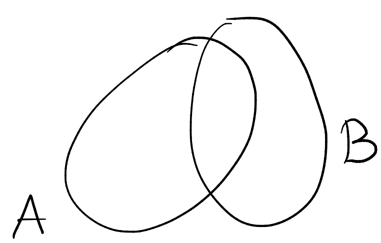
MATH 7241: Problems #1

Due date: Friday September 23

Reading: relevant background material for these problems can be found in the class notes, and in Ross (Sections 2.1 2.2, 2.3, 2.4) and in Grinstead and Snell (Chapters 1,2 3, 6).

Exercise 1 Let A and B be events such that $P(A) = 0.7$ and $P(B) = 0.9$. Find the largest and smallest possible values of $P(A \cup B) - P(A \cap B)$ (note: the event $A \cup B$ means either A or B or both are true, the event $A \cap B$ means both A and B are true).



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

$$\mathbb{P}(A \cup B) - \mathbb{P}(A \cap B)$$

$$= 2\,\mathbb{P}(A \cup B) - \mathbb{P}(A) - \mathbb{P}(B).$$

$\boxed{\textbf{Exercise 2}}$ Each of the following random variables is a well-known type. Identify each by name:
a) an airplane has four engines, and each engine may independently fail (very small probability!). X is the number of engines that fail.
b) flies are randomly landing on my pizza at a steady average rate. X is the number of flies that land on my pizza in the next five minutes.
c) a spammer sends a fake email to a new address every second. X is the number of attempts until somebody responds.
d) a farm raises several hundred thousand chickens. X is the weight of a randomly selected chicken.

$\boxed{\textbf{Exercise 3}}$ A town has five hotels; three people arrive and each randomly and independently selects a hotel. Find the probability that exactly two of them stay in the same hotel.

$$\frac{\#\text{ choices with two in same hotel}}{\#\text{ possible choices}}$$

3 people arrive in order :

$\boxed{\textbf{Exercise 4}}$ Find the mean of X, where the pdf is:

$$P(X = n) = (1-p)^2\, n\, p^{n-1}, \quad n = 1, 2, \ldots$$

[**Hint**: note that $\sum_{n=0}^{\infty} n(n-1)\, p^{n-2} = \frac{d^2}{dp^2} \sum_{n=0}^{\infty} p^n$].

$\mathbb{E}[X]$

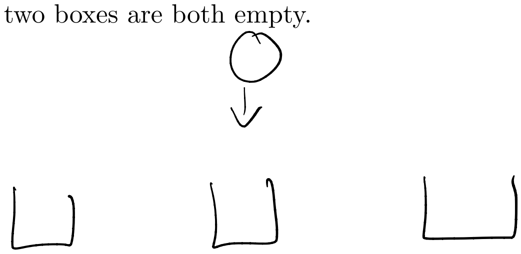
$$\underbrace{\sum_{n=0}^{\infty} p^n}\; = \frac{1}{1-p}$$

$$\sum_{n=1}^{\infty} n^2\, p^{n-1} \qquad \frac{d^2}{dp^2}\left(\quad\right) = \frac{d^2}{dp^2}\left(\frac{1}{1-p}\right).$$

$$\sum n\, p^n \longleftrightarrow \frac{d}{dp}\left(\sum p^n\right)$$

Exercise 5 Randomly distribute r balls in n boxes. Find the probability that the first box is empty. Find the probability that the first two boxes are both empty.

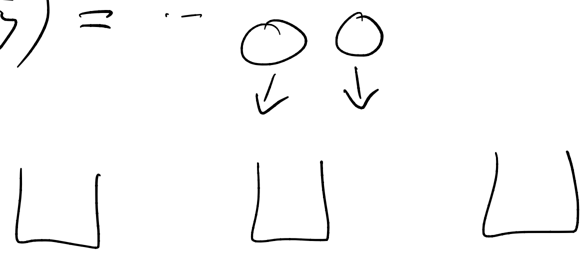
$r = 1, \ n = 3$



#1

$\mathbb{P}(\text{Box \#1 empty}) = \ldots$

$r = 2, \ n = 3$



<u>Counting</u>: all outcomes equally likely

$$\frac{\#\text{ with Box \#1 empty}}{\text{total \# outcomes}}.$$

<u>independent</u>:

