Semaine 3 : probabilités

1. Cours

1.1. Probabilité simple

Quand on a un évènement aléatoire, donc par exemple un lancer de dé, on va d'abord lister (quand c'est possible) l'ensemble des issues possibles. Par exemple, pour un dé, on peut avoir {1, 2, 3, 4, 5, 6}, pour une pièce {pile, face}.
Quand on veut calculer la probabilité d'un évènement, on fait $\text{probabilité}(\text{evenement}) = \frac{\text{nombre de cas dans lesquels l'évènement est réalisé}}{\text{nombre d'évènements total}}$

Par exemple pour un lancer de dé, la probabilité d'avoir un nombre pair est $\frac{3 \text{ (Nombre de nombre pairs dans les issues)}}{6 \text{ (Nombre d'issues totales)}} = \frac{1}{2}$

La probabilité d'avoir un 4 est $\frac{1 \text{ (Nombre de 4 dans les issues)}}{6 \text{ (Nombre d'issues totales)}} = \frac{1}{6}$

Evidemment, cela ne marche que lorsque les nombres sont équiprobables ($\leftrightarrow$ qu'ils ont la même chance de tomber).

1.2. Tableaux de probabilité

Maintenant on peut considérer le cas ou les évènements ne sont pas équiprobables.

Donc on a par exemple, pour un dé, le tableau suivant :

Numéros	1	2	3	4	5	6
Probabilité	0.1	0.1	0.1	0.1	0.1	0.5

Si on veut la probabilité d'avoir un nombre pair, on fait $p(\text{pair}) = p(2) + p(4) + p(6) = 0.1 + 0.1 + 0.5 = 0.7$

