1 TD2 - Probabilités conditionnelles

Objectif L'objectif de ce TD est de retrouver trois résultats fondamentaux de la théorie des probabilités qui sont :
- la formule des **probabilités totales**,
- la **règle de multiplication**, et
- la **formule de Bayes**.

Pour cela, nous introduirons les notions suivantes : loi de probabilité **jointe**, lois de probabilités **marginales**, loi de probabilité **conditionnelle**, **indépendance** de deux variables aléatoires.

3 Loi de probabilité jointe, lois de probabilités marginales

Dans cette série d'exercices, on dispose d'objets ayant une forme donnée : boule ou cube, et une couleur : rouge ou vert. On définit la variable aléatoire F, comme forme, qui peut prendre deux valeurs : b pour boule et c pour cube et la variable aléatoire C qui peut prendre deux valeurs : r pour rouge et v pour vert.

Les deux variables aléatoires F et C sont appelées variables aléatoires **catégorielles**, les valeurs qu'elles peuvent prendre ne sont pas comparables quantitativement : on ne peut pas dire que la valeur rouge est supérieure à la valeur vert par exemple.

Exercice 3.1: On met dans un sac 30 boules rouges, 10 boules vertes, 20 cubes rouges et 40 cubes verts.

1. Représentez le nombre d'objets de chaque catégorie dans un tableau dont les lignes correspondent aux formes (boule ou cube) et les colonnes correspondent aux couleurs (rouge et vert).
2. On tire du sac un objet, quelle est la probabilité que ce soit une boule rouge ?
3. Calculez de la même manière les 3 autres probabilités et représentez-les dans un tableau comme celui de la question 1.

Étant donné nos deux variables aléatoires F et C, la probabilité que F prenne la valeur b et que C prenne la valeur r est appelé **probabilité jointe**, notée $P(F = b, C = r)$. Les probabilités jointes sont celles se trouvant dans le tableau que vous avez construit à la question 3. Le tableau représente la **loi de probabilité jointe** des variables F et C, notée $P(F, C)$.

4. On tire du sac un objet, quelle est la probabilité que ce soit une boule ?
5. On tire du sac un objet, quelle est la probabilité qu'il soit rouge ?
6. Faites la somme de chaque ligne et de chaque colonne du tableau représentant la loi de probabilité jointe $P(F, C)$ (question 3) et inscrivez ces sommes dans les marges du tableau.
7. Indiquez, dans les marges du tableau, à quelles cellules correspondent les probabilités calculées à la question 4 et 5.

Les lois de probabilités des variables F et C sont appelées **lois de probabilités marginales** de la loi de probabilité jointe. On peut calculer les lois de probabilités marginales $P(F)$ et $P(C)$ à partir de la loi de probabilité jointe $P(F, C)$, comme vous l'avez fait en faisant la somme sur les lignes et sur les colonnes du tableau.

Exercice 3.2: À partir de cet exemple concret, essayons de formaliser la relation entre les lois marginales et la loi jointe.

1. Écrire la relation existant entre $P(F = b)$, $P(F = b, C = r)$ et $P(F = b, C = v)$
2. On considère maintenant que les objets peuvent avoir n couleurs différentes (représentées par c_1 à c_n). Quelle formule permet maintenant de calculer la probabilité de tirer une boule, $P(F = b)$?
3. Alternativement, on peut considérer que les objets peuvent avoir m formes différentes (représentées par f_1 à f_m). Quelle formule permet maintenant de calculer la probabilité de tirer un objet rouge, $P(C = r)$?

La relation que vous venez de trouver, permettant d'obtenir les lois de probabilités marginales à partir de la loi de probabilité jointe, est appelée **formule des probabilités totales**.

4 Variables aléatoires indépendantes

Dans l'exercice précédent, nous avons vu comment construire les lois marginales à partir d'une loi de probabilité jointe grâce à la formule des probabilités totales. Dans cet exercice nous verrons que, *dans certains cas*, il est aussi possible de construire une loi jointe à partir des lois marginales.

Exercice 4.1: On prend comme point de départ les lois marginales de l'exercice 3.1 : $P(F = b) = 0,4$, $P(F = c) = 0,6$, $P(C = r) = 0,5$, et $P(C = v) = 0,5$

1. A partir des lois marginales, $P(F)$ et $P(C)$, on pourrait imaginer que la probabilité jointe $P(F, C)$ de tirer un objet ayant une forme et une couleur correspondrait au produit $P(F) \times P(C)$. Vérifiez cette hypothèse pour $P(F = b, C = r)$.
2. À votre avis, pourquoi est ce que le calcul $P(F = b) \times P(C = r)$ n'aboutit pas au bon résultat $P(F = b, C = r)$? Autrement dit, quelle information présente dans la loi jointe est perdue dans les lois marginales ?
3. Formulez avec vos mots la condition permettant d'obtenir une loi jointe à partir du produit des lois marginales.
4. Trouvez une configuration de boules et de cubes de couleurs verte et rouge pour laquelle il est possible de retrouver la loi de probabilité jointe à partir des lois marginales (en d'autres termes, trouvez le bon nombre de boules rouges, de boules vertes, de cubes rouges et de cubes verts).
5. Dans ce cas-là, écrivez la relation qui lie la probabilité jointe $P(F = b, C = r)$ et les probabilités $P(F = b)$ et $P(C = r)$.
6. Vérifiez que cette relation est maintenue pour les 3 autres paires forme-couleur.

