

1 Statistique descriptive

Exercice 1.1. *La queue au péage.* À un poste de péage, on compte le nombre de voitures se présentant sur une période de 5 minutes, et ce plusieurs fois. Sur 100 observations de 5 minutes, on obtient les résultats suivants :

nombre de voitures	0	1	2	3	4	5	6	7	8	9	10	11	12
nombre d'observations	1	2	7	14	20	19	15	9	6	2	3	1	1

- 1) Préciser la population et le caractère étudiés. De quel type de caractère s'agit-il ?
- 2) Calculer les fréquences et les fréquences cumulées.
- 3) Représenter le diagramme en bâtons et le diagramme des fréquences cumulées.
- 4) Calculer la moyenne, la variance, l'écart-type, le mode, l'étendue, la médiane et les quartiles.
- 5) Par quel type de loi usuelle a-t-on l'habitude de modéliser le nombre de voitures dans une file d'attente ?

Exercice 1.2. *Transformation affine de la moyenne.* Soit X un caractère et $x = (x_1, x_2, \dots, x_n)$ la série statistique associée à l'observation de X sur un échantillon. Soit a et b deux réels, $Y = aX + b$ un nouveau caractère et $y = (y_1, y_2, \dots, y_n)$ la série statistique associée à l'observation de Y sur le même échantillon (autrement dit, $y_i = ax_i + b$).

- 1) Montrer que, si $\bar{x}$ est la moyenne de la série $(x_1, x_2, \dots, x_n)$, alors la moyenne $\bar{y}$ de la série $(y_1, y_2, \dots, y_n)$ se calcule par

$$\bar{y} = a\bar{x} + b.$$

Un laboratoire de physique nucléaire a réalisé 20 mesures de la masse d'un électron au repos (en yoctogrammes ; $1 \text{ yg} = 10^{-24} \text{ g}$) :

91 095,34	91 095,29	91 095,31	91 095,36	91 095,62
91 095,37	91 095,35	91 095,07	91 095,41	91 095,39
91 095,45	91 095,82	91 095,34	91 095,16	91 095,25
91 096,03	91 095,20	91 095,31	91 095,40	91 095,28

- 2) Déterminer la moyenne de ces mesures (qui sera une bonne estimation de la masse de l'électron au repos).

Moralité : La propriété d'affinité de la moyenne simplifie non seulement les calculs à la main, mais aussi la saisie de données dans les ordinateurs (en diminuant les risques d'erreur).

Exercice 1.3. *Ce que nous dit l'écart-type.* Soit X un caractère statistique et $x = (x_1, \dots, x_n)$ la série statistique associée sur un échantillon. On note $\bar{x}$ la moyenne de cette série et $\sigma(x)$ son écart-type. On se propose d'établir : pour tout $\alpha > 0$, la proportion de données dont l'écart à la moyenne est strictement supérieur à $\alpha\sigma(x)$ est inférieure à $1/\alpha^2$.

- 1) On se donne $\alpha > 0$. On note A_α l'ensemble des indices i tels que $|x_i - \bar{x}| > \alpha\sigma(x)$ et n_α le nombre d'éléments de A_α . En utilisant l'expression de la variance $\sigma(x)^2$ et en séparant les données de la série (soit i appartient à A_α , soit i n'appartient pas à A_α), montrer que

$$\sigma(x)^2 \geq \frac{n_\alpha}{n} (\alpha\sigma(x))^2$$

et conclure.

- 2) À qui doit-on cette inégalité ?
- 3) Montrer que, quelle que soit la série statistique de moyenne μ et d'écart-type σ , au moins 75% des données sont dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$.
- 4) Quel est le pourcentage maximal de données situées en dehors de l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$?

2 Modélisation probabiliste

2.1 Un peu de dénombrement

Exercice 2.1. *Nombre de chiffres.* Pour écrire tous les entiers de 100 à 999, on utilise 2 697 chiffres. Vrai ou faux ?