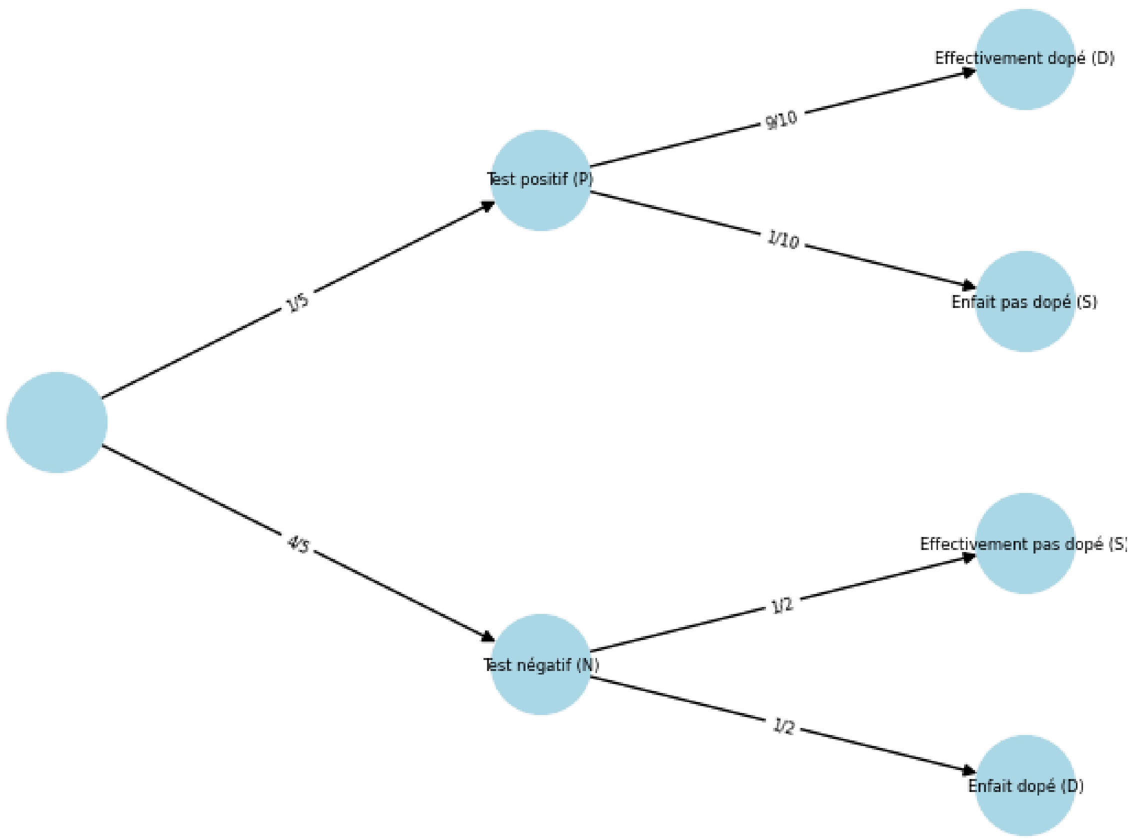
1.3. Arbres de probabilité

Prenons l'exemple suivant : on fait passer un test de dopage à tous les coureurs du peloton du tour de France. 20% d'entre eux on un test positif. Parmis ces coureurs au test positif, 90% le sont effectivement, et parmis les coureurs au test négatif, 50% sont en fait dopés.

Ici, on a deux évènements. Le premier, c'est le test de dopage, on note P pour un test positif, et N pour un test négatif. Ensuite, on a l'évènement pour savoir si le test de dopage est fiable ou pas. On note D pour si le coureur est effectivement dopé, et V si le coureur ne l'est pas.

On peut calculer les probabilités des évènements contraires en faisant $p(\text{event contraire}) = 1 - p(\text{event})$
Donc typiquement la probabilité d'avoir un coureur qui a un test positif est $1 - 0.2 = 0.8$

L'arbre de probabilité ressemble donc à ça :



On part de la base, et on affiche les probabilités comme elles viennent.

Pour expliquer exactement comment on a construit l'arbre, on va avoir besoin d'expliquer les probabilités conditionnelles.

