

	<p><b>Année 2017/2018, Session 1 de printemps</b>  L2 Info/MI, UE 4TIN404U, Probabilités,  Statistiques, Combinatoire  Lundi 7 mai 2018 <span style="float: right;">Durée : 1h30</span></p> <p><b>Document autorisé : une page A4 manuscrite</b></p>	<p style="text-align: center;"><b>Collège  Sciences et  Technologies</b></p>
---	--	--

Les exercices sont indépendants. Le barème est donné à titre indicatif. Les calculatrices sont autorisées, mais essentiellement inutiles ; tous les calculs peuvent être laissés sous forme de formules, éventuellement avec des fractions.

### Exercice 1 (3 pts)

- On considère une variable aléatoire  $X$ , pour laquelle on a

$$\mathbb{E}((X + 1)^2) = \mathbb{E}((X - 2)^2).$$

Déterminer la valeur de  $\mathbb{E}(X)$ .

On développe les espérances : il vient  $\mathbb{E}(X^2) + 2\mathbb{E}(X) + 1 = \mathbb{E}(X^2) = 4\mathbb{E}(X) + 4$ , soit en simplifiant,  $6\mathbb{E}(X) = 3$ , qui se résout en  $\mathbb{E}(X) = 1/2$ .

- On suppose de plus que l'on a  $\mathbb{E}((X - 2)^2) = 3$ . En déduire la variance de  $X$ .

De nouveau en développant, on obtient  $\mathbb{E}(X^2) - 4\mathbb{E}(X) + 4 = 3$ , soit en remplaçant  $\mathbb{E}(X)$  par  $1/2$  trouvé à la question précédente,  $\mathbb{E}(X^2) = 1$ . On applique alors l'une des formules pour la variance :  $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 1 - 1/4 = 3/4$ .

### Exercice 2 (5 pts)

On considère deux variables aléatoires **indépendantes**  $X$  et  $Y$ , à valeurs dans  $\mathbb{N}$ , et dont on suppose connues les lois : pour tous entiers  $k \geq 0$  et  $\ell \geq 0$ , on a  $p_k = \mathbb{P}(X = k)$  et  $q_\ell = \mathbb{P}(Y = \ell)$ . On définit la variable aléatoire  $Z$  par  $Z = X + Y$ .

- Dans un premier temps, on suppose  $p_k > 0$  pour tout  $k \geq 0$ , et  $q_k > 0$  pour tout  $k \geq 0$ . Déterminer l'ensemble des valeurs possibles pour  $Z$ .

$X$  et  $Y$  prennent des valeurs entières positives ou nulles, donc  $X + Y$  prend des valeurs aussi entières positives ou nulles. Il reste à vérifier si toute valeur positive ou nulle est possible. Par indépendance, si on se donne un entier  $n \geq 0$ , on a  $\mathbb{P}(X = n) = p_n > 0$  et  $\mathbb{P}(Y = 0) = q_0 > 0$ , donc l'événement  $\{X = n, Y = 0\}$  a, par indépendance, probabilité  $p_n q_0 > 0$ . Par conséquent, tout entier  $n \geq 0$  est bien une valeur possible de  $Z$ .

Donc l'ensemble des valeurs possibles pour  $Z$  est  $\mathbb{N}$ .

2. Toujours avec les mêmes hypothèses, et pour une valeur quelconque  $m$  parmi les valeurs possibles de  $Z$ , écrire l'événement  $\{Z = m\}$  comme une union d'événements disjoints, et dont on peut calculer la probabilité (définis à partir des variables  $X$  et  $Y$  donc).

$Z$  prend la valeur  $m$  si et seulement si  $X$  et  $Y$  prennent des valeurs dont la somme vaut  $m$ . Comme  $X$  et  $Y$  ne prennent que des valeurs entières positives ou nulles, cela signifie qu'on peut avoir  $X = k$  pour un entier  $0 \leq k \leq m$ , et en même temps,  $Z = m - k$  (pour le même entier  $k$ ).

On a donc, en termes d'événements,  $\{Z = m\} = \cup_{k=0}^m \{X = k\} \cap \{Y = m - k\}$ .

3. En déduire la loi de  $Z$  ; c'est-à-dire, pour toute valeur  $m$  possible pour  $Z$ , donner une formule, faisant intervenir les nombres  $p_k$  pour  $k \geq 0$  et les  $q_\ell$  pour  $\ell \geq 0$ , pour la probabilité de l'événement  $\{Z = m\}$ .

