SciencesPo

STATISTIQUES APPLIQUÉES AUX SCIENCES HUMAINES & SOCIALES

Niveau introductif

EXERCICES Séances 1 à 4

SÉANCE 1 INTRODUCTION À LA RECHERCHE EMPIRIQUE EN SCIENCES SOCIALES

Exercice 1

Classer les caractères suivants dans le tableau : sexe, surface d'une pièce, nombre de pièces d'un appartement, âge, nombre de salariés d'une entreprise, salaires, nationalité.

Caractère qualitatif	Caractère quantitatif	
	Discret	Continu
Sexe Nationalité	Nombre de pièces d'un appartement Nombre de salariés d'une entreprise Salaires	Surface d'une pièce Age

Exercice 2

Vous avez dû remarquer que dans les magasins de bricolage Cassetout", à la caisse, on demande aux clients leur code postal.

Pour cette situation, indiquer ce que sont les individus, quelle est la variable en précisant son type.

Solution

Les individus sont les **clients**. La variable est le **code postal** : elle est **quantitative, discrète, d'intervalle**.

Exercice 3

Dans les exemples suivants, donner la population, un individu de la population, la variable étudiée et le type de cette variable.

1. On souhaite connaître les catégories socioprofessionnelles de la population active en France en 1930.

2. On veut étudier la répartition des entreprises françaises selon leur taille (petite, moyenne ou grande) en 1973.

3. On mesure le P.N.B. dans chaque pays en 1991.

4. On s'intéresse aux billets en circulation dans le monde et à leur valeur à la Bourse (convertie en Euros) au 1/1/2000.

5. On s'intéresse à l'âge des étudiants inscrits en licence d'histoire.

6. On compte le nombre de personnes par logement dans la ville de Toulouse en 1999.

Solution

	Population	Individu	Variable	Type de variable
1	Population active en France en 1930	Actif	catégories socioprofessionnelles	Qualitative nominale
2	Ensemble des entreprises françaises	Entreprise	taille	Qualitative ordinale
3	Ensemble des pays	Pays	PNB	Quantitative discrète de rapport
4	Ensemble des billets en circulation	Billet	valeur à la Bourse	Quantitative discrète de rapport
5	Étudiants en licence d'Histoire	Étudiant	Age	Quantitative continue de rapport
6	Ensemble des logements	Logement	Nombre de personnes par logement	Quantitative discrète de rapport

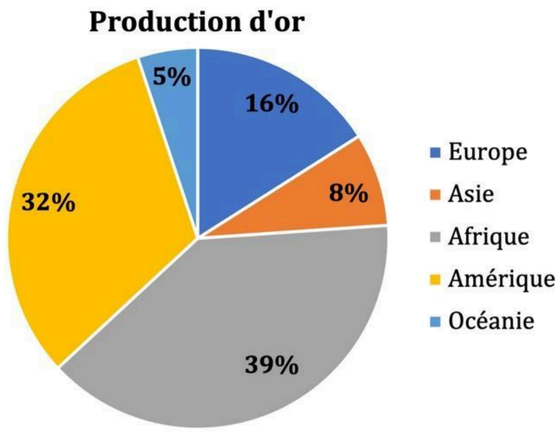
Exercice 4
Voici un tableau d'effectifs donnant la production mondiale d'or en 1937.

Continents	Europe	Asie	Afrique	Amérique	Océanie
Production en tonnes	176	87	431	350	56

Préciser la population, sa taille, un individu de la population, la variable et son type.
Faire le tableau de fréquence puis un diagramme circulaire.

Solution
La population est **l'ensemble des 5 continents**. Les individus sont les **continents**. La variable est la **production d'or en 1937**.
La variable est **quantitative, continue et de rapport**.

Continents	Europe	Asie	Afrique	Amérique	Océanie	Total
Production en tonnes	176	87	431	350	56	1100
Fréquence	**0,16**	**0,08**	**0,39**	**0,32**	**0,05**	**1**



Exercice 5
Le tableau suivant indique le nombre de chômeurs (exprimé en milliers) au sens du BIT (Bureau International du Travail), selon le sexe et l'âge, en France (Mars 1989) :

	Hommes	Femmes
Moins de 25 ans	249	342,6
De 25 à moins de 50 ans	565,1	827,9
50 ans et plus	168,5	155,2

1. *Définir la population étudiée. Quelles sont les variables et leur type ?*
2. *Calculer les fréquences marginales.*
3. a. *Quel est le pourcentage de femmes parmi les chômeurs ?*
b. *Quel est le pourcentage de jeunes de moins de 25 ans parmi les chômeurs ?*
c. *Quel est le pourcentage de femmes parmi les chômeurs de moins de 25 ans ?*

Solution
1. La population étudiée est **l'ensemble des chômeurs en France en mars 1989**. Le **sexe** et **l'âge** sont les variables. La première est **qualitative et nominale**, la deuxième est **quantitative, continue et de rapport**.
2. Je calcule les effectifs marginaux puis les fréquences marginales.

	Hommes	Femmes	Effectif marginaux
Moins de 25 ans	249	342,6	**591,6**
De 25 à moins de 50 ans	565,1	827,9	**1393**
50 ans et plus	168,5	155,2	**323,7**
Effectifs marginaux	**982,6**	**1325,7**	**2308,3**

	Hommes	Femmes	Fréquences marginales
Moins de 25 ans	**11%**	**15%**	**26%**
De 25 à moins de 50 ans	**25%**	**36%**	**60%**
50 ans et plus	**7%**	**7%**	**14%**
Fréquences marginales	**43%**	**57%**	**100%**

3. a. **57%** des chômeurs sont des femmes.
b. 26% des chômeurs sont des jeunes de moins de 25 ans.

c. **58%** des chômeurs de moins de 25 ans sont des femmes.

$$\frac{15}{26} \approx 0,58 \approx \frac{58}{100}$$

Exercice 6 L'effet de structure (Shift and Share)
On a relevé les salaires des cadres et des ouvriers dans deux entreprises A et B.

	Entreprise A	Entreprise B
Cadre	1600€	1800€
Ouvrier	1000€	1100€

On a parallèlement les structures suivantes.

	Entreprise A	Entreprise B
Cadre	100	50
Ouvrier	400	800

1. *Calculer le salaire moyen dans l'entreprise A, puis le salaire moyen dans l'entreprise B.*
2. *Dans quelle entreprise préféreriez-vous travailler ?*
Solution
1. *Le salaire moyen dans l'entreprise A est de :*

$$m_A = \frac{1600\times100+1000\times400}{100+400} = \frac{560000}{500} = 1120€$$

Le salaire moyen dans l'entreprise B est de :

$$m_B = \frac{1800\times50+1100\times800}{800+50} = \frac{970000}{850} \approx 1141€$$

2. Le salaire moyen et le salaire maximal est certes plus élevé dans l'entreprise A mais le nombre de cadres est plus important en proportion dans l'entreprise A.