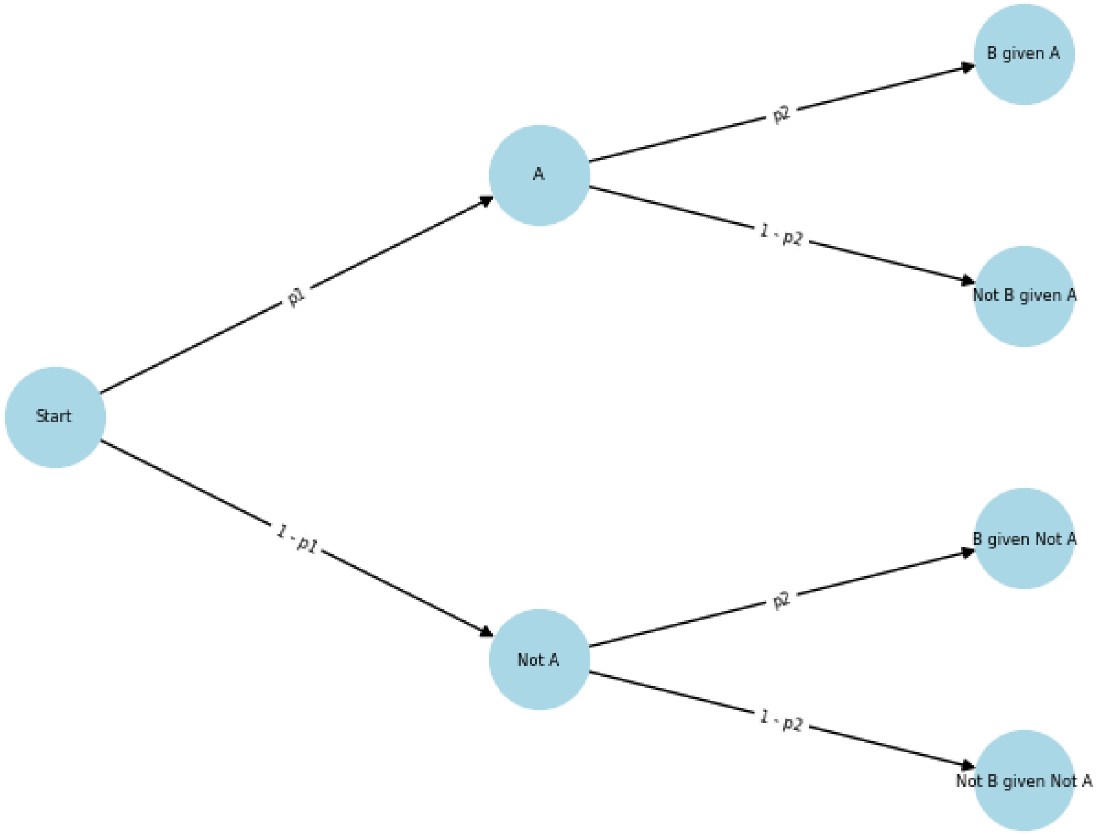
1.4. Probabilité conditionnelle

Une probabilité conditionnelle, c'est lorsque on sait déjà quelque chose, et qu'on veut tirer une probabilité "en sachant que" ce quelque chose est déjà vrai.

Par exemple, si on reprend l'exemple au-dessus, contrairement à ce que l'on peut penser à première vue, il faut faire quelques calculs un peu chiants pour arriver à calculer la probabilité qu'un coureur pris au hasard soit dopé.

Par contre la probabilité qu'il soit dopé sachant que son test est positif, ça c'est donné dans l'énoncé, et c'est 90%. Cela ne veut pas dire que 90% des coureurs sont dopés, loin de là, simplement que 90% de ceux qui sont testés positifs le sont.

Concrètement, dans l'arbre, on met donc les différents évènements qu'on a, avec leurs probabilités, comme dit dans l'image en-dessous



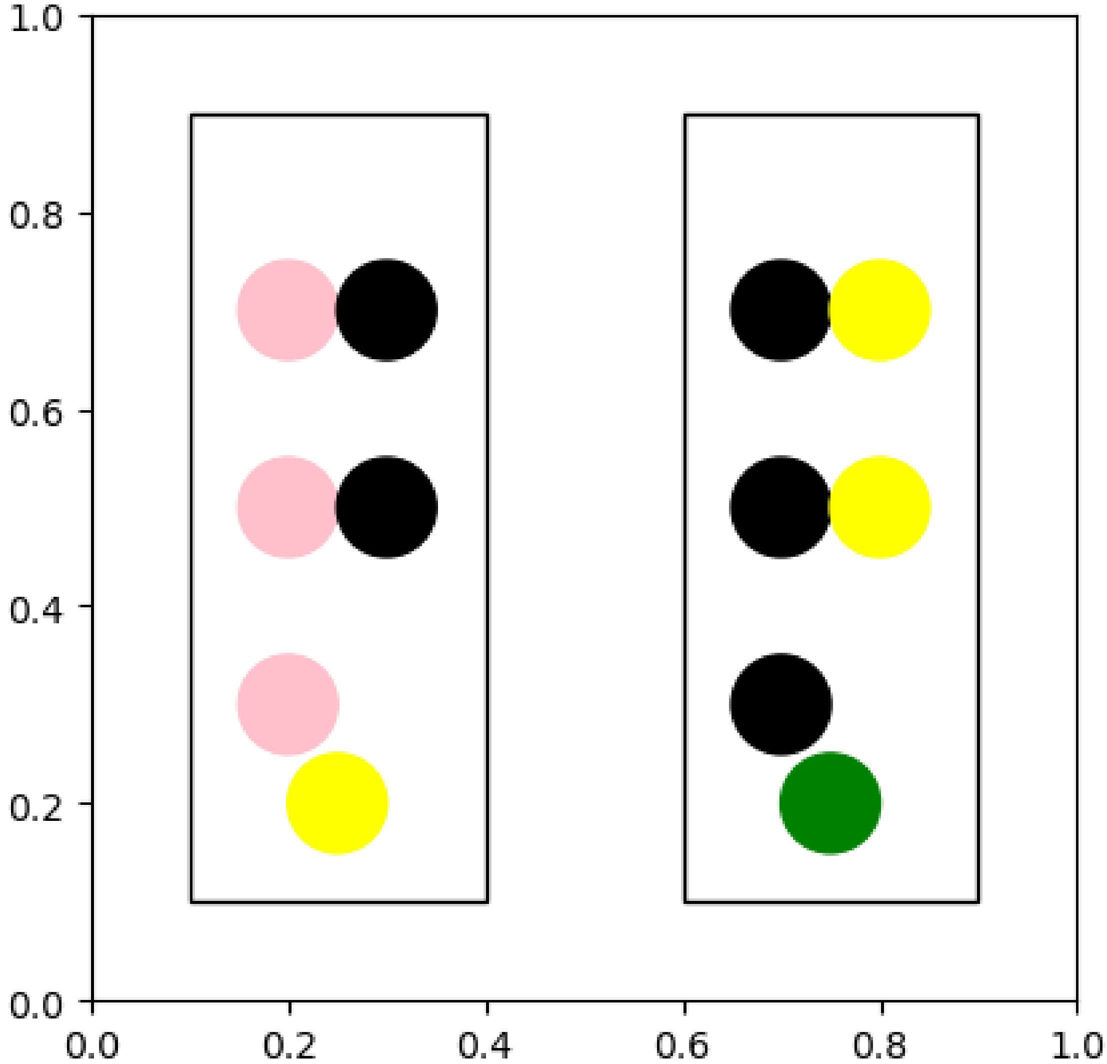
On construit donc ça comme ça.