La formule obtenue à la question précédente exprime l'événement  $\{Z = m\}$  comme une union d'événements disjoints, chacun étant l'intersection de deux événements indépendants (par indépendance de  $X$  et  $Y$ ). On la traduit donc en une formule pour les probabilités :

$$\mathbb{P}(Z = m) = \sum_{k=0}^m p_k q_{m-k}$$

4. On suppose maintenant que  $X$  et  $Y$  ne peuvent prendre chacune qu'un nombre fini de valeurs (il n'existe qu'un ensemble fini de  $k$  tels que  $p_k > 0$ , et de même un ensemble fini de  $\ell$  tels que  $q_\ell > 0$ ). Écrire une fonction Python `dicoSomme(dicX,dicY)` qui, prenant en entrée un dictionnaire décrivant la loi de  $X$  et un dictionnaire décrivant la loi de  $Y$ , retourne un nouveau dictionnaire décrivant la loi de  $Z$  (au sens de ce qui a été fait en TM : les clés d'un dictionnaire sont les valeurs possibles de la variable aléatoire, et la valeur associée à une clé est la probabilité de cette clé). **Il n'est pas demandé d'optimiser la complexité de votre fonction.**

Le plus simple :

```
def dicoSomme(dicX,dicY):
    dicS = {}
    for k in dicX:
        for l in dicY:
            if (k+l) in dicS:
                dicS[k+l] += dicX[k]*dicY[l]
            else:
                dicS[k+l] = dicX[k]*dicY[l]
    return dicS
```

### Exercice 3 (4 pts)

On suppose que dans une certaine population, le nombre d'enfants par famille suit la loi uniforme sur  $\{0, 1, 2, 3, 4\}$ .

1. On choisit *une famille* au hasard, de manière uniforme; quelle est l'espérance du nombre d'enfants de la famille?

On choisit une famille au hasard, donc on a probabilité  $1/5$  de choisir une famille ayant  $k$  enfants, pour chaque  $k \in \{0, 1, 2, 3, 4\}$ . Le calcul de l'espérance donne  $0 + 1/5 + 2/5 + 3/5 + 4/5 = 2$ .

2. Si l'on suppose que la population totale est composée de 100 familles, combien y a-t-il d'enfants au total? Parmi ceux-ci, combien sont issus de familles ayant chaque nombre possible d'enfants?

La question n'est pas censée être piégée... On a 100 familles au total, et le nombre d'enfants dans une famille prise au hasard dans cette population est uniforme sur  $\{0, 1, 2, 3, 4\}$  : il y a donc 20 familles sans enfant, 20 familles avec un seul enfant, 20 familles avec 2 enfants, 20 familles avec 3 enfants, et 20 familles avec 4 enfants. Soit 20 enfants de familles de 1, 40 enfants de familles de 2, 60 enfants de familles de 3, et 80 enfants de familles de 4. On trouve bien 200 enfants au total (c'est cohérent avec le fait que l'espérance du nombre d'enfants de la question précédente est de 2).

3. On sélectionne *un enfant* au hasard, de manière uniforme, et on note  $X$  le nombre total d'enfants de sa famille (l'enfant interrogé y compris; soit, 1 de plus que son nombre de frères et soeurs). Quelle est l'espérance de  $X$ ?

C'est la question la plus difficile : ici, on change d'expérience, on choisit un enfant uniforme parmi la population de 200 de la question précédente. On a donc probabilité  $20/200 = 1/10$  de tomber sur un enfant unique,  $40/200 = 2/10$  de tomber sur un enfant de famille de deux,  $60/200 = 3/10$  de tomber sur un enfant de famille de trois, et  $40/200 = 4/10$  de tomber sur un enfant de famille de quatre. La variable aléatoire  $X$  prend donc ses valeurs dans  $\{1, 2, 3, 4\}$ , avec probabilités respectives  $1/10$ ,  $2/10$ ,  $3/10$  et  $4/10$ . On calcule alors l'espérance de  $X$  :  $1/10 + 2 \times 2/10 + 3 \times 3/10 + 4 \times 4/10 = 3$ . On n'obtient pas la même chose qu'à la question 1, **parce que l'expérience n'est pas la même !**

4. Sans supposer qu'on est exactement dans la situation décrite dans le reste de l'exercice, on souhaite estimer le nombre moyen d'enfants par famille au sein d'une population. La méthode consistant à interroger une personne prise au hasard, à lui demander combien de frères et soeurs elle a, et à ajouter 1 (pour avoir la taille de sa fratrie), fournit-elle un estimateur sans biais de ce nombre moyen d'enfants par famille?

Interroger une personne prise au hasard revient à utiliser la variable  $X$ , et on a vu sur l'exemple des questions précédentes que l'espérance de  $X$  n'est pas systématiquement égale à ce qu'on cherche à évaluer, à savoir, le nombre moyen d'enfants par famille (et ce n'est pas seulement parce qu'il est impossible de tomber sur un enfant issu d'une famille de 0 enfants). Donc, l'estimateur n'est pas sans biais.

