw two cards without replacement



4 Aces.

Card 1 Card 2

$A_1 = \{\text{Card 1 is Ace}\}$

$A_2 = \{\text{Card 2 is Ace}\}.$

$\mathbb{P}(A_1) = \frac{4}{52} = \frac{1}{13}.$

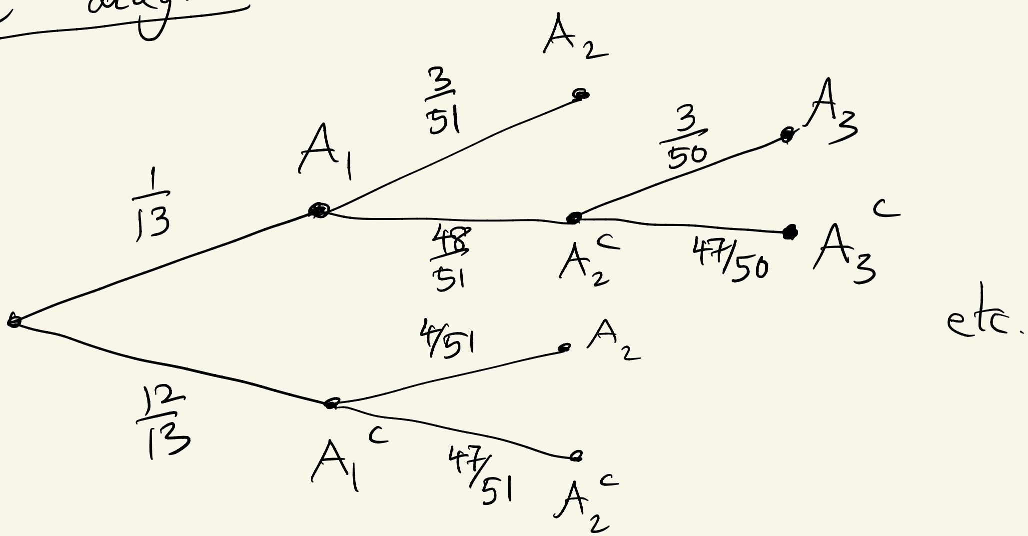
$\mathbb{P}(A_2|A_1) =$ conditional prob. that second card is Ace given that first card is Ace.

$$= \frac{3}{51}$$

$\mathbb{P}(A_2 \mid A_1^c)$ = cond. prob A_2 given that A_1 is false

$$= \frac{4}{51}$$

Tree diagram.



$$\mathbb{P}(A_1) = \frac{1}{13}$$

$$\mathbb{P}(A_2^c \mid A_1) = \frac{48}{51}$$

$$\mathbb{P}(A_3 \mid A_2^c \cap A_1) = \mathbb{P}(A_3 \mid A_2^c, A_1)$$

$$= \frac{3}{50}$$

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \ \underline{\text{and}} \ A_2)$$ joint probability

$$= \underbrace{\mathbb{P}(A_2 \mid A_1)\, \mathbb{P}(A_1)}$$

- this is the definition of conditional prob.
- product of probs. along the branches of the tree.

$$= \frac{3}{51} \cdot \frac{1}{13}$$

$$\mathbb{P}(A_1^c \cap A_2) = \text{joint prob.}$$

$$= \mathbb{P}(A_2 \mid A_1^c)\, \mathbb{P}(A_1^c)$$

$$= \frac{4}{51} \cdot \frac{12}{13}$$

Total probability formula

$$\mathbb{P}(A_2) = \mathbb{P}(A_2 \cap A_1) + \mathbb{P}(A_2 \cap A_1^c)$$

$$\boxed{\mathbb{P}(A_2) = \mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_1) + \mathbb{P}(A_2 \mid A_1^c)\,\mathbb{P}(A_1^c)}$$

Total probability formula

$$\Rightarrow \mathbb{P}(A_2) = \frac{3}{51} \cdot \frac{1}{13} + \frac{4}{51} \cdot \frac{12}{13}$$

$$= \frac{1}{13}$$

So $\mathbb{P}(A_2) = \mathbb{P}(A_1) = \frac{1}{13}$

similarly $\mathbb{P}(A_k) = \frac{1}{13}$ for any

$k = 1, 2, 3, \ldots, 52$.

Total probability

$$\mathbb{P}(A_2) = \mathbb{P}(A_2 | A_1)\, \mathbb{P}(A_1) + \mathbb{P}(A_2 | A_1^c)\, \mathbb{P}(A_1^c)$$

Note: $A_1 \cup A_1^c = S$ (sample space = all outcomes)

$A_1 \cap A_1^c = \phi$ (null event = empty set)

- events A_1, A_1^c are exhaustive
- events A_1, A_1^c are mutually exclusive

Generalize:

events $B_1, B_2, \ldots, B_n$ are

- exhaustive $B_1 \cup B_2 \cup \cdots \cup B_n = S$.