On a une égalité assez intéressante, qu'il faut globalement apprendre par coeur, et si vous comprenez pourquoi c'est comme ça, tant mieux, sinon tant pis. c'est $p(B|A) = \frac{p(A \cap B)}{p(A)}$. Donc la probabilité de B sachant A ($\leftrightarrow$ La probabilité que B soit réalisé sachant que A est réalisé) est égale à la probabilité que A et B soit réalisés ensemble, sur la probabilité que A soit réalisée.

On l'a aussi sous cette forme, qui peut être plus logique : $p(A \cap B) = p(B|A)p(A)$ Donc la probabilité que A et B soient réalisés ensemble, vaut la probabilité que A soit réalisé fois la probabilité que B se réalise sachant que A est réalisé.

1.5. Exemple concret d'examen, des boîtes et des boules.

On a deux boîtes, une avec 3 billes roses, 2 billes noire et 1 bille jaune. L'autre contient 3 billes noires, 2 billes jaunes, et 1 bille verte. Concrètement, la situation peut donc être modélisée ainsi :



Maintenant, on va essayer de calculer les probabilités. On note U_1 et U_2 les évènements de sélection de la boîte 1 et 2 respectivement.

On note R_1, N_1, J_1, V_1 les évènements que sont “tirer une boule rouge au premier tirage”, “tirer une boule noire au premier tirage”, “tirer une boule jaune au premier tirage”, “tirer une boule verte au premier tirage”

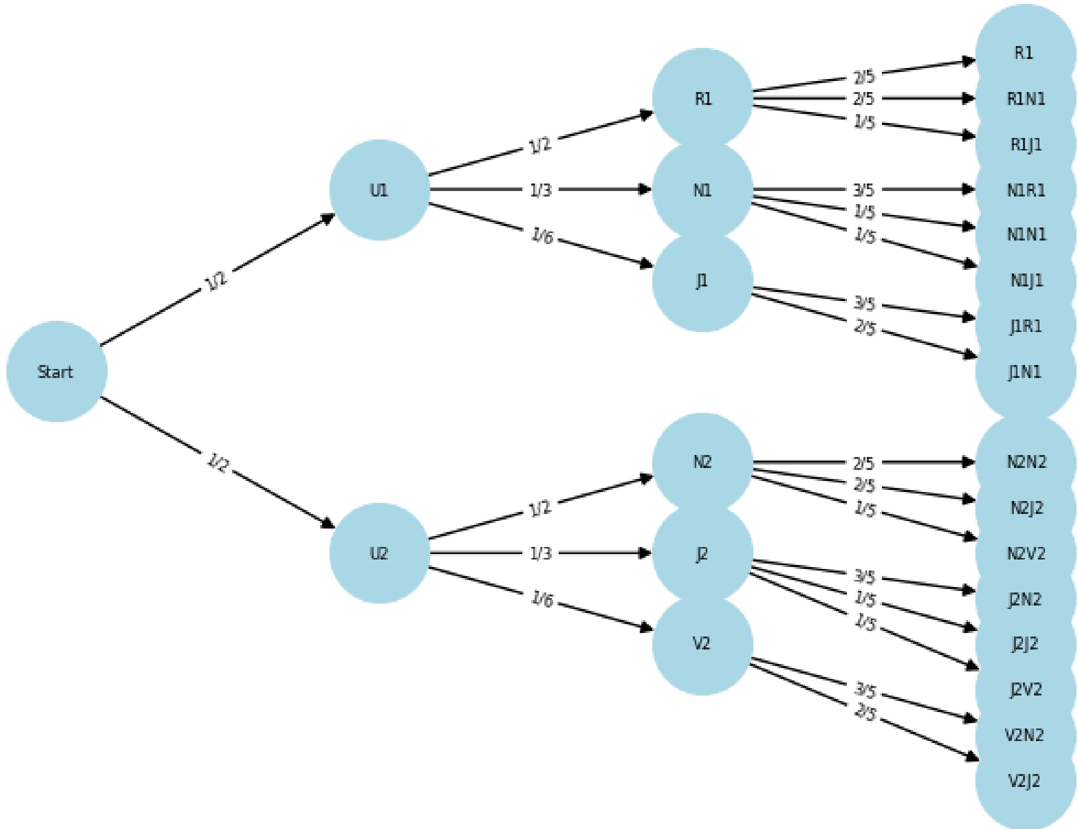
On note R_2, N_2, J_2, V_2 les évènements que sont “tirer une boule rouge au deuxième tirage”, “tirer une boule noire au deuxième tirage”, “tirer une boule jaune au deuxième tirage”, “tirer une boule verte au deuxième tirage”