#### Exercice 4 (4 pts)

Un test de dépistage est conçu pour dépister une maladie, mais il est imparfait : il peut donner un résultat "positif" (détection de la maladie) pour un sujet sain, mais également

un résultat “négatif” (absence de la maladie) pour un sujet porteur de la maladie. On a déterminé les valeurs suivantes : appliqué à un porteur sain, le test a 5% de chances de donner un résultat positif (“faux positif”) ; appliqué à un porteur infecté, il a 1% de chances de donner un résultat négatif (“faux négatif”). La proportion de personnes infectées dans la population est estimée à 2%.

1. On applique le test à une personne, qu’on suppose prise uniformément au hasard. En donnant un nom aux événements “la personne testée est porteuse de la maladie”, “la personne testée est saine”, “le résultat du test est positif”, et “le résultat du test est négatif”, interpréter précisément chacune des données chiffrées comme une probabilité ou une probabilité conditionnelle (en précisant de quelle probabilité ou probabilité conditionnelle il s’agit).

On peut définir 4 événements :  $S$  “la personne testée est saine”,  $M$  “la personne testée est porteuse de la maladie”,  $P$  “le résultat du test est positif”, et  $N$  “le résultat du test est négatif”. On a donc naturellement  $M = \bar{S}$  et  $N = \bar{P}$ .

L’énoncé nous fournit comme informations :  $\mathbb{P}(M) = .02$  (d’où on déduit  $\mathbb{P}(S) = 0.98$ ),  $\mathbb{P}(P|S) = 0.05$  (d’où on déduit  $\mathbb{P}(N|S) = 0.95$ ), et  $\mathbb{P}(N|M) = 0.01$  (d’où  $\mathbb{P}(P|M) = 0.99$ ).

2. Le résultat du test est négatif. Quelle est la probabilité que la personne testée soit porteuse de la maladie ?

La question nous demande  $\mathbb{P}(M|N)$ . La formule de Bayes nous donne  $\mathbb{P}(M|N) = \mathbb{P}(N|M)\mathbb{P}(M)/\mathbb{P}(N)$  ; il nous manque  $\mathbb{P}(N)$ .

On calcule donc  $\mathbb{P}(N)$  par la formule des probabilités totales :

$$\mathbb{P}(N) = \mathbb{P}(N|S)\mathbb{P}(S) + \mathbb{P}(N|M)\mathbb{P}(M) = 0.95 \times 0.98 + 0.01 \times 0.02 = 0.9312 = \frac{9312}{10000}$$

Revenant au calcul précédent, on obtient  $\mathbb{P}(M|N) = \frac{1 \times 2}{9312} \simeq 0.000215$ .

3. On dispose d’un second test de dépistage, qui a les caractéristiques suivantes : il a un taux de 10% de faux positifs, et de 0.5% de faux négatifs ; on suppose que si les deux tests sont administrés à une personne donnée, les résultats des deux tests sont indépendants. On n’applique le second test qu’aux personnes pour lesquelles le premier test a donné un résultat positif.

Quelle est la probabilité que la personne testée soit porteuse de la maladie, si les deux tests ont donné un résultat positif ? Quelle est la probabilité que, pour une personne donnée, on administre les deux tests et qu’ils donnent des résultats contraires ?

Ici, l’expérience change : il faut redéfinir les événements qui décrivent les résultats. On peut par exemple poser  $PP$  “les résultats des deux tests sont positifs”, “PN” “le premier test donne un résultat positif, le second un résultat négatif”, et  $N$  “le premier test donne un résultat négatif” ; ces trois événements forment une partition de l’univers puisque un et un seul des trois événements se produira toujours.

On réinterprète la description de l’énoncé :  $\mathbb{P}(PP|M) = 0.99 \times 0.995$ ,  $\mathbb{P}(PP|S) = 0.01 \times 0.005$ ,  $\mathbb{P}(PN|M) = 0.99 \times 0.005$ ,  $\mathbb{P}(PN|S) = 0.02 \times 0.995$ ,  $\mathbb{P}(N|S)$  et  $\mathbb{P}(N|M)$  restent inchangés.

On nous demande  $\mathbb{P}(M|PP)$  et  $\mathbb{P}(PN)$ , qu'on obtient de nouveau par application de la formule de Bayes et de celle des probabilités totales :

$$\begin{aligned}\mathbb{P}(PP) &= \mathbb{P}(PP|S)\mathbb{P}(S) + \mathbb{P}(PP|M)\mathbb{P}(M) = 0.01 \times 0.005 \times 0.98 + 0.99 \times 0.995 \times 0.02 = 0.019 \\ \mathbb{P}(PN) &= \mathbb{P}(PN|S)\mathbb{P}(S) + \mathbb{P}(PN|M)\mathbb{P}(M) = 0.01 \times 0.995 \times 0.98 + 0.99 \times 0.005 \times 0.02 = 0.009 \\ \mathbb{P}(M|PP) &= \frac{\mathbb{P}(PP|M)\mathbb{P}(M)}{\mathbb{P}(PP)} \simeq .9975\end{aligned}$$