Deux variables X et Y prenant leurs valeurs dans les ensembles S_X et S_Y. X et Y sont **indépendantes** lorsque la condition $P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j)$ est vérifiée pour tout couple $(x_i, y_j) \in S_X \times S_Y$.

5 Probabilités conditionnelles

Il est possible, et c'est souvent le cas, que deux variables aléatoires ne soient pas indépendantes. Dans ce cas là, le fait de connaître la valeur d'une des deux variables aléatoire, nous donne de l'information sur la valeur de la seconde. On se demande alors comment intégrer cela dans le calcul de la probabilité jointe.

Exercice 5.1: On reprend ici l'exemple de la question 1 de l'exercice 3.1, avec le premier tableau comportant le *nombre* de boules et de cubes des différentes couleurs.

1. Utilisez les marges du tableau pour indiquer les sommes du nombre d'objets sur les lignes et sur les colonnes.
2. On tire un objet du sac et on vous dit qu'il est vert. A votre avis quelle est la probabilité que ce soit une boule ?
3. Comparer la valeur obtenue avec la probabilité marginale de tirer une boule $P(F = b)$. Pourquoi sont elles différentes ?
4. Retrouvez la probabilité que vous avez calculé à la question 2 à l'aide du tableau que vous avez fait à la question 1.

On note $P(F = b|C = v)$ la probabilité que la variable aléatoire F prenne la valeur b sachant que la variable aléatoire C vaut v. Cette probabilité est appelée **probabilité conditionnelle**.

Exercice 5.2: La probabilité conditionnelle $P(F = b|C = v)$ représente la proportion d'éléments d'intérêt (ici, les boules : $F = b$) parmi ceux correspondant à une information connue (ici, les objets de couleur verte : $C = v$).[1] Les probabilités conditionnelles peuvent être obtenues à partir des probabilités jointe et marginale, et vice-versa.

1. À partir de vos réponses aux questions précédentes, écrivez la relation existant entre la probabilité conditionnelle $P(F = b|C = v)$, la probabilité jointe $P(F = b, C = v)$, et la probabilité marginale $P(C = v)$.
2. À partir de vos réponses aux questions précédentes, écrivez la relation existant entre la probabilité conditionnelle $P(C = v|F = b)$, la probabilité jointe $P(C = v, F = b)$, et la probabilité marginale $P(F = b)$.

Notez que la probabilité jointe des deux questions ci-dessus est la même, car $P(F = b, C = v) = P(C = v, F = b)$. Cependant, la probabilité conditionnelle est différente, car $P(C = v|F = b) \neq P(F = b|C = v)$.

3. Organisez les probabilités jointes, les probabilités marginales et les probabilités conditionnelles que vous venez de voir sous la forme d'un arbre de probabilité, comme vous avez appris à le faire au lycée.
4. A partir du tableau de la question 1 de l'exercice 5.1, proposez une méthode pour calculer les lois de probabilités conditionnelles $P(F|C)$ et $P(C|F)$.

On note $P(X|Y)$ la **loi de probabilité conditionnelle** de X sachant Y.

1. Au lycée, vous avez probablement noté la la probabilité conditionnelle $P_C(F)$ au lieu de $P(F|C)$. Cette notation est à éviter dans ce cours.

6 Formule de Bayes

Exercice 6.1: On considère maintenant qu'un seul type d'objets, des boules, pouvant être vertes ou rouges. On considère deux boîtes, la boîte 1 qui contient 60 boules rouges et 40 boules vertes et la boîte 2 contenant 30 boules rouges et 70 boules vertes. L'expérience consiste à choisir une boîte et à en extraire un objet.

On considère deux variables aléatoires B et C qui désignent respectivement la boîte choisie et la couleur de l'objet extrait de la boîte choisie. B vaut 1 si l'on choisit la boîte 1 et 2 si on choisit la boîte 2. C vaut r si on extrait une boule rouge et v si on extrait une boule verte. La variable B suit la loi de probabilité suivante : $P(B = 1) = 0,8$ et $P(B = 2) = 0,2$.

1. Calculez la loi de probabilité conditionnelle $P(C|B)$ pour toutes les valeurs de C et de B possibles.
2. Calculez la loi de probabilité $P(C)$ pour toutes les valeurs de C possibles.
3. Calculez la loi de probabilité jointe $P(B,C)$ pour toutes les valeurs de C et de B possibles.

Exercice 6.2: On suppose que l'on ne peut plus reconnaître les deux boîtes : on ne sait plus laquelle est la boîte 1 et laquelle est la boîte 2. On extrait une boule d'une des deux boîtes et on aimerait savoir de quelle boîte il s'agit étant donné la couleur de la boule que l'on a extrait.

1. Quelle est la loi de probabilité que l'on aimerait connaître ? Exprimez-la en fonction des variables aléatoires définies.
2. En utilisant la règle des probabilités conditionnelles (exercice 6.1), écrire de deux manières la loi de probabilité jointe $P(B,C)$.
3. En déduire une formule permettant de calculer la loi de probabilité qui nous intéresse.
4. A l'aide de la formule de la question précédente, calculez la loi de probabilité qui nous intéresse (question 1).

La formule que vous venez de trouver s'appelle **formule de Bayes**. Elle s'appuie sur la définition de probabilité conditionnelle et sur la formule des probabilités totales, et sert à trouver $P(Y|X)$ alors que le problème nous donne la loi "inverse" $P(X|Y)$.