On va observer les probabilités dans ce contexte : “on choisit une urne au hasard, ensuite, on tire deux boules de celle-ci sans remise, et on observer la couleur de ces deux boules”.

L’arbre de probabilité est en-dessous, mais on va calculer quelques probabilités (je ne vais pas absolument tout faire, je n’ai malheureusement pas le temps, le ML m’appelle à grand pas).

Donc on va faire le premier, concrètement, $p(U_1) = p(U_2) = \frac{1}{2}$. Ensuite, $p(R_1|U_1) = \frac{3}{6} = \frac{1}{2}$, puis $p(R_2|U_1 \cap R_2) = \frac{2}{5}$, puisqu'il ne reste que 5 boules dans l'urne 1 au moment de faire le deuxième tirage, car on en a déjà tirée une (qui dans ce cas précis est une bille rouge, donc il ne reste que 2 billes rouges parmis les 5 restantes)

Donc en gros, on a cet arbre de probabilité (tout est pas parfait dedans, mais flemme de me battre avec matplotlib)



Et à partir de là, ce qui est bien c'est qu'on peut absolument tout faire.

Concrètement, on va avoir plusieurs types de questions, à chaque fois il nous suffit de prendre les "chemins"

Prouvez que la probabilité d'avoir 2 boules jaunes est 1/30

Pour cette question, on voit qu'il n'y a qu'un seule chemin possible vu que dans U_1, il n'y a que une seule boule jaune, il est matériellement impossible de faire en sorte qu'on puisse tirer deux boules.

donc le chemin vaut $p(2 \text{ boules jaunes}) = \frac{1}{2} * \frac{1}{3} * \frac{1}{5} = \frac{1}{30}$ Comme demandé

Quelle est la probabilité d'avoir une boule rouge et une boule noire

on a deux chemins, tous les deux en haut, où on a boule rouge puis boule noire, et boule noire puis boule rouge
$p(2 \text{ boules noires}) = \frac{1}{2} * \frac{1}{2} * \frac{2}{5} + \frac{1}{2} * \frac{1}{3} * \frac{3}{5} = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \frac{1}{5}$

Quelle est la probabilité d’avoir deux boules de même couleur ?