(On a donc  $\mathbb{P}(S|PP) \simeq 0.0025$  : si on attend le résultat du second test pour annoncer à une personne que le test est positif, on n'a qu'une chance sur 400 d'annoncer à tort la mauvaise nouvelle ; et la probabilité de  $PN$  est de l'ordre de 1% : ce sont les cas où le résultat des deux tests est encore assez incertain)

### Exercice 5 (4 pts)

Au jeu de bridge, on distribue aléatoirement l'ensemble des cartes d'un jeu de 52 entre les quatre joueurs, de sorte que chacun en détienne 13, qui constituent sa "main". Les cartes ont chacune une valeur (13 valeurs possibles) et une "couleur" parmi Pique, Coeur, Carreau et Trèfle (et chaque combinaison d'une valeur et d'une couleur correspond à une unique carte). Le paquet de cartes est censé être "bien mélangé", de sorte que toutes les répartitions de 13 cartes à chaque joueur soient équiprobables. Le jeu se joue à 4, en deux équipes (fixes) de deux joueurs.

Dans cet exercice, les valeurs à calculer peuvent être laissées sous la forme de formules faisant apparaître des coefficients binomiaux  $\binom{a}{b}$  ; il n'est pas demandé de valeurs numériques.

1. Combien y a-t-il de mains possibles pour un joueur de bridge ? De couples de mains possibles pour un joueur et son partenaire ?

La description d'une main pour un joueur correspond à un sous-ensemble de 13 cartes d'un jeu de 52 : le nombre de mains possibles est donc de  $\binom{52}{13}$ .

Pour les mains de deux partenaires, on peut donner deux formules (équivalentes) : soit on dit qu'une fois déterminée la main du premier joueur, la main du second constitue un sous-ensemble à 13 éléments de l'ensemble des 39 ( $52 - 13$ ) cartes restantes, pour un total de  $\binom{52}{13} \times \binom{39}{13}$  paires de mains ; soit on dit qu'il y a  $\binom{52}{26}$  façons de déterminer l'ensemble des 26 cartes détenues par les partenaires, puis  $\binom{26}{13}$  façons de sélectionner, parmi les 26, les 13 cartes du premier joueur, pour un total de  $\binom{52}{26} \times \binom{26}{13}$ . Les deux formules sont équivalentes : elles se ramènent à  $\frac{52!}{26! \times 13! \times 13!}$ .

2. Combien y a-t-il de façons de sélectionner exactement 4 cartes parmi les 13 cartes de Pique ? Exactement 3 cartes parmi les 13 cartes de Coeur ?

Pour 4 cartes parmi les 13 Piques :  $\binom{13}{4}$ . Pour 3 cartes parmi les 13 Coeurs :  $\binom{13}{3}$ .

3. Donner une formule pour la probabilité qu'un joueur ait une main formée d'exactly 4 Pique, 3 Coeur, 3 Carreau, 3 Trèfle (on parle alors de "répartition 4-3-3-3").

On commence par compter le nombre de mains de ce type "4-3-3-3" : une telle main est composée de 4 Piques parmi 13, mais aussi 3 Coeurs parmi 13, 3 Carreaux parmi 13, et enfin 3 Trèfles parmi 13, pour un nombre total de mains de  $\binom{13}{4} \times \left(\binom{13}{3}\right)^3$ .

Pour donner une probabilité, on divise par le nombre total de mains possibles calculé à la première question ; on obtient comme probabilité,  $\binom{13}{4} \times \left(\binom{13}{3}\right)^3 / \binom{52}{13}$  (numériquement : environ 2.63%)

4. Donner une formule pour la probabilité que votre partenaire ait une répartition “4 – 3 – 3 – 3” (avec 4 cartes à Pique), **sachant que** vous en avez vous-même une (avec 4 cartes à Carreau).

Le mieux est de faire directement un calcul de probabilité conditionnelle : si on se place dans le cas où le premier joueur a bien une main 4-3-3-3 avec 4 Piques, la main du second joueur est tirée dans un jeu contenant 39 cartes, dont 9 Piques, 10 Coeurs, 10 Carreaux et 10 Trèfles. Le nombre de mains possibles est  $\binom{39}{13}$  ; et le nombre de mains 4-3-3-3 avec 4 Carreaux, est  $\binom{9}{3} \times \binom{10}{3} \times \binom{10}{4} \times \binom{10}{3}$ . On obtient donc comme formule,  $\binom{9}{3} \times \binom{10}{3} \times \binom{10}{4} \times \binom{10}{3} / \binom{39}{13} \simeq 0.0313$