Exercice 2.2. *Drôle de dictionnaire.* Avec un alphabet de vingt-six lettres, combien peut-on former de mots de quatre lettres :

- 1) en admettant une répétition des lettres ?
- 2) sans lettre double ?
- 3) finissant par une voyelle (les voyelles de l'alphabet étant A, E, I, O, U et Y) et en admettant une répétition des lettres ?

Exercice 2.3. *Loto.* Un tirage du loto consiste en cinq boules différentes parmi des boules numérotées de 1 à 49 et un numéro chance, entre 1 et 10. Combien y a-t-il de tirages différents au loto ?

Exercice 2.4. *Plaques d'immatriculation.* En France, les plaques d'immatriculation sont de la forme AA-001-AA. Les chiffres évoluent en premier, puis le bloc de droite, et enfin le bloc de gauche. Il faut savoir que les lettres I, O et U ne sont pas employées, ni la série SS ; et, dans le bloc de gauche, la série WW n'est pas employée.

- 1) Montrer que l'on peut former 277 977 744 plaques d'immatriculations.
- 2) J'ai repéré une voiture immatriculée AG-007-JB. Quel est le rang de sa plaque, autrement dit la combienième plaque est-ce ?
- 3) Je ne sais pas si la plaque CA-001-AA existe déjà, mais quel est ou sera le rang de cette plaque ?

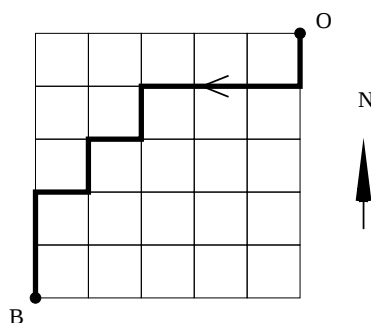
Exercice 2.5. *Triangles.* On place dix points dans le plan de sorte que trois quelconques ne soient jamais alignés. Combien peut-on tracer de triangles différents dont les sommets sont parmi ces dix points ?

Exercice 2.6. *Est-ce dans vos cordes ?* Soit n un entier strictement positif. On place n points distincts sur un cercle.

- 1) Combien déterminent-ils de cordes ?
- 2) Quel est le nombre de points d'intersection de ces cordes, en supposant qu'en chaque point d'intersection passent au plus deux cordes ?

Exercice 2.7. *Anagrammes.* Combien y a-t-il d'anagrammes du mot MIRE ? Et du mot BAOBAB ?

Exercice 2.8. *Promenade dans New York.* Combien y a-t-il de plus courts chemins entre le point O et le point B .



Exercice 2.9. *Course en tête.* Au cours d'une série de huit lancers d'une pièce de monnaie, nous dirons que PILE « fait la course en tête » si, tout au long de la série, le nombre de PILE obtenus est toujours supérieur ou égal au nombre de FACE. Combien y a-t-il de telles séries ?

Exercice 2.10. *Péage.* Un péage automatique d'autoroute n'accepte que les pièces de 2 €, 1 €, 50 cent, 20 cent et 10 cent. De combien de façon peut-on acquitter un montant de 3 € ?

2.2 Modélisation

Exercice 2.11. *Différentes façons de choisir.* Dans la grille suivante, on doit choisir une case blanche au hasard. On propose trois procédures pour ce choix :

- (A) On choisit une case blanche au hasard parmi les cinq cases blanches.
- (B) On choisit au hasard une ligne puis, dans la ligne, une case blanche au hasard.
- (C) On choisit au hasard une colonne puis, dans la colonne, une case blanche au hasard.

a		
b		c
	d	e

- 1) Modéliser la première procédure par une variable aléatoire X . On précisera l'univers Ω_X et la loi de X .
- 2) Modéliser la deuxième procédure par une variable aléatoire Y (*idem*).
- 3) Modéliser la troisième procédure par une variable aléatoire Z (*idem*).