Il faut simplement additionner toutes les probabilités, en faisant tous les chemins (notons qu’on a déjà la probabilité d’avoir 2 boules jaunes, et que la probabilité d’avoir 2 boules vertes vaut 0, puisqu’il n’y a en pas dans l’urne 1, et une seule dans l’urne 2) donc

$p(2 \text{ boules rouges}) + p(2 \text{ boules noires}) + p(2 \text{ boules jaunes}) + p(2 \text{ boules vertes}) = \left(\frac{1}{2} * \frac{1}{2} * \frac{2}{5}\right) + \left(\frac{1}{2} * \frac{1}{3} * \frac{1}{5} + \frac{1}{2} * \frac{1}{2} * \frac{2}{5}\right) + \left(\frac{1}{30}\right) + 0 = \frac{2}{20} + \frac{1}{30} + \frac{2}{20} + \frac{1}{30} + 0 = \frac{4}{20} + \frac{2}{30} = \frac{1}{5} + \frac{2}{30} = \frac{6}{30} + \frac{2}{30} = \frac{8}{30} = \frac{4}{15}$

Quelle est la probabilité d’avoir deux boules de couleur différentes ?

Là pour cette question, il y a deux possibilités. Soit on prend toutes les possibilités (ce qui est un peu long), soit on se rend compte que en fait, l’évènement “avoir deux boules de couleur différentes” et l’évènement “avoir deux boule de même couleur” sont enfait les opposées l’un de l’autre. On peut le voir car ils sont mutuellements exclusifs (on ne peut pas avoir les deux en même temps), et que la réunion des deux vaut 1 (puisque deux boules sont soit de même couleur soit de couleur différentes).

Donc on va faire la méthode 2.

Un évènement A qui est l’opposé d’un autre évènement B a pour probabilité : $p(A) = 1 - p(B)$

Donc ici, on a $p(\text{deux boules couleurs différentes}) = 1 - p(\text{deux boules même couleur}) = 1 - \frac{4}{15} = \frac{11}{15}$

Quelle est la probabilité qu’on ait sélectionné l’urne 1, sachant que les deux boules qu’on a tiré sont noires ?

Cette question est probablement la plus difficile, car on ne peut pas y répondre à l’aide de l’arbre de probabilité.

Pour faire ça, on va évidemment simplifier la chose, mais en gros il faut utiliser un théorème qui s’appelle le Bayes Theorem.

Premièrement, on commence par voir les chemins compatibles : il y en a un en haut, de probabilité 1/30, et un en bas, de probabilité 1/20.
Celui qui nous intéresse est celui du haut. Concrètement, on va reprendre la logique des probabilités en mettant $\frac{\text{cas favorables}}{\text{nombre de cas total}}$

Donc là, on fait $p(U_1|N_1 \cap N_2) = \frac{\frac{1}{30}}{\frac{1}{30}+\frac{1}{20}} = \frac{\frac{1}{30}}{\frac{5}{60}} = \frac{1}{30} * \frac{60}{5} = \frac{2}{5}$

On fait maintenant l’expérience un nombre n de fois, et on veut le n minimum pour lequel la probabilité d’avoir au moins une fois deux boules jaunes est supérieure à 90% :

Cette question est probablement la plus dûre, mais également celle qui rapporte le plus de points de l’exercice de probabilité.

Concrètement, voici les étapes à suivre :
1. Isoler la probabilité mentionnée (sans le “au moins”) avant

2. Appliquer la formule $1-(1-p)^n > \%$ (formule qui marchera dans 90% des cas, à condition de faire attention au $\geq$)
3. Résoudre l'inéquation à l'aide du logarithme

Maintenant, dans notre cas à nous :

1. On mentionne "probabilité d'avoir deux boules jaunes", ça on la connait, c'est $\frac{1}{30}$.
2. Appliquer la formule : $1-\left(1-\frac{1}{30}\right)^n > 0.9$
3. Résoudre l'inéquation à l'aide du logarithme :

$1-\left(1-\frac{1}{30}\right)^n > 0.9$
$1-\left(\frac{29}{30}\right)^n > 0.9$
$0.1 \geq \left(\frac{29}{30}\right)^n$
$\ln(0.1) > \ln\left(\left(\frac{29}{30}\right)^n\right)$ On applique le ln pour casser la puissance
$\ln(0.1) > n\ln\left(\frac{29}{30}\right)$
$\frac{\ln(0.1)}{\ln\left(\frac{29}{30}\right)} < n$ Ici il faut faire attention car $\ln\left(\frac{29}{30}\right)$ est négatif, puisque $\frac{29}{30}$ est plus petit que 1, donc diviser par ce nombre change le signe de l'inégalité
$67.919755 < n$
$n = 68$