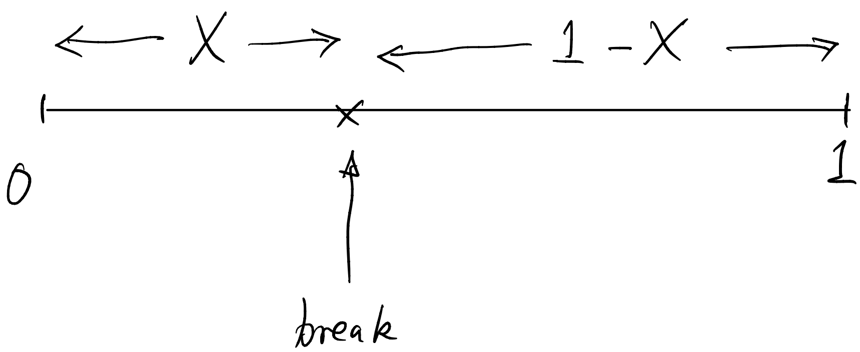
$P(\text{Box 1 empty}) = P(\text{ball 1 } \underline{\text{not}} \text{ in Box 1}$

$\underline{\text{and}} \text{ ball 2 } \underline{\text{not}} \text{ in Box 1})$

$= P(\text{ball 1 not in Box 1}) \cdot$

$\mathbb{P}(\text{ball 2 not in Box 2}).$

$\boxed{\textbf{Exercise 6}}$ We start with a stick of length 1, and break it in two pieces at a randomly chosen position (chosen uniformly over its length). Find the mean length of the longer end of the broken stick.



X = break point $\sim U[0,1]$.

$L = \max(X, 1-X)$.

$E[L] = \mathbb{E}[\max(X, 1-X)]$.

$$\mathbb{E}[g(X)] = \int g(x) f_X(x)\,dx$$

any function

same function

Exercise 7 The current in a resistor is a random variable X. The pdf of X is $f(x) = e^{-(x-1)}$ for $x \geq 1$. The power dissipated in the resistor is $Y = X^2$. Find the pdf of Y.

cdf of Y: $F_Y(y) = \mathbb{P}(Y \leq y)$

$\Rightarrow$ pdf is $f_Y(y) = \frac{d}{dy} F_Y(y)$.

$\boxed{\textbf{Exercise 8}}$ Derive the formula

$$\mathrm{VAR}[X_1 + X_2 + \cdots + X_n] = \sum_{k=1}^{n} \mathrm{VAR}[X_k] + 2\sum_{i<j} \mathrm{COV}(X_i, X_j) \tag{1}$$

$n=2$

$n=3$

$\boxed{\textbf{Exercise 9}}$ Find a random number generator that generates uniformly on $[0, 1]$ (for example the command *rand* in Matlab). Using this generator, estimate the volume of the region under the surface

$$z = \frac{1}{3} \cosh \sqrt{x^2 + y^2}$$

and above the unit square $0 \le x \le 1$, $0 \le y \le 1$. [Note: generate three independent uniform random variables for each run, corresponding to the three coordinates of a random point in the unit cube. Do enough runs to be confident that you have an accurate estimate of the first two decimal places].

(U, V, W) uniform on $[0,1]$.

$\downarrow$

generate sequence of random points.

count number below the surface.

$\boxed{\textbf{Exercise 10}}$ In class we considered this problem: "An urn contains n Red balls and m Black balls. Suppose that k balls are withdrawn from the urn, and let X be the number of Red balls among these. Find $\mathbb{E}[X]$ assuming (i) replacement, and (ii) no replacement." Using the same reasoning as in class, compute $\mathrm{VAR}[X]$ assuming (i) replacement, and (ii) no replacement. [Hint: use the formula from Exercise 8 above. The answers will be different for the two cases].

$\boxed{\textbf{Exercise 11}}$ A typing firm has three typists A,B and C. The number of errors per 100 pages made by typist A is a Poisson random variable with mean 2.6; the number of errors per 100 pages made by typist B is a Poisson random variable with mean 3; the number of errors per 100 pages made by typist C is a Poisson random variable with mean 3.4. A manuscript of 300 pages is sent to the firm. Let X denote the number of errors in the typed manuscript.

a) Assume that one typist is randomly selected to do all the work. Find the mean and variance of X.

b) Assume instead that the work is divided into three equal parts which are given to the three typists. Find the mean and variance of X in this case.

