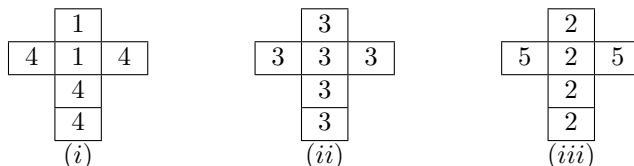


## Semaine 1 - Notions élémentaires de probabilités

### Exercice 1 – Dés de Gardner

Dans un numéro de la revue *Scientific American* de 1974, M. Gardner proposait un jeu consistant à choisir un dé parmi les trois dés à 6 faces non pipés ci-dessous, de manière à essayer d’obtenir le nombre le plus élevé en lançant le dé une seule fois.



**Q 1.1** On vous propose de jouer au jeu à 2 joueurs suivant : chaque joueur mise  $M$  euros. Puis on vous demande de choisir un des dés ci-dessus, votre adversaire en choisit ensuite un autre et enfin chacun lance son dé. Celui qui obtient le nombre le plus élevé remporte la mise.

**Q 1.1.1** Calculez, pour chaque couple  $(x, y)$  de dés la probabilité qu’en jouant avec le dé  $x$  on obtienne un résultat plus élevé qu’avec  $y$ .

**Q 1.1.2** Sachant que la mise est de 30 euros, devez-vous accepter de jouer et, le cas échéant, quel dé devez-vous choisir ? Formellement, quel critère vous permet de statuer ?

### Exercice 2 – La roulette

Dans les casinos, la roulette contient 37 numéros : 18 rouges, 18 noirs et un vert. Quand la roulette tourne, la bille a autant de chances de tomber sur chacun des 37 numéros. Si l’on mise €1 sur le rouge et que ce dernier sort, on gagne €1, sinon on perd la mise de €1.

**Q 2.1** La roulette vous sera t-elle profitable ?

**Q 2.1.1** Soit  $X$  la variable aléatoire représentant le résultat d’une mise de €1. Quelle est la distribution de probabilité de  $X$  ? Quelle est l’espérance de  $X$  ?

**Q 2.1.2** En moyenne combien gagnerez-vous ou perdrez-vous par mise ?

**Q 2.1.3** Combien gagnerez-vous ou perdrez-vous si vous jouez 100 fois en misant €1 à chaque fois ? 1000 fois ? Peut-on en déduire que la roulette n’est pas un jeu profitable ? Justifiez votre réponse.

### Exercice 3 – Paradoxe de Simpson

Le recensement des jugements prononcés dans l’état de Floride entre 1973 et 1978 a permis d’établir le tableau suivant, qui présente les sentences en fonction de la couleur de peau de l’accusé :

meurtrier	peine de mort	autre sentence
noir	59	2547
blanc	72	2185

**Q 3.1** Calculez la probabilité d’obtenir la peine de mort sachant que l’on est noir, puis sachant que l’on est blanc. Qu’en concluez-vous ?

**Q 3.2** En fait le tableau ci-dessus est une synthèse du tableau ci-dessous :

victime	meurtrier	peine de mort	autre sentence
blanche	noir	48	238
	blanc	72	2074
noire	noir	11	2309
	blanc	0	111

Calculez la probabilité d'obtenir la peine de mort conditionnellement à la couleur de peau de l'accusé et de la victime. La justice est-elle clémente envers les noirs dans l'état de Floride? Justifiez votre réponse.

#### Exercice 4 – Sport et age

Dans un échantillon aléatoire de 240 personnes, on a recueilli l'information suivante sur l'âge et sur le type de sport le plus fréquemment pratiqué à Jussieu :

âge \ activité sportive	moins de 20 ans	[20; 25[ ans	[25; 30[ ans	plus de 30 ans
jogging	15	20	15	30
natation	15	10	20	25
ping pong	20	10	30	30

Q 4.1 Quelles sont les deux variables aléatoires étudiées ?

Q 4.2 Estimer la loi jointe de ces deux variables.

Q 4.3 Calculer la probabilité qu'un individu de faire de la natation (dans cet échantillon). Quelle est la probabilité qu'un individu de cet échantillon tiré au hasard ait entre 20 et 25 ans ?