1.6. Equations logarithmiques

Lorsque on a une équation avec les logarithmes, qu'ils soient en log ou en ln (voir semaine 2), on essaye de les regrouper pour avoir quelques chose de la forme :
$\log(\text{quelque chose}) = \log(\text{autre chose})$
Et on peut alors simplifier en faisant
$e^{\log(\text{quelque chose})} = e^{\log(\text{autre chose})}$
$\text{quelque chose} = \text{autre chose}$

Donc concrètement, si on a

$1+\log(x+2)+\log(x-1) = \log(23x-17)$
$\log(10)+\log(x+2)+\log(x-1) = \log(23x-17)$ Car $\log(10) = 1$ (par définition)
$\log(10(x+2)(x-1)) = \log(23x-17)$
$e^{\log(10(x+2)(x-1))} = e^{\log(23x-17)}$
$10(x+2)(x-1) = 23x-17$
$10x^2-10x-20 = 23x-17$
$10x^2-33x-3 = 0$

En cherchant les solutions on obtient =
$x_1 = \frac{33-\sqrt{1209}}{20}$
$x_2 = \frac{33+\sqrt{1209}}{20}$

Seule x_2 est positive, donc on ne garde que x_2, sinon on va avoir des ln(x) avec x négatif, ce qui n'est pas possible (c'est comme diviser par 0, c'est inderdit)

2. Exercises

2.1. Probabilités

On prend un élève de l'EPFL au hasard, la probabilité qu'il ait fait des maths avant de venir est de 90%. Si il n'a pas fait de math, la probabilité qu'il échoue est de 50%, Si il en a fait, la probabilité qu'il échoue n'est que de 20%.

Niveau 1

2.1.1. Quelle est la probabilité que l'élève n'ait pas fait de math avant de venir ?

2.1.2. S'il n'a pas fait de math, quelle est la probabilité qu'il réussisse

2.1.3. S'il a fait des math, quelle est la probabilité qu'il réussisse.

Niveau 2

2.1.4. Dessiner l'arbre de probabilité

2.1.5. Quelle est la probabilité que l'élève ait fait des math avant de venir et échoue

2.1.6. Quelle est la probabilité que l'élève n'ait pas fait de math avant de venir et réussisse

Niveau 3

2.1.7. On prend un élève qui a réussi, quelle est la probabilité qu'il ait fait des math avant ?

Niveau 4

2.1.8. On prend plusieurs élèves à la suite, n pour être précis, quel est le plus petit n pour lequel la probabilité d'avoir au moins un élève qui n'a pas fait de math est supérieure à 0.99

2.2. Logarithmes

Résoudre les équations suivantes

2.2.1. $1 + \log(2x + 4) + \log(4x) = \log(3)$

2.2.2. $\log(x^2 + 1) + \log(4) = \log(2)$

3. Solutions

3.1. Probabilités

3.1.1.

$1 - p(\text{math avant de venir}) = 1 - 90\% = 1 - 0.9 = 0.1$

3.1.2.