Joint pdf: Mean of an independent product

We noted that the expected value is a *linear operator* on random variables:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[aX+bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

However the expected value of a product is generally *not* the product of expected values, except in one important case: *if* X, Y *are independent* then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \qquad \Leftrightarrow \mathrm{COV}[X,Y] = 0.$$

This can be seen as follows:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{i,j} x_i y_j\, P(X=x_i, Y=y_j) \\
&= \sum_{i,j} x_i y_j\, P(X=x_i) P(Y=y_j) \\
&= \sum_i x_i P(X=x_i) \sum_j y_j P(Y=y_j) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}$$

Similarly for any functions of independent random variables we get

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$$

[Note that if X and Y are independent then $\mathrm{CORR}[X,Y] = 0$. In general the converse is false, but in many cases it is 'mostly true' ...]

Joint pdf: Variance of an independent sum

As a consequence, *if X, Y are independent* then

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$$

To see this, note that

$$\begin{aligned}
\mathrm{VAR}[X+Y] &= \mathbb{E}[X+Y]^2 - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad -(\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2
\end{aligned}$$

This same result applies to a sum of many independent random variables. If $X_1, \ldots, X_n$ are *independent* then

$$\mathrm{VAR}[X_1 + \cdots + X_n] = \mathrm{VAR}[X_1] + \cdots + \mathrm{VAR}[X_n]$$

X, Y independent:

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$$

$$\mathrm{VAR}[X-Y] = \mathrm{VAR}[X + (-Y)]$$

independent

$$\begin{aligned}
&= \mathrm{VAR}[X] + \mathrm{VAR}[-Y] \\
&= \mathrm{VAR}[X] + (-1)^2\mathrm{VAR}[Y] \\
&= \mathrm{VAR}[X] + \mathrm{VAR}[Y].
\end{aligned}$$

c = number.

$$\begin{aligned}
\mathrm{VAR}[cX] &= \mathbb{E}[(cX)^2] - (\mathbb{E}[cX])^2 \\
&= \mathbb{E}[c^2X^2] - (c\,\mathbb{E}[X])^2
\end{aligned}$$

$$= c^2 \mathbb{E}[X^2] - c^2 (\mathbb{E}[X])^2$$

$$= c^2 \left[\mathbb{E}[X^2] - \mathbb{E}[X]^2 \right]$$

$$= c^2 \, VAR[X].$$

Example 18 *Binomial r.v. X can be written as a sum of Bernoulli r.v.'s. If $X \sim Bin(n,p)$ then*

$$X = X_1 + X_2 + \cdots + X_n$$

where each X_i is Bernoulli with probability $P(X_i = 1) = p$. Then it easily follows that

$$\mathbb{E}[X] = np, \quad \text{VAR}[X] = np(1-p)$$

pdf. $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k=0,1,\ldots,n)$

$n=5$ H T T H T $X=2$

$X_1=1 \quad X_2=0 \quad X_3=0 \quad X_4=1 \quad X_5=0$

$$X_k = \begin{cases} 1 & \text{if } k^{th} \text{ toss is Heads} \\ 0 & \text{—— " —— Tails.} \end{cases}$$

$$X = X_1 + X_2 + X_3 + X_4 + X_5.$$

$$\Rightarrow \mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] + \mathbb{E}[X_4] + \mathbb{E}[X_5].$$

X_k	0	1
prob.	$1-p$	p

$\Rightarrow \mathbb{E}[X_k] = p$

$\mathbb{E}[X_k^2] = p$

$\text{VAR}[X_k] =$

$$\mathrm{VAR}[X_k] = \mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2$$

$$= p - p^2$$

$$= p(1-p).$$

$$\Rightarrow \quad \mathbb{E}[X] = 5p$$

$$\mathrm{VAR}[X] = \mathrm{VAR}[X_1 + \cdots + X_5]$$

$$= \mathrm{VAR}[X_1] + \cdots + \mathrm{VAR}[X_5]$$

b/c coin tosses are independent!

$$= 5p(1-p).$$

Example 19 *Coupon collecting problem: there are m types of coupons. Every cereal box has one randomly selected coupon. Let N be the number of boxes collected until we get at least one coupon of every type. Question: what is $\mathbb{E}[N]$?*

As we collect boxes, we encounter new types. Let X_i be the number of boxes after encountering the $(i-1)^{st}$ new type until we encounter the i^{th} new type. Then