- mutually exclusive = disjoint

$$B_i \cap B_j = \phi \quad \text{any } i, j$$

Total prob. formula:

for any event A

$$\boxed{\begin{aligned}\mathbb{P}(A) = \ &\mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\,\mathbb{P}(B_2) \\ &+ \cdots + \mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n).\end{aligned}}$$

Total prob. formula

Note: such events arise naturally from a discrete random variable X:

$$\text{Ran}\, X = \{x_1, x_2, \dots, x_n\}.$$

Let $B_i = \{X = x_i\}$

Then the events $\{B_1, B_2, \dots, B_n\}$ are exhaustive and mutually exclusive.

In the above example it is perhaps not immediately obvious how to compute $\mathbb{P}(A_2)$. We can use the formula for total probability, which in this case says that

$$\begin{aligned}\mathbb{P}(A_2) &= \mathbb{P}(A_2 \,\text{and}\, A_1) + \mathbb{P}(A_2 \,\text{and}\, A_1^c) \\ &= \mathbb{P}(A_2|A_1)\,\mathbb{P}(A_1) + \mathbb{P}(A_2|A_1^c)\,\mathbb{P}(A_1^c) \\ &= \frac{3}{51}\,\frac{4}{52} + \frac{4}{51}\,\frac{48}{52} \\ &= \frac{4}{52}\end{aligned}$$

The general formula for total probability is this: suppose that there is a collection of events $A_1, A_2, \ldots, A_n$ which are mutually disjoint, so $A_i \text{and} A_j = \emptyset$ for all $i \neq j$, and also exhaustive, meaning they include every outcome in the sample space S, so that $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Then for any event B,

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \cdots + \mathbb{P}(B \cap A_n) \\ &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)P(A_n)\end{aligned}$$

$A = \{\text{applicant is non-smoker}\}$

$B = \{\text{applicant says they are non-smoker}\}$

$\mathbb{P}(A|B) = \mathbb{P}(\text{app is non-smoker given say are non-smoker}\}$.

Example 2 Bayes Rule *is a useful application of conditional probability. The formula is*

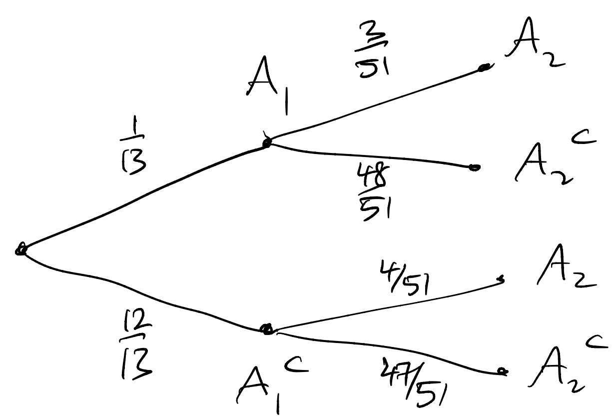
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}$$

Suppose an insurance application contains the question "Is the applicant a smoker?". Assume that 30% of the population smokes, and that 40% of smokers will lie about it. Assuming no non-smokers will lie, what percentage of applicants who say they are non-smokers actually are non-smokers?

<u>Situation</u>: know $\mathbb{P}(A|B)$ and $\mathbb{P}(A|B^c)$

but want $\mathbb{P}(B|A)$ and $\mathbb{P}(B|A^c)$.

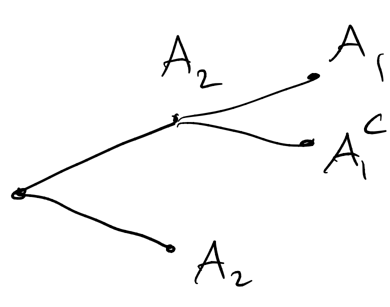
<u>Ex</u>: Draw 2 cards.



Find $\mathbb{P}(A_1 | A_2)$.

Use definition to start:

$$\mathbb{P}(A_1 | A_2) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)}$$



$$= \frac{\left(\frac{3}{51}\right)\left(\frac{1}{13}\right)}{\left(\frac{1}{13}\right)} = \frac{3}{51}$$

Bayes Rule

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)} \qquad \text{(definition)}$$

$$= \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} \qquad \text{(trivial)}$$

$$= \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A)} \qquad \text{(definit)}$$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\;\mathbb{P}(B)}{\mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)}$$

(total prob)

Example (non-smoker question)

First step: name the events

Second step: identify question $\mathbb{P}(A|B)$.