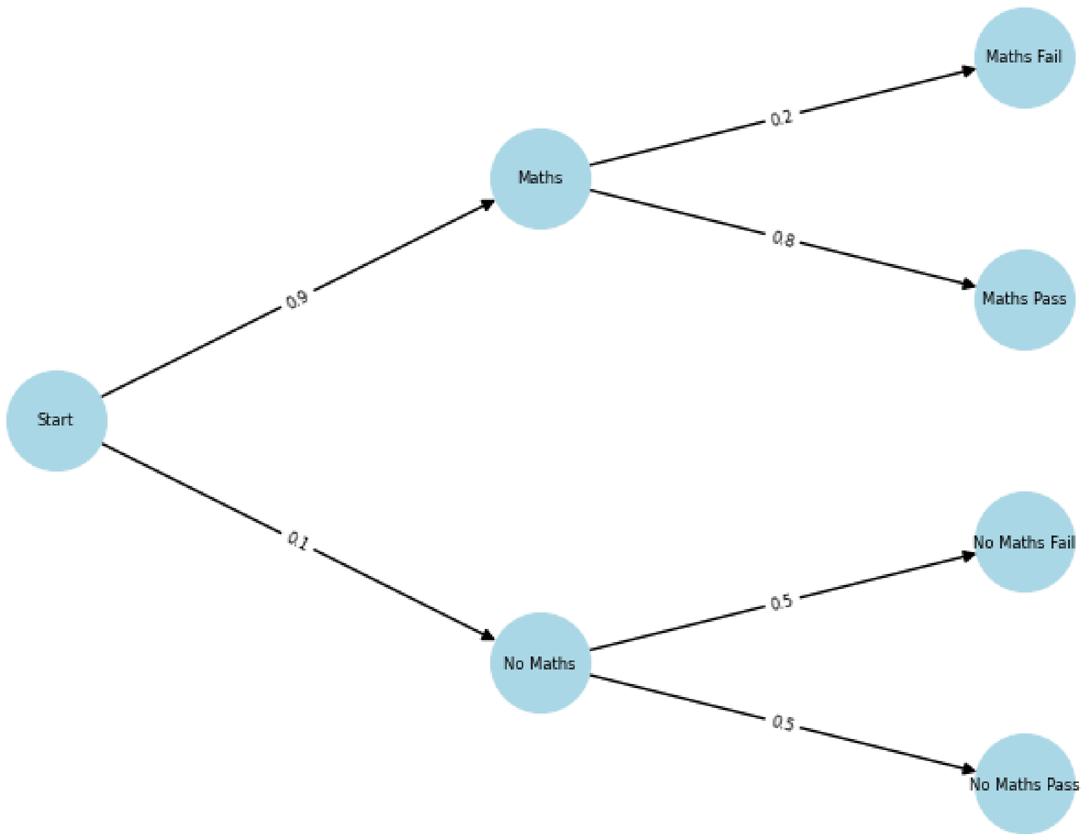
$1 - p(\text{échec}|\text{pas math}) = 1 - 50\% = 1 - 0.5 = 0.5$

3.1.3.

$1 - p(\text{échec}|\text{math}) = 1 - 20\% = 1 - 0.2 = 0.8$

3.1.4.

L'arbre est le suivant



3.1.5.

La probabilité qu'il ait fait des math avant de venir et qu'il échoue est $p(\text{math} \cap \text{échoue}) = p(\text{math}) * p(\text{échoue}|\text{math}) = 0.9 * 0.2 = 0.18$

3.1.6.

$p(\text{pas math} \cap \text{réussisse}) = p(\text{pas math}) * p(\text{réussisse}|\text{pas math}) = 0.1 * 0.5 = 0.05$

3.1.7.

$p(\text{math}|\text{réussi}) = \frac{p(\text{math} \cap \text{réussi})}{p(\text{réussi})}$
$p(\text{math}|\text{réussi}) = \frac{p(\text{math} \cap \text{réussi})}{p(\text{réussi}|\text{math})p(\text{math}) + p(\text{réussi}|\text{pas math})p(\text{pas math})}$
$p(\text{math}|\text{réussi}) = \frac{0.9*0.8}{0.9*0.8+0.1*0.5}$
$p(\text{math}|\text{réussi}) = 0.935$

Pour réussir à l'EPFL, il y a tout interêt à faire des math avant

3.1.8.

La probabilité dont il est question ici est celle de ne pas avoir d'élève qui n'a pas fait dde math.

$1 - (1 - 0.1)^n > 0.99$
$1 - (0.9)^n > 0.99$
$0.01 > (0.9)^n$
$\ln(0.01) > n \ln(0.9)$
$\frac{\ln(0.01)}{\ln(0.9)} < n$
$43 < n$

3.1.9.

On obtient $80x^2 + 160x - 3 = 0$. Donc $x_1 = \frac{-160+8\sqrt{415}}{160}$ (on ne conserve que la solution positive, sinon on a des log négatifs, ce qui est interdit)

3.1.10.

On obtient

$x^2 = \frac{1}{2}$

donc $x = \frac{1}{4}, -\frac{1}{4}$ (car on ne va pas avoir de logarithmes inférieurs à 0)