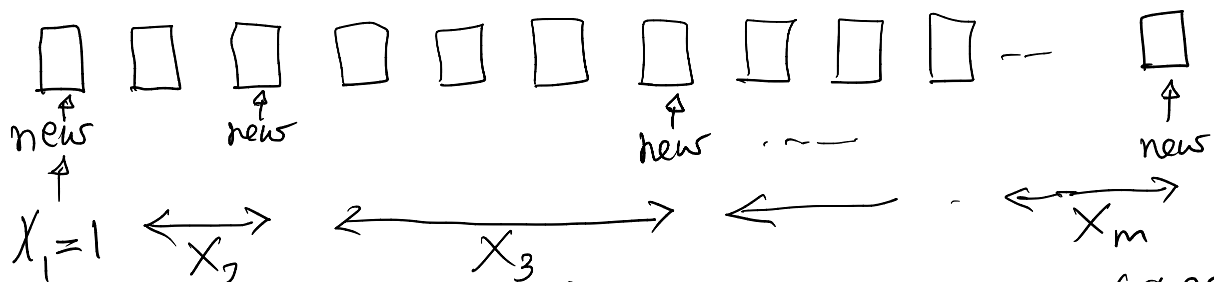
$$N = X_1 + X_2 + \cdots + X_m$$

and each X_i is geometric with probability of success

$$p_i = 1 - \left(\frac{i-1}{m}\right) = \frac{m-i+1}{m}$$

Then

$$\mathbb{E}[N] = \sum_{i=1}^{m} \frac{1}{p_i} \simeq m \log m$$



Collect boxes, noting when we get a new coupon.

X_i = number of boxes after $(i-1)^{st}$ coupon until get a new coupon.

$$N = X_1 + X_2 + X_3 + \cdots + X_m$$

Want $\mathbb{E}[N] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_m]$.

$X_1 = 1 \Rightarrow \mathbb{E}[X_1] = 1$ ✓.

X_2: collect boxes until we find a new coupon

There are m coupons

$\Rightarrow$ we want one of the $m-1$ other ones.

X_2 = number of "trials" until first "success"

← get a new coupon.

$$\mathbb{P}(\text{success}) = \mathbb{P}(\text{get a new coupon})$$

$$= \frac{m-1}{m}$$

X_2 : geometric r.v , $p = \frac{m-1}{m}$

$$\mathbb{E}[X_2] = \frac{1}{p} = \frac{m}{m-1}$$

X_3 : number of trials until you get a new coupon, different from previous two.

$X_3 \sim$ geometric , $p = \frac{m-2}{m}$

$$\mathbb{E}[X_3] = \frac{m}{m-2}$$

Similarly

$$\mathbb{E}[X_4] = \frac{m}{m-3}$$

$$\vdots$$

$$\mathbb{E}[X_m] = \frac{m}{m-(m-1)} = m$$

Summation:

$$\mathbb{E}[N] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_m]$$

$$= 1 + \frac{m}{m-1} + \frac{m}{m-2} + \frac{m}{m-3} + \cdots + m$$

$$= m\left[\frac{1}{m} + \frac{1}{m-1} + \frac{1}{m-2} + \frac{1}{m-3} + \cdots + \frac{1}{1}\right]$$

$$= m \cdot \sum_{k=1}^{m} \frac{1}{k}$$

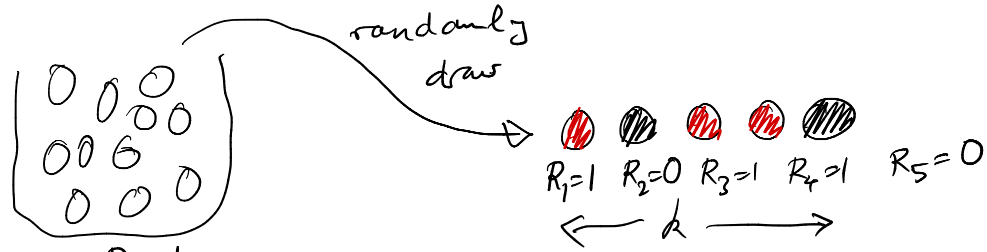
$$\lim_{m\to\infty}\left(\sum_{k=1}^{m}\frac{1}{k}-\int_{1}^{m}\frac{1}{x}\,dx\right)=C$$

Euler's constant

$$\int_{1}^{m}\frac{1}{x}\,dx=\ln x\Big|_{1}^{m}=\ln m$$

so $\boxed{\mathbb{E}[N]\simeq m\ln(m)}$

Example 20 *An urn contains n Red balls and m Black balls. Suppose that k balls are withdrawn from the urn, and let X be the number of Red balls among these. Find $\mathbb{E}[X]$ assuming (i) replacement, and (ii) no replacement.*



n Red
m Black

$X = \#$ red balls drawn.