Q 4.4 Calculer la probabilité qu'un individu qui fait du jogging d'avoir plus de 30 ans (dans cet échantillon).

Q 4.5 Ces deux variables aléatoires semblent-elles indépendantes ?

#### Exercice 5 – Notes d'examen

Les étudiants d'un même groupe de TD ont obtenu chacun à une UE une note de partiel (sur 50) et une note globale (sur 100, qui intègre note de CC, de partiel et d'examen) ; les données sont les suivantes :

N°Etudiant	<i>partiel</i>	<i>global</i>
1	45	92
2	23	86
3	50	97
4	46	95
5	33	87
6	21	76
7	13	72
8	30	84
9	34	85
10	50	98

Q 5.1 Calculer le coefficient de corrélation linéaire  $r$  entre  $X$  (note de partiel) et  $Y$  (note globale).

Q 5.2 Représenter graphiquement le nuage de points correspondant aux données. Est-ce que la relation entre les variables semble être linéaire ? Y a-t-il des points aberrants ?

Q 5.3 On a oublié les notes d'un 11<sup>ème</sup> étudiant qui sont : 40 au partiel et 70 de note globale. Si on les incorpore, quel sera l'effet sur  $r$  ?

**Exercice 6 – Estimation de la variance d’une loi de probabilité**

Soit  $X = (X_1, X_2, \dots, X_k, \dots, X_n)$  l’échantillon i.i.d. empirique tiré de  $X_0$ , variable dont l’espérance  $E(X_0) = m$  et la variance  $V(X_0) = \sigma^2$  sont deux paramètres inconnus ; on note  $\theta = (m, \sigma^2)$  le paramètre bi-dimensionnel.

**Q 6.1** Quelle est l’espérance  $E_\theta(\bar{X})$  de la moyenne empirique  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$  ?

**Q 6.2** Quelle est sa variance  $V_\theta(\bar{X})$  ?

**Q 6.3** En déduire que  $\bar{X}$  est un estimateur sans biais et convergent de  $m$ , c’est-à-dire que :

$$E_\theta(\bar{X}) = m \text{ et } \lim_{n \rightarrow \infty} E_\theta[(\bar{X}_n - m)^2] = 0.$$

[*rappel : lorsque  $n$  varie, on écrit  $\bar{X}_n$  au lieu de  $\bar{X}$ .*]

**Q 6.4** Montrer que  $\frac{1}{n} E_\theta[\sum_{k=1}^n (X_k - m)^2] = \sigma^2$ .

**Q 6.5** Pourquoi ne peut-on pas prendre  $\frac{1}{n} \sum_{k=1}^n (X_k - m)^2$  comme estimateur de  $\sigma^2$  ?

**Q 6.6** On considère la statistique  $Y = \sum_{k=1}^n (X_k - \bar{X})^2$ . En utilisant la décomposition  $X_k - \bar{X} = (X_k - m) - (\bar{X} - m)$ , montrer que  $Y = \sum_{k=1}^n (X_k - m)^2 - n(\bar{X} - m)^2$  puis que  $E_\theta(Y) = (n - 1)\sigma^2$ .

**Q 6.7** En déduire que la *variance empirique corrigée*  $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$  est un estimateur sans biais de  $\sigma^2$ .

**Exercice 7 – Modélisation et introduction aux réseaux bayésiens**

Je vis à Paris, j’ai un enfant. J’ai de la fièvre, je ne tousse pas, je suis parti à l’étranger la semaine dernière. Est-ce que j’ai la grippe A ?

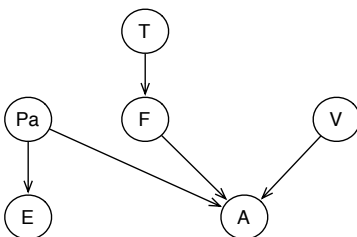
**Q 7.1** Proposer une modélisation reposant sur 6 variables aléatoires bimodales.