Exercice 7
Considérons deux régions **R1** et **R2** et trois secteurs économiques : l'agriculture (**S1**), l'industrie et le BTP (**S2**) et les services et commerces (**S3**). La **productivité du travail** PT est mesurée par le rapport entre la valeur ajoutée VA et le nombre d'emplois E

$$PT = \frac{VA}{E}$$

Dans un rapport, on peut lire que la productivité du travail dans la région **R1** est strictement supérieure à la productivité du travail dans la région **R2**.
*Peut-on en conclure que la région **R1** a une meilleure productivité que la région **R2** ?*

Les données sont les suivantes (VA en millions d'€).

Secteur / Région	S1		S2		S3	
	E	VA	E	VA	E	VA
R1	500	40	6 000	1 300	500	45
R2	1 500	170	1 000	380	1 000	120

Solution
La productivité moyenne dans la région 1 est de :

$$PT_1 = \frac{40\times\frac{40}{500}+1300\times\frac{1300}{6000}+45\times\frac{45}{500}}{40+1300+45} \approx 0,209$$

La productivité moyenne dans la région 2 est de :

$$PT_2 = \frac{170\times\frac{170}{1500}+380\times\frac{380}{1000}+120\times\frac{120}{1000}}{170+380+120} \approx 0,266$$

Si l'on tient compte de la valeur ajoutée, la région 2 a une meilleure productivité que la région 1.

Exercice 8 Sensibilité aux valeurs extrêmes

Imaginons deux entreprises A et B dont les salaires mensuels en € sont les suivants.

	Salaires	Effectifs	Salaires	Effectifs	Salaires	Effectifs
Entreprise A	1 000	100	3 000	10	50 000	1
Entreprise B	1 100	100	3 500	10	5 000	1

1. *Déterminer les salaires moyens dans les deux entreprises A et B.*
2. *Quels commentaires pouvez-vous faire concernant cette répartition des salaires ?*
Solution
1. *Je détermine le salaire moyen dans l'entreprise A.*

$$\overline{x}_A = \frac{100\times1\,000+10\times3\,000+1\times50\,000}{100+10+1} = \frac{180\,000}{111} \approx 1621,62€$$

Je détermine le salaire moyen dans l'entreprise B.

$$\overline{x}_B = \frac{100\times1\,100+10\times3\,500+1\times5\,000}{100+10+1} = \frac{150\,000}{111} \approx 1351,35€$$

2. Le salaire moyen dans l'entreprise A est bien plus important mais c'est lié au seul gros salaire de l'entreprise. Les employés de l'entreprise B gagne plus en général.

Exercice 9
On cherche le nombre d'habitants pour un médecin dans une région comportant 4 départements. Dans ces quatre départements, on a trouvé 1 médecin pour 100, pour 200, pour 250, pour 500 habitants. *Combien y a-t-il d'habitants pour un médecin dans cette région ?*
Solution
Il nous **manque la répartition de la population dans chaque département** pour pouvoir répondre. Si leur population est identique, la fréquence moyenne des médecins est de :

$$\frac{1}{4}\times\left(\frac{1}{100}+\frac{1}{200}+\frac{1}{250}+\frac{1}{500}\right)=\frac{1}{4}\times\left(\frac{10}{1\,000}+\frac{5}{1\,000}+\frac{4}{1\,000}+\frac{2}{1\,000}\right)=\frac{1}{4}\times\frac{21}{1\,000}=\frac{21}{4\,000}$$

Cela correspond à 21 médecins pour 4 000 habitants, soit à peu près 1 médecin pour 200 habitants.

Exercice 10
Un salaire augmente de 10% puis l'année suivante de 20% puis la dernière année de 30%. *Quelle est la hausse annuelle moyenne ?*
Solution
Soit x le coefficient lié à l'augmentation annuelle.

$$x^3 = 1,1\times1,2\times1,3x^3 = 1,716x = \sqrt[3]{1,716}x\approx1,197$$

Cela correspond donc à une hausse annuelle d'environ **19,7%**.

Exercice 11
Dans une classe, il y a 22 garçons qui ont obtenu 12 de moyenne à un devoir, les filles ont obtenu 10 de moyenne et la moyenne de la classe est de 11,1. *Combien y a-t-il de filles dans la classe ?*
Solution
Soit x le nombre de filles dans la classe. Le problème se traduit par l'équation suivante :

$$22\times12 + 10x = 11,1(x + 22)264 + 10x = 11,1x + 244,21,1x = 19,8x = 18$$

Il y a **18 filles** dans la classe.

Exercice 12 (Examen de rattrapage, Nancy 2011)
Un apiculteur amateur fait le bilan de la production de miel de ses ruches pour l'année 2010. Pour chacune d'elles, il note la quantité de miel produite (en kg). Il obtient les résultats suivants.

Production de miel (en kg)	18	20	21	22	23	24	26	28
Nombre de ruches	2	4	3	3	1	3	1	3

Quelle est sa production moyenne de miel ?
Solution

La production moyenne de miel est de :

$$\bar{x} = \frac{18\times2+\ldots+28\times3}{2+\ldots+3} = \frac{450}{20} = 22,5\ kg$$

Exercice 13

Déterminer les médianes, les quartiles 1 et 3 ainsi que l'écart interquartile des séries suivantes

Série 1 : 2/2/2/8/8/8/9/10/11/12/14/15/16/16/17/18/19/19/20

Série 2 : 4/5/5/6/6/7/8/9/10/10/10/12/12/13/13/15/18/19/19/20

Série 3 : 4/5/5/6/6/7/8/9/10/10/10/12/12/13/13/15/18/19/19/20/25/26/25/28

Série 4 : 4/5/5/6/6/7/8/9/10/10/10/12/12/13/13/15/18/19/19/20/21

Solution

Série 1

L'effectif total est 19. $19:2 = 9,5$. La médiane est la 10ème valeur, soit **12**.

$19:4 = 4,75$. Le premier quartile est donc la 5ème valeur, soit **8**.

$19\times\frac{3}{4} = 14,25$. Le troisième quartile est donc la 15ème valeur, soit **17**.

L'écart interquartile est de $17 - 8 = 9$.

Série 2

L'effectif total est 20. $20:2 = 10$. La médiane est comprise entre la 10ème valeur et la 11ème valeur, soit **10**.

$20:4 = 5$. Le premier quartile est donc compris entre la 5ème et la 6ème valeur, soit **6,5**.

$20\times\frac{3}{4} = 15$. Le troisième quartile est donc compris entre la 15ème et la 16ème valeur, soit **14**.

L'écart interquartile est de $14 - 6,5 = 7,5$.