Want $\mathbb{E}[X]$.

i) replacement $\Rightarrow$ ball goes in after drawing

ii) without replacement $\Rightarrow$ ball stays outside.

The key to solving this problem is to rewrite X as a sum of many small random variables which can each be easily analyzed, and then use the distributive properties of the expected value to put these together for X. For each $i = 1, \ldots, k$ define

$$R_i = \begin{cases} 1 & \textit{if the ith ball is Red} \\ 0 & \textit{if the ith ball is Black} \end{cases}$$

Then it should be clear that

$$X = R_1 + R_2 + \cdots + R_k$$

i) replacement $\Rightarrow$ $R_1, R_2, R_3, \ldots, R_k$ are independent, and they are identical.

So in this case X is binomial.

R_i	0	1
prob	$\frac{m}{n+m}$	$\frac{n}{n+m}$

$n = \#$ Red
$m = \#$ Black

$$\Rightarrow \quad \mathbb{E}[R_i] = \frac{n}{n+m}.$$

$$\Rightarrow \quad \mathbb{E}[X] = \frac{kn}{n+m}.$$

ii) without replacement:

$$X = R_1 + R_2 + R_3 + \cdots + R_k.$$

Not independant, but still have

$$\mathbb{E}[X] = \mathbb{E}[R_1] + \mathbb{E}[R_2] + \cdots + \mathbb{E}[R_k].$$

R_1	0	1
prob.	$\frac{m}{n+m}$	$\frac{n}{n+m}$

n Red

m Black.

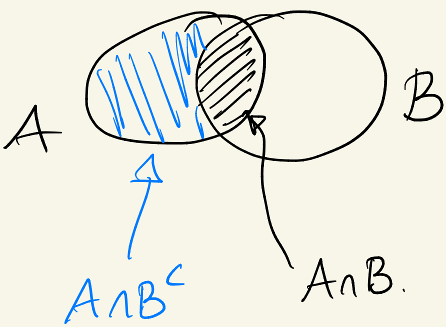
$$\Rightarrow \mathbb{E}[R_1] = \frac{n}{n+m}$$

R_2	0	1
prob		

$$\mathbb{P}(R_2 = 1) = \mathbb{P}(\text{second ball is Red})$$

$$\mathbb{P}(R_2 = 1) = \mathbb{P}(R_2 = 1, R_1 = 1) + \mathbb{P}(R_2 = 1, R_1 = 0)$$

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$$



$$= \mathbb{P}(R_1 = 1, R_2 = 1) + \mathbb{P}(R_1 = 0, R_2 = 1)$$

Note that

$$\mathbb{P}(R_1 = 0, R_2 = 1) = \mathbb{P}(\text{first is Black and second is Red})$$

$$= \mathbb{P}(\text{first is Red and second is Black}).$$

$$= \mathbb{P}(R_1 = 1, R_2 = 0).$$

$$\Rightarrow \mathbb{P}(R_2 = 1) = \mathbb{P}(R_1 = 1, R_2 = 1) + \mathbb{P}(R_1 = 1, R_2 = 0)$$

$$= \mathbb{P}(R_1 = 1)$$

R_2	0	1
prob	$\frac{m}{n+m}$	$\frac{n}{n+m}$

$$\mathbb{E}[R_2] = \frac{n}{n+m}.$$

Similarly,

$$\mathbb{E}[R_i] = \frac{n}{n+m} \quad \text{for all } i = 1, 2, \ldots, k.$$

<u>Conclusion</u>: without replacement

$$\mathbb{E}[X] = \frac{kn}{n+m} \quad \text{(same formula).}$$

Consider R_1: its pdf is

R_1	0	1
Probability	$\frac{m}{n+m}$	$\frac{n}{n+m}$

So we compute

$$\mathbb{E}[R_1] = \frac{n}{n+m}$$

What about $R_2, R_3, \ldots$? Assuming replacement, it is clear that these all have the same pdf's (and they are independent), so we also have