$A = \{$ app. is non-smoker$\}$

$B = \{$ app. says they are non-smoker$\}$.

Third step: apply Bayes Rule

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B|A)\,\mathbb{P}(A) + \mathbb{P}(B|A^c)\,\mathbb{P}(A^c)}$$

$\mathbb{P}(B|A) = 1$ (non-smokers don't lie!)

$\mathbb{P}(B|A^c) = 0.4$ (40% of smokers lie)

$\mathbb{P}(A) = 0.7$ (30% of pop. smokes)

$\mathbb{P}(A^c) = 0.3$

$$\Rightarrow \mathbb{P}(A|B) = \frac{(1)(0.7)}{(1)(0.7) + (0.4)(0.3)} = 0.85$$

If the events A, B are independent then the formula $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ implies that both of the following are true:

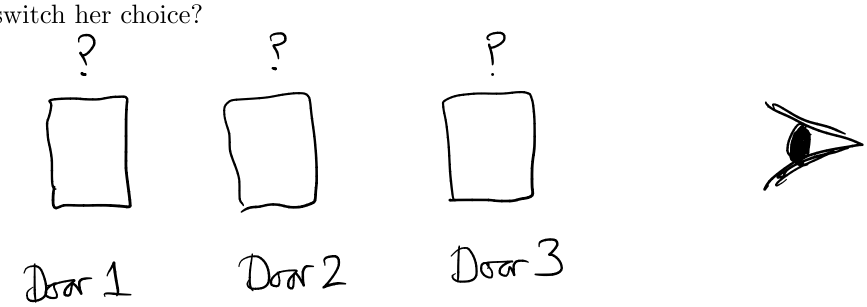
$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(B|A) = \mathbb{P}(B)$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

$$\Rightarrow \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

doesn't care if B is true or not.

The famous *Monty Hall* gameshow problem. There are 3 doors, and a prize is hidden behind one door. The contestant chooses a door. The host then opens *one of the other doors* to show that it does not conceal a prize. The contestant may now change her choice to the third remaining door. Should she switch her choice?



T = door concealing the prize

G = door first guessed by contestant.

S = {contestant wins after switching}.

Total probability:

$$\mathbb{P}(S) = \mathbb{P}(S \mid G = T)\, \mathbb{P}(G = T) + \mathbb{P}(S \mid G \neq T)\, \mathbb{P}(G \neq T)$$

$$= 0 \cdot \left(\frac{1}{3}\right) + 1 \cdot \left(\frac{2}{3}\right)$$

$$\mathbb{P}(S) = \frac{2}{3}$$

$D = \{$contestant wins without switching$\}$.

$$\mathbb{P}(D) = \mathbb{P}(D \mid G = T)\, \mathbb{P}(G = T)$$

$$+ \mathbb{P}(D \mid G \neq T)\, \mathbb{P}(G \neq T)$$

$$= (1)\left(\tfrac{1}{3}\right) + 0 \cdot \left(\tfrac{2}{3}\right)$$

$$\mathbb{P}(D) = \tfrac{1}{3}.$$

Conditional expectation

If X and Y are discrete r.v.'s then we can compute conditional probabilities as above:

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

There is also the formula for total probability

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x|Y = y)\,\mathbb{P}(Y = y)$$

where the sum runs over all possible values of Y. Conditioning is a very useful method for solving problems in probability, because it is often much easier to compute conditional probabilities and then sum over the result to find the 'unconditioned' probability.

Total prob.

$$\mathbb{P}(X = x) = \mathbb{P}(X=x \mid Y=y_1)\,\mathbb{P}(Y=y_1) + \mathbb{P}(X=x \mid Y=y_2)\,\mathbb{P}(Y=y_2) + \cdots + \mathbb{P}(X=x \mid Y=y_n)\,\mathbb{P}(Y=y_n)$$

where $\mathrm{Ran}(Y) = \{y_1, y_2, \ldots, y_n\}$.