Exercice 14

On a relevé les pourcentages de réussite obtenus par 49 élèves à un examen, rangés par classes de 10 %:

Classe	**[0 ; 10[**	**[10 ; 20[**	**[20 ; 30[**	**[30 ; 40[**	**[40 ; 50[**	**[50 ; 60[**	**[60 ; 70[**	**[70 ; 80[**	**[80 ; 90[**
Effectif	2	4	5	8	7	9	6	6	2
Effectif cumulé	2	6	11	19	26	35	41	47	49

1. *Comment trouver la médiane ?*

2. *À quelle classe la médiane appartient-elle ?*

3. *Si on sait que le contenu de cette classe est 46 – 42 – 45 – 44 – 40 – 43 – 49, quelle est la valeur de la médiane ?*

4. *Si on ne connaît pas le contenu de cette classe, par approximation affine, quelle est la valeur la médiane ?*

Solution

1. $49:2 = 24,5$. La médiane est donc la **25ème valeur** de la série.

2. La médiane appartient à la classe **[40 ; 50[**.

3. D'après le tableau, 40 est la 20ème valeur, 49 la 26ème valeur donc **46 est la médiane**.

4. Si on ne connaît pas le contenu de cette classe, par approximation affine, la k-ième valeur de la classe [40 ; 50[est donnée par la formule $40 + \frac{10(k-1)}{7}$. La médiane est la 6ème valeur de la classe, soit $40 + \frac{50}{7} = 47 + \frac{1}{7}$

Exercice 15 Déterminer une médiane et des quartiles (cas discret)

Un apiculteur amateur fait le bilan de la production de miel de ses ruches pour l'année 2010. Pour chacune d'elles, il note la quantité de miel produite (en kg). Il obtient les résultats suivants :

Production de miel (en kg)	**18**	**20**	**21**	**22**	**23**	**24**	**26**	**28**
Nombre de ruches	2	4	3	3	1	3	1	3

1. *Déterminer la médiane et les quartiles de cette série en expliquant votre méthode.*

2. L'apiculteur a retrouvé les résultats qu'il avait établis pour l'année 2009 :

Minimum	**Q1**	**Médiane**	**Q3**	**Maximum**
10	15	20	25	30

a. *En 2009, à quel pourcentage peut-on estimer la part du nombre de ruches ayant produit plus de 25 kg de miel ?*

b. *En 2009, à quel pourcentage peut-on estimer la part du nombre de ruches ayant produit moins de 20 kg de miel ?*

c. *À l'aide des résultats obtenus dans la question 1 et du tableau précédent, comparer les productions des deux années.*

3. L'apiculteur cherche à estimer la production en 2011. On fait l'hypothèse que, par rapport à l'année 2010, la production de chacune des ruches augmente de 3 kg. *Pour les deux questions ci-dessous, déterminer sans justification la bonne réponse parmi A, B et C. en reportant le numéro de la question accompagné de la lettre correspondant à votre réponse sur votre copie.*

Question	Réponse A	Réponse B	Réponse C
La moyenne de la série…	ne change pas	augmente de $\frac{3}{21}$ kg	augmente de 3 kg
La médiane de la série…	ne change pas	augmente de $\frac{3}{21}$ kg	augmente de 3 kg

Solution

1. Je détermine la médiane et les quartiles de cette série en calculant les effectifs cumulés croissants.

Production de miel (en kg)	18	20	21	22	23	24	26	28
Nombre de ruches	2	4	3	3	1	3	1	3
Effectifs cumulés croissants	2	6	9	12	13	16	17	20

$\frac{20}{2} = 10$ *donc la médiane est comprise entre la* $10^{ème}$ *et la* $11^{ème}$ *valeur, soit* $M_e = 22$

$\frac{20}{4} = 5$ *donc le* 1^{er} *quartile est compris entre la* $5^{ème}$ *et la* $6^{ème}$ *valeur, soit* $Q_1 = 20$

$20\times\frac{3}{4} = 15$ *donc le* $3^{ème}$ *quartile est compris entre la* $15^{ème}$ *et la* $16^{ème}$ *valeur, soit* $Q_3 = 24$

2. L'apiculteur a retrouvé les résultats qu'il avait établis pour l'année 2009 :

Minimum	Q1	Médiane	Q3	Maximum
10	15	20	25	30

a. En 2009, **25%** des ruches ont produit plus de 25 kg de miel ($Q_3 = 25$).

b. En 2009, **50%** des ruches ont produit moins de 20 kg de miel ($M_e = 20$).

c. En 2009 et en 2010, la médiane et le $3^{ème}$ quartile sont quasiment identiques. Par contre, le 1^{er} quartile est bien inférieur en 2009. On en déduit que la récolte a été **moins bonne en 2009**.

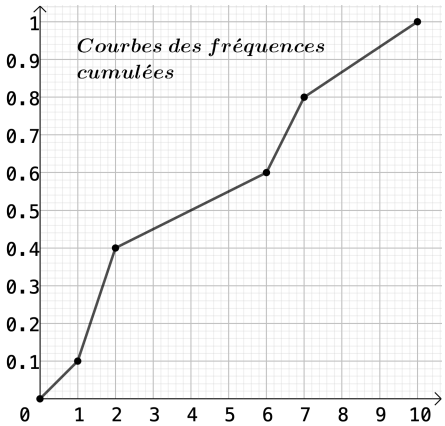
3. Moyenne et médiane augmentent autant que chacune des valeurs.

Question	Réponse A	Réponse B	Réponse C
La moyenne de la série…	ne change pas	augmente de $\frac{3}{21}$ kg	**augmente de 3 kg**
La médiane de la série…	ne change pas	augmente de $\frac{3}{21}$ kg	**augmente de 3 kg**

Exercice 16

On considère la courbe des fréquences cumulées croissantes suivante correspondant à la distribution des notes sur 10 obtenues par des étudiants.

Avec la précision permise par le graphique, déterminer Q_1, Q_3, l'écart interquartile, la médiane, le pourcentage d'étudiants ayant obtenus entre 4 et 7.