**Q 7.2** Traduire la question posée sous forme de probabilité (conditionnelle)

**Q 7.3** Exprimer cette probabilité en fonction de la loi jointe

**Q 7.4** Combien de paramètres doit-on estimer pour exprimer la loi jointe ?

**Q 7.5** Soit le graphe de dépendance entre les variables suivant : exprimer la loi jointe en tenant compte de ces hypothèses. Combien de paramètres reste-t-il à estimer ?



---

**Exercice 8 – Inégalités aux États-Unis**


---

*Le Monde, 5/09/2014 - Les inégalités continuent de se creuser aux États-Unis*

Les inégalités se sont encore accrues aux États-Unis, selon une étude publiée jeudi 4 septembre par la Réserve Fédérale (Fed). Les revenus des 10% les plus riches ont augmenté de 10% entre 2010 et 2013 pour s'inscrire à 397 500 dollars par an (307 000 euros). Dans le même temps, ceux des 40% les moins aisés, ajustés de l'inflation, ont décliné, indique le rapport publié tous les trois ans. Pour les vingt premiers centiles situés au bas de l'échelle, la chute atteint 8% à 15 200 dollars annuels. Si le revenu moyen global a augmenté de 4% au cours des trois dernières années, le revenu médian [...], lui a chuté de 5%. Une tendance qui « correspond à un accroissement de la concentration des revenus durant cette période », indique la Fed.

Ainsi, les 3% les plus riches américains concentrent 30,5% du revenu total en 2013 contre 27,7% en 2010, tandis que la part des 90% les moins riches, elle, a reculé. Par ailleurs, cette catégorie des 3% les plus riches détient 54,4% de la richesse globale (revenu plus patrimoine) contre 44,8% en 1989. A l'autre bout de l'échelle, les 90% les moins riches ont vu leur part tomber à 24,7% contre 33,2% en 1989.

[...]

#### ORIGINES DES MÉNAGES

Lorsqu'on regarde l'origine des ménages, les inégalités sont encore plus criantes. Le revenu moyen de la population blanche, propriétaire et diplômée a augmenté entre 2010 et 2013, tandis que celui des noirs, des hispaniques, des locataires et des sans diplôme a baissé dans le même temps. De la même façon, le revenu médian des noirs et des hispaniques a chuté de 9% sur la période, quand il ne baissait que de 1% pour les blancs.

Par ailleurs, le rapport indique que le taux de propriétaires de leur logement parmi les ménages américains est tombé à 65,2%. Il s'agit du plus bas niveau constaté depuis 1995. Quand aux familles propriétaires de leurs entreprises, le pourcentage est tombé à 11,7%. Du jamais vu depuis 25 ans.

La thèse de l'économiste français Thomas Piketty développée dans son livre *Le capital au XXIe siècle* sur l'accroissement des inégalités, a beau avoir été contestée par une partie de la doxa libérale, les chiffres semblent têtus.

**Q 8.1** Notons,  $R_a$  la variable aléatoire du revenu des salariés américains, indexée par l'année concernée. Nous avons un tirage aléatoire uniforme sur les individus de la population américaine et que nous nous intéressons au revenu de la personne tirée. Nous avons alors :

$$P(R_{2010} > \alpha_{10}^{2010}) = 0.1, \quad P(R_{2013} > \alpha_{10}^{2013}) = 0.1$$

Que valent  $\alpha_{10}^{2010}$  et  $\alpha_{10}^{2013}$  ?

**Q 8.2** Sans tenir compte de l'inflation, donner une traduction probabiliste de la phrase concernant les 4 premiers déciles des distributions de  $R_{2010}$  et  $R_{2013}$ .

**Q 8.3** Même question sur les 2 premiers déciles.

**Q 8.4** En imaginant que nous disposons d'une formule analytique pour  $P(R_a)$ ,  $a \in \{2010, 2013\}$  exprimer l'espérance de  $R_a$ .

**Q 8.5** Donner une modélisation de la phrase suivante : *les 3% les plus riches américains concentrent 30,5% du revenu total en 2013*

**Q 8.6** Introduire de nouvelles variables aléatoires (*Origine*, 3 modalités et *Diplôme*, 2 modalités) et utiliser les probabilités conditionnelles pour modéliser le premier paragraphe de la seconde partie du texte.