On a simulé n choix d'une case blanche selon l'une des trois procédures ci-dessus, pour $n = 100$, puis $n = 1000$ et enfin $n = 10000$. On a obtenu les fréquences ci-dessous :

Case	a	b	c	d	e
Fréquence ($n = 100$)	0,19	0,16	0,22	0,31	0,12
Fréquence ($n = 1000$)	0,150	0,159	0,174	0,368	0,149
Fréquence ($n = 10000$)	0,167	0,168	0,164	0,336	0,165

- 4) Au vu de ce tableau, peut-on deviner la procédure utilisée ?

Remarque : L'observation des résultats simulés ne permet pas de remonter au modèle à coup sûr. Mais l'un des objectifs de la statistique est d'évaluer le risque de se tromper en « remontant d'une distribution de fréquences à une loi de probabilité ».

Exercice 2.12. *Pièce étrange.* On lance une pièce truquée, pour laquelle « face » sort dix fois plus souvent que « pile ». Modéliser un lancer de cette pièce (autrement dit, introduire une variable aléatoire...).

Exercice 2.13. *Dé étrange.* Un dé est pipé, de telle sorte que la probabilité de sortie d'un numéro est proportionnelle à ce numéro. Modéliser un lancer de ce dé.

Exercice 2.14. *Coup de Vénus.*



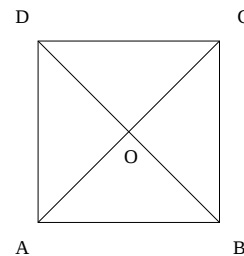
Les Grecs et les Romains jouaient aux « dés » avec des osselets du pied appelés astragales. Ils lançaient l'astragale qui pouvait retomber sur l'un de ses quatre côtés (tous marqués différemment) : deux petits et deux plus larges. L'expérience permet d'estimer à 0,1 la probabilité d'obtenir chacune des petites faces et à 0,4 celle d'obtenir chacune des grandes faces.

En lançant quatre astragales, quelle est la probabilité de réussir le « coup de Vénus », c'est-à-dire d'obtenir les quatre faces différentes ? On modélisera le lancer de quatre astragales pour répondre à la question.

Exercice 2.15. *Promenade aléatoire.*

Une fourmi part du point A et effectue trois déplacements suivant les segments de la figure, en choisissant, depuis chaque intersection, un segment au hasard.

- 1) Écrire tous les déplacements à l'aide d'un arbre.
- 2) Quelle est la probabilité que la fourmi se retrouve en A ? En B ? En C ? En D ? En O ?
- 3) Synthétiser ces résultats en dressant le tableau de la loi de la variable aléatoire : point d'arrivée de la fourmi.



Remarque : Les promenades aléatoires ont été imaginées au début du XX^e siècle par les probabilistes, tels que le Russe Andreï Markov ou le Français Henri Poincaré, comme des modèles :

- en physique, pour étudier le mouvement brownien d'une particule (grain de pollen dans l'air, molécule dans l'air...);
- en génétique, pour étudier la transmission d'un caractère héréditaire.

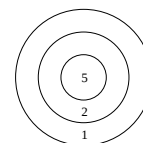
2.3 Variables aléatoires

Petit lexique des jeux — Etant donné un jeu (d'argent, ou de pièces en chocolat si vous préférez...), on note G la variable aléatoire "gain".

- *Espérance de gain* : l'espérance de gain $\mathbb{E}[G]$ est un bon prédicteur du gain moyen par partie, sur un grand nombre de parties.
- *Honnête, favorable ou défavorable* : on dit qu'un jeu est honnête si $\mathbb{E}[G] = 0$, favorable si $\mathbb{E}[G] > 0$ et défavorable si $\mathbb{E}[G] < 0$.
- *Risqué* : la dispersion de la variable G mesure le risque de perdre gros (ou de gagner gros, certes...). On pourra comparer les risques de deux jeux en comparant les variances des gains de ces deux jeux.

Exercice 2.16. *Sensible, mais pas sans...*

On suppose qu'un joueur de fléchette visant la cible ci-contre l'atteint toujours. La probabilité que le tir d'un joueur lambda arrive dans chacune des zones est proportionnelle à son aire (les cercles ont pour rayons 1, 2 et 3).