Solution

$Q_1 = 1,5$ $M_e = 4$ $Q_3 = 6,8$ $EI = 6,8 - 1,5 = 5,3$

On pourra noter que le troisième quartile a été obtenu par interpolation affine (toute valeur entre 6 et 7 convenait). Les quartiles et la médiane correspondent aux abscisses des points d'ordonnées 0.25 , 0.5 et 0.75 (25%, 50% et 75%)

30% des étudiants ont obtenu entre 4 et 7.

$$0,8 - 0,5 = 0,3 = \frac{30}{100}$$

Exercice 17

Des élèves ont participé à une dictée. Voici le nombre de fautes relevées pour chacun :

8 3 5 1 4 2 8 7 9 4 3 1 0 2 8 2 1 8 4 0 2 8 5 1 0 5 7 4 3 9 6 8 4 9 2 8 8 7 0 3 6 8 4 9 2 8 7 7 0 3

1. *Déterminer la variable et son type. Quelle est la taille de l'échantillon ?*
2. *Déterminer le mode.*
3. *Déterminer l'étendue.*
4. *Déterminer la médiane et les quartiles de cette série statistique ainsi que l'intervalle interquartile.*
5. *Déterminer la moyenne et l'écart-type.*
6. *À partir de ces deux derniers indices que peut-on dire de ce groupe d'élèves ? Est-il homogène ?*

Solution

1. La variable est le **nombre de fautes dans la dictée** : c'est une variable quantitative discrète et de rapport. La taille de l'échantillon est de **50**.

Reprenons les données dans un tableau.

Nombre de fautes	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Effectif	5	4	6	5	6	3	2	5	10	4	50
ECC	5	9	15	20	26	29	31	36	46	50	
Produits moyens	0	4	12	15	24	15	12	35	80	36	233

2. La valeur de plus grand effectif est 8 donc le mode est **8**.
3. L'étendue égale $9 - 0 = 9$.
4. $50 : 2 = 25$. La médiane est donc comprise entre la $25^{\text{ème}}$ et la $26^{\text{ème}}$ valeur, soit **4**.
$50 : 4 = 12,5$. Le 1^{er} quartile de cette série statistique est donc la $13^{\text{ème}}$ valeur, soit **2**.
$\frac{3}{4}\times 50 = 37,5$. Le $3^{\text{ème}}$ quartile de cette série statistique est donc la $38^{\text{ème}}$ valeur, soit **8**.
L'intervalle interquartile est donc $8 - 2 = 6$.
5. La moyenne égale :
$$m = \frac{0\times 5+\ldots+9\times 4}{50} = \frac{233}{50} = 4,66$$
L'écart-type égale :
$$\sigma = \sqrt{\frac{5\times(0-m)^2+\ldots+4\times(9-m)^2}{50}} \approx 2,96$$
6. L'écart-type est très important par rapport à la moyenne, on en déduit que les valeurs sont très hétérogènes.

Exercice 18

Les résultats d'un concours sont donnés dans le tableau suivant :

Notes	**[0 ; 5[**	**[5 ; 8[**	**[8 ; 10[**	**[10 ; 12[**	**[12 ; 16[**	**[16 ; 20]**
Effectif	17	16	26	30	24	11

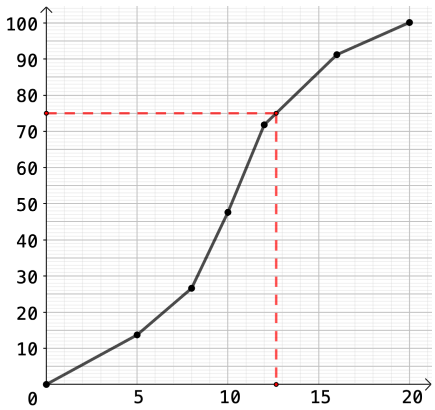
1. *Quelle est la population étudiée ? Quelle est la variable (ou caractère) et le type de cette variable ?*
2. *Faire le tableau de distribution des fréquences (en %).*
3. On sait que 25**%** des candidats sont admis. *Déterminer à l'aide d'un graphique approprié la note à partir de laquelle un candidat est admis. Peut-on calculer ce nombre ?*
4. *Calculer la note moyenne m et l'écart-type.*

Solution

1. La population étudiée est **l'ensemble des participants au concours**. La variable étudiée est **la note obtenue au concours**. Cette variable est **quantitative discrète**.

2. Je trace le tableau de distribution des fréquences (en **%**).

Notes	**[0 ; 5[**	**[5 ; 8[**	**[8 ; 10[**	**[10 ; 12[**	**[12 ; 16[**	**[16 ; 20]**	**Total**
Effectif	17	16	26	30	24	11	124
Fréquence (en %)	**13,7**	**12,9**	**21**	**24,2**	**19,4**	**8,9**	
FCC	13,7	26,6	47,6	71,8	91,2	100,1	



3. On utilise un graphique où sont indiquées les fréquences cumulées croissantes. Par interpolation affine, on peut estimer que les candidats seront admis à partir de **12,6**. Nous connaissons pas la distribution des notes dans l'intervalle [12 ; 16[donc nous ne pouvons pas connaître la note exacte qu'il faut avoir pour obtenir le concours. Par contre, on peut retrouver la valeur 12,6 par le calcul.

$$\frac{75}{100}\times 124 = 93$$

Donc 93 candidats ne seront pas admis. Il faut donc trouver la note du $94^{ème}$ candidat par interpolation affine. On estime que les 24 notes entre 12 et 16 sont également réparties. Avec cette interpolation, la première note de cet intervalle sera

$$12$$

Puis

$$12 + \frac{1}{24} \times (16 - 12) = 12 + \frac{4}{24} = 12 + \frac{1}{6}$$

Puis

$$12 + \frac{2}{24} \times 4 = 12 + \frac{8}{24} = 12 + \frac{2}{6}$$

jusqu'à

$$12 + \frac{23 \times 4}{24} = 12 + \frac{23}{6}$$

$16 - 12 = 4$ est l'amplitude de l'intervalle [12 ; 16[.

La première note de cet intervalle correspond à la note du $90^{ème}$ candidat

$$17 + 16 + 26 + 30 + 1 = 90$$

On en déduit, à l'aide de cette interpolation affine, que le $94^{ème}$ candidat a obtenu

$$12 + \frac{4}{6} = 12 + \frac{2}{3} \approx 12,66$$

Qui est la note d'admission au concours (si l'on suit un modèle linéaire, ce qui n'est jamais le cas en réalité…).

4. Pour calculer la note moyenne m et l'écart-type (dont on ne peut calculer une valeur exacte), on va utiliser le centre des classes et faire comme si chaque candidat d'une classe avait obtenu la même note.

Notes	**[0 ; 5[**	**[5 ; 8[**	**[8 ; 10[**	**[10 ; 12[**	**[12 ; 16[**	**[16 ; 20]**	**Total**
Centre de la classe	**2,5**	**6,5**	**9**	**11**	**14**	**18**	
Effectif	17	16	26	30	24	11	124

On peut alors calculer la moyenne

$$m = \frac{2,5 \times 17 + \ldots 18 \times 11}{124} = \frac{1\,244,5}{124} \approx 10,04$$

Et l'écart-type σ

$$\sigma = \sqrt{\frac{17(2,5-m)^2 + \ldots + 11(18-m)^2}{124}} \sigma = \sqrt{\frac{17 \times 2,5^2 + \ldots + 11 \times 18^2}{124} - m^2} \sigma = \sqrt{\frac{14\,786,25}{124} - \left(\frac{1\,244,5}{124}\right)^2} \sigma = \frac{\sqrt{14\,786,25 \times 124 - 1\,244,5^2}}{124} \sigma = \frac{\sqrt{2}}{}$$

SÉANCE 3 PROBABILITÉS

Exercice 19

soit P une probabilité et A, B, C trois événements.