Note: if X and Y are independent then

$$\mathbb{P}(X = x \mid Y = y_i) = \mathbb{P}(X = x).$$

Example 3 $\boxed{\textit{Best prize:}}$ *n distinct prizes arrive in sequence, all have different values, and one is the best. You must pick a prize or else move on to the next one (no going back to earlier ones). Your knowledge consists of the values of the previous prizes. You want to use a strategy that will maximize the probability of selecting the best prize. The prizes are randomly arranged in sequence.*

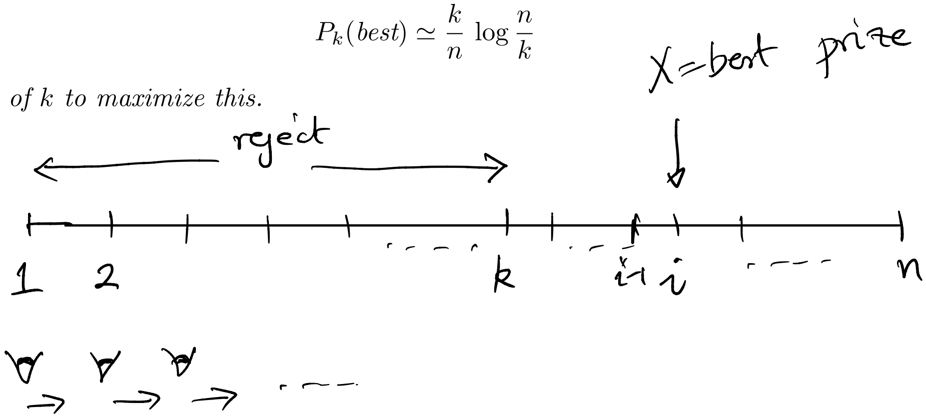
$\boxed{\textit{Strategy:}}$ *reject the first k prizes, then select the first one which is better than all of these previous ones. Let X be the position of the best prize. Use*

$$P_k(best) = \sum_{i=1}^{n} P_k(best|X=i)\, P(X=i)$$

to deduce

$$P_k(best) \simeq \frac{k}{n} \log \frac{n}{k}$$

Find value of k to maximize this.



$P_k(best) = \mathbb{P}($choose best prize using this strategy$)$.

$X =$ position of best prize.

$$\mathbb{P}_k(\text{win} \mid X=i) = \begin{cases} 0 & \text{if } i \le k \\ \mathbb{P}\left(\begin{matrix}\text{best among } \{1,2,\dots,i-1\} \\ \text{located in } \{1,\dots,k\}\end{matrix}\right) & k+1 \le i \le n \end{cases}$$

$$= \begin{cases} 0 & \text{if } i \le k \\ \dfrac{k}{i-1} & k+1 \le i \le n. \end{cases}$$

$$\mathbb{P}(X=i) = \frac{1}{n}$$

$$\Rightarrow \mathbb{P}_k(\text{win}) = \sum_{i=1}^{n} \mathbb{P}_k(\text{win} \mid X=i)\, \mathbb{P}(X=i)$$

$$= \sum_{i=k+1}^{n} \frac{k}{i-1} \cdot \frac{1}{n}$$

$$= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1}$$

$$= \frac{k}{n} \left\{ \frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \cdots + \frac{1}{n-1} \right\}.$$

Scenario where k, n are large:

$$\sum_{i=k+1}^{n} \frac{1}{i-1} \approx \int_{k+1}^{n} \frac{1}{x-1}\, dx$$

$$= \ln\left(\frac{n-1}{k}\right).$$

let $x = \frac{k}{n}$, and $\ln\left(\frac{n-1}{k}\right) \simeq \ln\left(\frac{1}{x}\right)$.

$$\mathbb{P}_k(\text{win}) = x \ln\left(\frac{1}{x}\right)$$

$x = \frac{k}{n} \Rightarrow 0 \leq x \leq 1.$

What value of x maximizes this?

$$f(x) = x \ln\left(\frac{1}{x}\right) = -x \ln x.$$

$$f' = -\ln x - x \cdot \left(\frac{1}{x}\right) = -\ln x - 1$$

Max. $f' = 0 \Rightarrow x = \frac{1}{e}.$

$$x = \frac{k}{n} \Rightarrow \boxed{k \simeq \frac{1}{e} n \simeq \frac{1}{3} n}$$

Best strategy is:

- reject the first third of prizes
- pick the first one that

is better than everything

previous.

We define the conditional expectation of X conditioned on the value $Y = y$ as

$$\mathbb{E}[X|Y = y] = \sum_x x\,\mathbb{P}(X = x|Y = y)$$

This number is defined for each possible value of Y. Putting these all together we get the r.v. $\mathbb{E}[X|Y]$ as a function of Y. You should think of $\mathbb{E}[X|Y]$ as a random variable which is determined by the random variable Y, like Y^2 or e^{tY}: if you know the value of Y, then you know the value of $\mathbb{E}[X|Y]$. There is a very useful relation between the conditional expectation $\mathbb{E}[X|Y]$ and the 'unconditioned' expectation $\mathbb{E}[X]$.

Theorem 1

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X]$$

Note that on the left side we are first averaging over X, with Y fixed, and then we average over Y. On the right side we do it all in just one step.

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X]$$

average over Y

average over X with Y fixed

average over X