$$\mathbb{E}[R_i] = \frac{n}{n+m} \quad \textit{for all } i = 1, \ldots, k$$

and this implies

$$\mathbb{E}[X] = \frac{k\,n}{n+m}$$

Suppose we do not replace the balls after drawing. Then the variables $R_1, R_2, \ldots$ *are not independent, however they all have the same marginal pdf. To see this, consider* R_2*:*

$$\mathbb{P}(R_2 = 1) \;=\; \mathbb{P}(R_1 = 0, R_2 = 1) + \mathbb{P}(R_1 = 1, R_2 = 1)$$

Now we have

$$\mathbb{P}(R_1 = 0, R_2 = 1) = \mathbb{P}(R_1 = 1, R_2 = 0)$$

(To see this, think about drawing the first two balls at the same time, and suppose one is Red and the other is Black. Which ball was drawn first? It doesn't matter, so both orders are equally likely). Then we have

$$\begin{aligned} \mathbb{P}(R_2 = 1) &= \mathbb{P}(R_1 = 1, R_2 = 0) + \mathbb{P}(R_1 = 1, R_2 = 1) \\ &= \mathbb{P}(R_1 = 1) \end{aligned}$$

This means that R_1 *and* R_2 *have the same pdf, so in particular*

$$\mathbb{E}[R_2] = \mathbb{E}[R_1] = \frac{n}{n+m}$$

The same argument can be extended to show that if $k \leq n + m$ *then*

$$\mathbb{E}[R_i] = \frac{n}{n+m} \quad \textit{for all } i = 1, \ldots, k$$

and so $\mathbb{E}[X]$ *is the same as in the case with replacement. For the homework problem you will analyze the variance of* X*, which does turn out to be different for the two cases!*

MTH 7241 Fall 2022: Prof. C. King

Notes 2: Conditioning

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

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

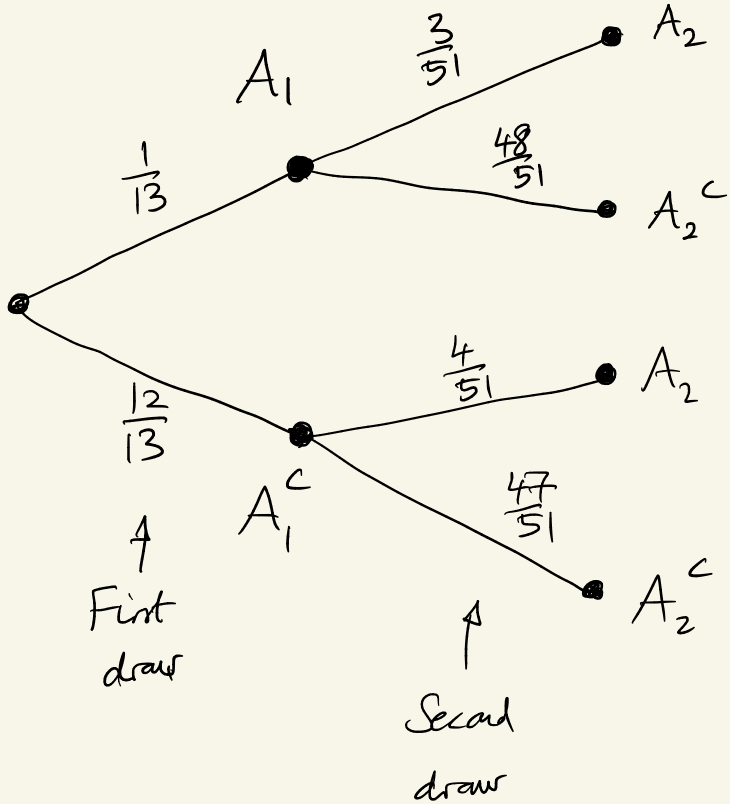
Standard deck of cards: 52 cards
4 Aces.

Randomly draw a card, Prob. to draw Ace $= \frac{4}{52} = \frac{1}{13}$.

Draw 2 cards:

A_1 = first card is Ace

A_2 = second card is Ace



$$\mathbb{P}(A_2 \mid A_1) = \frac{3}{51}$$

$$\mathbb{P}(A_2 \mid A_1^c) = \frac{4}{51} \qquad \text{etc..}$$