Exprimer en fonction de P(A), P(B), P(C), $P(A \cap B)$, $P(A \cap C)$, $P(B \cap C)$, $P(A \cap B)$ et $P(A \cap B \cap C)$ les probabilités des événements suivants :

1. A seul se produit
2. A et B se produisent mais pas C
3. les trois événements se produisent simultanément
4. aucun des trois événement ne se produit

Solution

1. A seul se produit

$$P(A \cap \overline{B} \cap \overline{C}) = P(A \cap \overline{B}) - P(A \cap \overline{B} \cap C) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

2. A et B se produisent mais pas C $\quad P(A \cap B \cap \overline{C}) = P(A \cap B) - P(A \cap B \cap C)$

3. les trois événements se produisent simultanément $P(A \cap B \cap C)$

4. aucun des trois événement ne se produit

$$P(\overline{A} \cap \overline{B} \cap \overline{C}) = P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C)$$

Or,

$P(A \cup B \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C)$
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P((A \cap C) \cup (B \cap C))$
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P((A \cap C) \cap (B \cap C))$
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Donc

$P(\overline{A} \cap \overline{B} \cap \overline{C}) = 1 - P(A) - P(B) - P(C) + P(A \cap B) + P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)$

Exercice 20

Dans une urne, il y a 3 boules blanches, 4 boules vertes, 2 boules rouges et une boule jaune. On tire deux boules simultanément et on regarde la couleur des deux boules.

1. *Quel est l'ensemble fondamental lié à cette expérience ? Citer un événement impossible et un événement certain liés à cette expérience.*

2. *Est-ce une situation d'équiprobabilité ?*

3. *Quelle est la probabilité d'obtenir deux boules rouges (le tirage s'apparente au tirage successif de deux boules sans remise) ?*

Solution

1. L'ensemble fondamental lié à cette expérience est $\Omega = \{BB\ ; BV\ ; BR\ ; BJ\ ; VV\ ; VR\ ; VJ\ ; RR\ ; RJ\}$

« Obtenir deux boules jaunes » est un événement impossible.

« Obtenir une boule de couleur » est un événement certain.

2. Ce n'est pas une situation d'équiprobabilité car toutes les issues n'ont pas la même probabilité de se réaliser.

3. La probabilité d'obtenir deux boules rouges est de

$$\frac{2}{10} \times \frac{1}{9} = \frac{2}{90} = \frac{1}{45}$$

Exercice 21

Une urne contient 5 boules bleues, 4 boules vertes et 1 boule rouge.

1. La première expérience consiste à tirer une boule au hasard et regarder la couleur.

a. *Quelles sont les issues de l'expérience ? Cite un événement certain, un événement impossible, deux événements contraires.*

b. *Quelle est la probabilité d'obtenir une boule bleue ?*

2. La deuxième expérience consiste à tirer une boule au hasard et regarder la couleur : si l'on obtient une boule bleue, on a gagné. Si on tire une boule verte, on l'écarte et on retire une autre boule. Enfin, si l'on tire la boule rouge, on a perdu.

a. *Quelles sont les issues de l'expérience ? Est-ce une situation d'équiprobabilité ?*

b. *Quelle est la probabilité de perdre ?*

Solution

1. a. « Tirer une boule bleue », **« tirer une boule verte »** et **« tirer une boule rouge »** sont les **issues** de l'expérience.

« Tirer une boule de couleur » est un événement **certain**.

« Tirer une boule jaune » est un événement **impossible**.

« Tirer une boule de couleur primaire » et **« tirer une boule verte »** sont deux événements contraires.

1. b. La probabilité d'obtenir une boule bleue est de

$$\frac{5}{10} = \frac{1}{2}$$

2. a. « Gagner » et **« perdre »** sont les deux issues de l'expérience. Il y a plus de boules bleues que de rouges donc ce n'est **pas une situation d'équiprobabilité**.

b. Le tirage d'une boule verte n'entre pas en jeu dans le fait de perdre ou de gagner, on peut donc faire comme si elles n'existaient pas. La probabilité de perdre est donc celle de tirer une boule rouge parmi 5 bleues, soit une probabilité de

$$P(P) = \frac{1}{6}$$

Remarque

Pour s'en convaincre, on peut regarder tous les événements qui amènent à perdre et additionner leurs probabilités :

$$P(P) = P(R) + P(VR) + P(VVR) + P(VVVR) + P(VVVVR)$$

$$P(P) = \frac{1}{10} + \frac{4}{10} \times \frac{1}{9} + \frac{4}{10} \times \frac{3}{9} \times \frac{1}{8} + \frac{4}{10} \times \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} + \frac{4}{10} \times \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} \times \frac{1}{6}$$

On réduit au même dénominateur en mettant tout sur $10\times9\times8\times7\times6 = 30\ 240$

$$P(P) = \frac{3024+1344+504+144+24}{30\ 240} = \frac{5\ 040}{30\ 240} = \frac{1}{6}$$

Exercice 22

Le jeu du loto consiste à deviner les 6 entiers qui vont être tiré au hasard parmi les entiers $\{1 ; \ldots ; 49\}$.

1. *Comment modéliser le tirage de 6 entiers parmi {1 ; . . . ; 49} (Expliciter l'univers Ω et la probabilité P sur Ω) ?*

2. On considère les événements suivants :

A_k = « *Avoir deviné exactement k bon(x) numéro(s)* », $k\in\{1 ; \ldots ; 6\}$

P = « *Perdre* » = « *Avoir 0, 1 ou 2 bon(s) numéro(s)* »

G = « *Gagner* » = « *Avoir au moins 3 bons numéros* »

Calculer la probabilité de ces événements.

Solution

1. $\Omega = \{\{x_1; x_2; x_3; x_4; x_5; x_6\},\ i < j \Rightarrow x_i < x_j,\ x_i \in [\![1; 49]\!]\}$. Si $\omega\in\Omega$ est une issue,

$$P(\omega) = \frac{6\times5\times4\times3\times2\times1}{49\times48\times47\times46\times45\times44} = \frac{720}{10\ 068\ 347\ 520} = \frac{1}{13\ 983\ 816}$$

2. On calcule la probabilité des événements suivants :

$$P(A_0) = \binom{6}{0} \times \frac{43\times42\times41\times40\times39\times38}{49\times48\times47\times46\times45\times44} = \frac{4\ 389\ 446\ 880}{10\ 068\ 347\ 520}$$

$$P(A_1) = \binom{6}{1} \times \frac{6\times43\times42\times41\times40\times39}{49\times48\times47\times46\times45\times44} = \frac{4\ 158\ 423\ 360}{10\ 068\ 347\ 520}$$

$$P(A_2) = \binom{6}{2} \times \frac{6\times5\times43\times42\times41\times40}{49\times48\times47\times46\times45\times44} = \frac{1\ 332\ 828\ 000}{10\ 068\ 347\ 520}$$

$$P(A_3) = \binom{6}{3} \times \frac{6\times5\times4\times43\times42\times41}{49\times48\times47\times46\times45\times44} = \frac{177\ 710\ 400}{10\ 068\ 347\ 520}$$

$$P(A_4) = \binom{6}{4} \times \frac{6\times5\times4\times3\times43\times42}{49\times48\times47\times46\times45\times44} = \frac{9\ 752\ 400}{10\ 068\ 347\ 520}$$

$$P(A_5) = \binom{6}{5} \times \frac{6\times5\times4\times3\times2\times43}{49\times48\times47\times46\times45\times44} = \frac{185\ 760}{10\ 068\ 347\ 520}$$

$$P(A_6) = \binom{6}{6} \times \frac{6\times5\times4\times3\times2\times1}{49\times48\times47\times46\times45\times44} = \frac{720}{10\ 068\ 347\ 520}$$

$$P(P) = P(A_0) + P(A_1) + P(A_2) = \frac{9\ 880\ 698\ 240}{10\ 068\ 347\ 520}$$

$$P(G) = P(\overline{P}) = 1 - P(P) = \frac{187\ 649\ 280}{10\ 068\ 347\ 520}$$

Exercice 23

Une urne contient une boule blanche et une boule noire. On fait deux tirages avec remise. *Quelle est la probabilité d'obtenir deux boules noires ?*

Solution

La probabilité d'obtenir deux boules noires est de

$$P(NN) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Exercice 24

Une urne contient une boule blanche, une boule noire, une boule rouge et une boule jaune. On fait trois tirages avec remise. *Quelle est la probabilité d'obtenir au moins deux boules blanches ?*

Solution

Pour obtenir au moins deux boules blanches, il y a 10 issues possibles : BBR, BBN, BBJ, BRB, BNB, BJB, RBB, NBB, JBB, BBB.

La probabilité de chacune de ces issues est

$$\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{64}$$

Donc la probabilité d'obtenir au moins deux boules blanches est de

$$P(BB) = \frac{10}{64}$$

Exercice 25

Une urne contient deux boules bleues et trois boules vertes.

1. On fait deux tirages sans remise.

a. *Quelle est la probabilité d'obtenir deux boules bleues ?*

b. *Quelle est la probabilité d'obtenir deux boules vertes ?*

c. *Quelle est la probabilité d'obtenir une boule verte et une boule bleue ?*

2. *Reprendre la question 1 avec deux tirages avec remise.*

Solution

Sans remise

1. a. La probabilité d'obtenir deux boules bleues est de

$$P(BB) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}$$

b. La probabilité d'obtenir deux boules vertes est de

$$P(VV) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

c. La probabilité d'obtenir une boule verte et une boule bleue est de

$$P(BV) = 1 - P(BB) - P(VV) = \frac{6}{10}$$

Avec remise

2. a. La probabilité d'obtenir deux boules bleues est de

$$P(BB) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$$

b. La probabilité d'obtenir deux boules vertes est de

$$P(VV) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$$

c. La probabilité d'obtenir une boule verte et une boule bleue est de

$$P(BV) = 1 - P(BB) - P(VV) = \frac{12}{25}$$

Exercice 26

Un virus est présent dans une population. Un individu étant tiré au hasard, on note V l'événement « l'individu est porteur du virus ». On effectue un test de dépistage. On note T l'événement « le test est positif ». On obtient expérimentalement les valeurs suivantes sur un échantillon de la population.

$$P(V) = p \; P_V(T) = r \; P_{\overline{V}}(T) = s$$

1. *Montrer que* $P(T) = pr + (1-p)s$

2. *Calculer* $P_T(V)$ *en fonction de p, r et s.*

3. Application numérique. On donne

$$r = 0,99 \; s = 0,001$$

Calculer numériquement $P_T(V)$ *dans les deux cas suivants.*

$$a.\; p = 0,001 \; b.\; p = 0,3$$

4. *Dans lequel de ces deux cas peut-on dire que le test est bon ?*

Solution

1. Montrons que $P(T) = pr + (1-p)s$

D'après la formule des probabilités totales, on peut écrire

$$P(T) = P(V \cap T) + P(\overline{V} \cap T) = P(V) \times P_V(T) + P(\overline{V}) \times P_{\overline{V}}(T) = pr + (1-p)s$$

2. Calculons $P_T(V)$ en fonction de p, r et s.

$$P_T(V) = \frac{P(V \cap T)}{P(T)} = \frac{pr}{pr+(1-p)s}$$

3. a. Calculons $P_T(V)$ lorsque $p = 0,001$

$$P_T(V) = \frac{pr}{pr+(1-p)s} = \frac{0,001 \times 0,99}{0,001 \times 0,99 + 0,999 \times 0,001} = \frac{110}{221} \approx 0,498$$

3. a. Calculons $P_T(V)$ lorsque $p = 0,3$

$$P_T(V) = \frac{pr}{pr+(1-p)s} = \frac{0,3 \times 0,99}{0,3 \times 0,99 + 0,999 \times 0,001} = \frac{11\,000}{11\,037} \approx 0,997$$

4. D'après la question précédente, on peut dire que le test est excellent si $p = 0,3$, très incertain si $p = 0,001$.

Exercice 27

1. Un test médical a pour but de comparer deux méthodes de traitements de calculs (rénaux), la méthode par chirurgie notée C et la méthode par ultra-sons notée U. 700 patients ont été traités, 350 par chirurgie et 350 par ultrasons. 273 traités par la méthode C ont guéri et 289 par la méthode U. On note R l'événement « une personne prise au hasard parmi les patients traités a été guérie ».

Calculer $P_C(R)$ et $P_U(R)$. *Que peut-on en déduire ?*

2. Une analyse plus fine ne tenant compte de la taille des calculs traités donne les valeurs suivantes.

	Méthode C	**Méthode U**
Gros calculs C_g	191 réussites/71 échecs	60 réussites/26 échecs
Petits calculs C_p	82 réussites/6 échecs	229 réussites/35 échecs

Calculer $P_{C \cap C_g}(R)$ et $P_{U \cap C_g}(R)$, $P_{C \cap C_p}(R)$ et $P_{U \cap C_p}(R)$. *Que peut-on en déduire ? Comment expliquer cette différence ?*

Solution

1. On calcule

$P_C(R) = \frac{273}{350} = 0,78$ *et* $P_U(R) = \frac{289}{350} \approx 0,826$

On peut en déduire que la méthode par **ultra-sons** semble plus efficace.

2. On calcule

$P_{C \cap C_g}(R) = \frac{191}{262} \approx 0,729$ $P_{U \cap C_g}(R) = \frac{60}{86} \approx 0,698$ $P_{C \cap C_p}(R) = \frac{82}{88} \approx 0,932$ $P_{U \cap C_p}(R) = \frac{229}{264} \approx 0,867$

On peut en déduire que la méthode par **chirurgie** est **plus efficace**. La différence provient du fait que l'on a traité en majorité des gros calculs par chirurgie alors que ceux-ci ont un taux de succès moindre, ce qui influe sur le taux général de réussite de la chirurgie.

Exercice 28
Dans l'article «Causal schemas in judgment under uncertainty» (*Progress in Social Psychology*, Martin Fishbein éditeur, 1980), Amos Tversky et Daniel Kahneman présentent une application intéressante du théorème de Bayes. Décrivons d'abord le problème traité par ces auteurs.
Dans une ville, deux compagnies opèrent des taxis : les Verts, qui représentent 85% de la flotte totale, et les Bleus, qui ont les 15% restants. Un accident survient la nuit et le véhicule impliqué quitte la scène. Un témoin affirme qu'il s'agissait d'un taxi Bleu.
Quelle est la probabilité que le taxi responsable soit effectivement un Bleu, sachant que la fiabilité des témoignages dans des conditions analogues à celles prévalant lors de l'accident est évaluée à 80% (les témoins identifient correctement la couleur du taxi dans 80% des cas, et se trompent dans 20%) ?
Avant de résoudre le problème, mentionnons que les auteurs ont demandé à plusieurs sujets d'évaluer intuitivement cette probabilité. La réponse médiane et modale est 80%. Ainsi, les gens, spontanément, sont prêts à donner une crédibilité élevée à la parole du témoin.
Qu'en est-il en réalité ? En notant T l'événement « le témoin affirme que le taxi est bleu », on pourra calculer dans cet ordre $P(V)$, $P(B)$, $P_B(T)$, $P_V(T)$, $P(T\cap B)$, $P(T\cap V)$, $P(T)$ puis finalement $P_T(B)$.

Solution
D'après l'énoncé,
$P(V) = 0,85$ $P(B) = 0,15$ $P_B(T) = 0,8$ $P_V(T) = 0,2$
En utilisant les probabilités conditionnelles, on obtient
$P(T\cap B) = P(B) \times P_B(T) = 0,15\times 0,8 = 0,12$ $P(T\cap V) = P(V) \times P_V(T) = 0,85\times 0,2 = 0,17$
Les événements B et V étant contraires, on peut appliquer la formule des probabilités totales.
$$P(T) = P(T\cap B) + P(T\cap V) = 0,12 + 0,17 = 0,29$$
Ainsi,
$$P_T(B) = \frac{P(T\cap B)}{P(T)} = \frac{0,12}{0,29} = \frac{12}{29} \approx 0,414$$
On définitive, si le témoin affirme que le taxi est bleu, il y a une probabilité plus forte qu'il soit vert !

Exercice 29

Une urne contient 3 sortes de boules de masses différentes : 7 boules de masse 1 kg, 5 boules de masse 3 kg et 3 boules de masse 5 kg. On tire au hasard une boule de l'urne et on note X sa masse.

1. *Déterminer la loi de la variable X.*

2. *Calculer l'espérance et la variance de X.*

Solution

1. Je détermine la loi de la variable X.

$$P(X = 1) = \frac{7}{15}$$
$$P(X = 3) = \frac{5}{15}$$
$$P(X = 5) = \frac{3}{15}$$

2. Je calcule l'espérance et la variance de X.

$E(X) = 1\times\frac{7}{15} + 3\times\frac{5}{15} + 5\times\frac{3}{15} = \frac{37}{15}$

$Var(X) = E\left(X^2\right) - E^2(X) = 1^2 \times \frac{7}{15} + 3^2 \times \frac{5}{15} + 5^2 \times \frac{3}{15} - \left(\frac{37}{15}\right)^2 = \frac{127}{15} - \frac{1369}{225} = \frac{536}{225}$

Exercice 30

La proportion des groupes sanguins en France est environ :

$A = 44\%$ $B = 13\%$ $AB = 3\%$ $O = 40\%$

On considère la répartition de ces différents groupes sur 50 étudiants.

1. *Donner la loi de la variable X égale au nombre d'étudiants de groupe O. Calculer la probabilité d'avoir exactement 20 étudiants du groupe O parmi ces 50 étudiants.*

2. *Donner la loi de la variable Y égale au nombre d'étudiants de groupe AB.*

3. *Calculer $P(Y\leq5)$, l'espérance et la variance de Y.*

Solution

1. La loi de X s'apparente à la succession de 50 épreuves aléatoires indépendantes de même loi de Bernoulli de paramètre 0,4 donc $X\sim B(50; 0,4)$

$$P(X = 20) = \binom{50}{20} \times 0,4^{20} \times 0,6^{30} \approx 0,115$$

2. La loi de Y s'apparente à la succession de 50 épreuves aléatoires indépendantes de même loi de Bernoulli de paramètre 0,03 donc $Y\sim B(50; 0,03)$

3. Je calcule $P(Y\leq5)$ et l'espérance et la variance de Y.

$Y\sim B(50; 0,03)\sim P(\lambda)$ où $\lambda = 50\times0,03 = 1,5$

$P(Y\leq5) = \sum_{k=0}^{5} \binom{50}{k} \times 0,03^{k} \times 0,97^{50-k} \approx 0,996$

$E(X) = 50\times0,03 = 1,5$

$Var(X) = 50\times0,03\times0,97 = 1,455$

Exercice 31

Considérons les enfants de parents hétérozygotes de génotype Aa. La distribution des enfants est

$P(AA) = 1/4$ $P(Aa) = 1/2$ $P(aa) = 1/4$

On choisit de façon aléatoire 240 de ces enfants. On définit N_1, N_2, N_3 le nombre d'enfants de génotype AA, Aa et aa respectivement.

1. *Quelle est la loi de N_1 ? N_2 ? N_3 ? Calculer l'espérance et la variance pour chaque génotype.*

2. *Quel est le lien entre ces différentes variables ?*

Solution

1. N_1, N_2 et N_3 suivent des schémas de Bernoulli constitués de 240 épreuves indépendantes de même loi de de Bernoulli de paramètres respectifs 0,25, 0,5 et 0,25 donc

$N_1\sim B(240; 0,25)$ $E\left(N_1\right) = 240\times0,25 = 60$ $Var\left(N_1\right) = 240\times0,25\times0,75 = 45$

$N_2\sim B(240; 0,5)$ $E\left(N_2\right) = 240\times0,5 = 120$ $Var\left(N_2\right) = 240\times0,5\times0,5 = 60$

$N_3\sim B(240; 0,25)$ $E\left(N_3\right) = 240\times0,25 = 60$ $Var\left(N_3\right) = 240\times0,25\times0,75 = 45$

2. $N_1 + N_2 + N_3 = 240$

Exercice 32
Une machine à embouteiller peut tomber en panne. La probabilité d'une panne à chaque emploi est de 0,01. La machine doit être utilisée 100 fois.
Soit X le nombre de pannes obtenues après 100 utilisations.
1. *Quelle est la loi de X ?*
Calculer $P(X = 0)$*,* $P(X = 1)$ *et* $P(X \geq 4)$*.*
2. On estime le coût d'une réparation à 500 F. Soit la variable aléatoire Y représentant la dépense pour les réparations après 100 utilisations.
Exprimer Y en fonction de X et calculer E(Y) et Var(Y).
Solution
1. La loi de X s'apparente à la succession de 100 épreuves aléatoires indépendantes de même loi de Bernoulli de paramètre 0,01 donc $X \sim B(100; 0,01)$
$P(X = 0) = \binom{100}{0} \times 0,01^0 \times 0,99^{100} = 0,99^{100} \approx 0,366$
$P(X = 1) = \binom{100}{1} \times 0,01^1 \times 0,99^{99} = 0,99^{99} \approx 0,370$
$P(X \geq 4) = 1 - P(X \leq 3) \approx 0,018$
2. $Y = 500 \cdot X$ donc $E(Y) = 500 \cdot E(X) = 500 \cdot 100 \times 0,01 = 500$
$Var(Y) = 500^2 \cdot Var(X) = 250\,000 \cdot 100 \times 0,01 \times 0,99 = 247\,500$

Exercice 33
Dans une pépinière, 95% des scions (jeunes arbres greffés) sont supposés sans virus. Par commodité, les scions sont rangés par paquets de 2. Un paquet est dit sain si les 2 scions le sont.
1. *Quelle est la probabilité d'avoir un paquet sain ?*
2. Soit X le nombre de paquets sains sur un lot de 10. *Quelle est la loi de X ?*
3. Un lot de 10 est accepté par l'acheteur si 9 au moins des paquets sont sains. *Quelle est la probabilité qu'un lot soit accepté ?*
Solution
1. La probabilité d'avoir un paquet sain est de $0,95^2 = 0,9025$
2. $X \sim B(10; 0,9025)$
3. La probabilité qu'un lot soit accepté égale
$P(X \geq 9) = P(X = 9) + P(x = 10) = \binom{10}{9} \bullet 0,9025^9 \bullet 0,0975^1 + \binom{10}{10} \bullet 0,9025^{10} \bullet 0,0975^0$
$P(X \geq 9) = 10 \bullet 0,9025^9 \bullet 0,0975^1 + 0,9025^{10} \approx 0,746$

Exercice 34
Lors d'un sondage portant sur un grand nombre de personnes, on sait que 2% des personnes interrogées acceptent de ne pas rester anonymes. *Sachant que l'un des sondeurs a interrogé 250 personnes (en les choisissant de manière indépendante), calculer la probabilité :*
1. *Que ces 250 personnes souhaitent rester anonymes.*
2. *3 personnes acceptent de ne pas rester anonymes.*
3. *Plus de 10 personnes acceptent de ne pas rester anonymes.*
Solution
1. Soit X la variable aléatoire qui compte le nombre de personnes qui acceptent de ne pas rester anonymes dans cet échantillon de 250 personnes. La loi de X s'apparente à la succession de 250 épreuves aléatoires indépendantes de même loi de Bernoulli de paramètre 0,02 donc $X \sim B(250; 0,02)$
$P(X = 0) = \binom{250}{0} \times 0,02^0 \times 0,98^{250} = 0,98^{250} \approx 0,0064$
2. La probabilité que 3 personnes acceptent de ne pas rester anonymes est de
$P(X = 3) = \binom{250}{3} \times 0,02^3 \times 0,98^{247} \approx 0,14$
3. La probabilité que plus de 10 personnes acceptent de ne pas rester anonymes est de
$P(X \geq 11) = \sum_{k=11}^{250} (250\ k) \times 0,02^k \times 0,98^{250-k} \approx 0,0128$

Exercice 35
Dans un hôpital parisien, il arrive en moyenne 1,25 personnes à la minute aux urgences entre 9 h et 12 h.
Soit X le nombre de personnes observées à la minute à l'entrée de ce service. On admet que
$$X \sim P(\lambda) \text{ avec } \lambda = 1,25$$
Déterminer les probabilités suivantes.
1. *En 1 mn, il arrive 2 personnes.*
2. *En 1 mn, il arrive 4 personnes au plus.*
3. *En 1 mn, il arrive 3 personnes au moins.*
Solution
1. $P(X = 2) = \frac{1,25^2}{2!} e^{-1,25} \approx 0,224$

2. $P(X \leq 4) = \sum_{k=0}^{4} \frac{1,25^k}{k!} e^{-1,25} \approx 0,991$

3. $P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} \frac{1,25^k}{k!} e^{-1,25} \approx 0,132$

Exercice 36
Une enquête a permis d'établir que 1% des Français ont un IMC (indice de masse corporelle) supérieur à 40. Dans cette situation, la personne est atteinte d'obésité massive. On interroge au hasard et de façon indépendante, 400 français. On note X la variable aléatoire réelle qui associe le nombre de personnes atteintes d'obésité massive à ce tirage.
1. *Quelle est la loi suivie par X. Préciser les paramètres.*
Cette loi peut être approchée par une loi de Poisson.
2. *Quelle est son espérance ? En donner une interprétation.*
3. *Quelle est la probabilité que le nombre de personnes atteintes d'obésité massive soit au moins égal à 5 ?*
Solution
1. La loi de X s'apparente à la succession de 400 épreuves aléatoires indépendantes de même loi de Bernoulli de paramètre 0,01 donc $X \sim B(400; 0,01)$
2. X peut être approchée par une loi de Poisson de paramètre $400 \times 0,01 = 4$, correspondant à l'espérance de cette loi. Cela veut dire qu'en moyenne, il y aura 4 personnes atteintes d'obésité massive dans cet échantillon.
3. La probabilité que le nombre de personnes atteintes d'obésité massive soit au moins égal à 5 est de
$$P(X \geq 5) \approx 0,371 \text{ (avec la loi binomiale)}$$
$$P(X \geq 5) \approx 0,371 \text{ (avec la loi de Poisson)}$$