- 1) Modéliser l'expérience.

Le patron du bar veut faire de ce jeu de fléchette un jeu d'argent. Quand un client jouera, il devra miser m euros ; il gagnera, en retour le montant en euros inscrit dans la zone touchée.

- 2) Quelle mise m le patron doit-il choisir pour que le jeu soit honnête (cf. petit lexique des jeux) pour un joueur λ ?

Exercice 2.17. *Joueur ?* « Lancez 15 fois une pièce de monnaie. Si vous obtenez 15 fois FACE, vous gagnez 700 millions d’euros, sinon vous perdez 1 000 euros. »

- 1) Montrez que l'espérance de gain est supérieure à 20 000 euros.
- 2) Jouez-vous ?

Exercice 2.18. *Au casino.* Au jeu de la roulette, les 37 issues **0**, **1**, **2**, ..., **36** sont équiprobables.

- 1) Modéliser l'expérience.
- 2) Donner la probabilité des événements :

- 2a) "rouge";
- 2b) "noir";
- 2c) "pair" = "obtenir **2, 4, ..., ou 36**";
- 2d) "impair";
- 2e) "manque" = "obtenir **1, 2, ..., ou 18**";
- 2f) "passe" = "obtenir **19, 20, ..., ou 36**";
- 2g) " 12^P " = "obtenir **1, 2, ..., ou 12**";
- 2h) " 12^M " = "obtenir **13, 14, ..., ou 24**";
- 2i) " 12^D " = "obtenir **25, 26, ..., ou 36**".

- 3) Donner la probabilité des événements :
- 3a) "rouge et manque";
 - 3b) "noir ou passe";
 - 3c) "noir, impair et passe";
 - 3d) " 12^P , rouge et pair".

Envisageons trois stratégies de jeu :

- (A) On mise 1 euro sur "rouge". Si un numéro rouge sort, on gagne alors 1 euro (en plus de récupérer sa mise), sinon on perd sa mise.
- (B) On mise 1 euro sur un numéro (mettons **13**). Si le numéro (ici **13**) sort, on gagne 35 euros (en plus de récupérer sa mise), sinon on perd sa mise.
- (C) On mise 1 euro sur la case "12^P". Si on gagne, on récupère 2 euros (en plus de sa mise), sinon on perd sa mise.

- 4) Pour chaque stratégie, donner la loi de la variable aléatoire gain, calculer son espérance et sa variance.
- 5) Comparer ces trois stratégies à l'aide du petit lexique des jeux.

PASSE			0			MANQUE		
			1	2	3			
			4	5	6			
			7	8	9			
PAIR			10	11	12	IMPAIR		
			13	14	15			
			16	17	18			
			19	20	21			
 (NOIR)	 (ROUGE)			22	23	24		
				25	26	27		
				28	29	30		
				31	32	33		
12 P	12 M	12 D	34	35	36	12 D	12 M	12 P

Exercice 2.19. On lance une pièce jusqu'à obtenir PILE. On appelle N le nombre de lancers effectués.

- 1) Pour tout $k \in \mathbb{N}$, calculer $\mathbb{P}(N > k)$.

- 2) En déduire la fonction de répartition de N , puis la densité de probabilité de N .
- 3) En sommant les lignes, puis les colonnes du tableau suivant, montrer que $\mathbb{E}[N] = 2$:

1/2	0	0	...
1/4	1/4	0	...
1/8	1/8	1/8	...
$\vdots$	$\vdots$	$\vdots$	$\ddots$

Exercice 2.20. *Match de tennis.*

Deux joueurs de tennis, de force égale, se disputent un jeu. Il y a égalité et on souhaite estimer le nombre d'échanges N avant que le jeu se termine. On modélise l'évolution du jeu par une marche suivant le schéma ci-contre.

- 1) Montrer que l'espérance du nombre d'échanges est donnée par

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} (2n)2^{-n}.$$

- 2) Pour $x \in]-1, 1[$, on définit

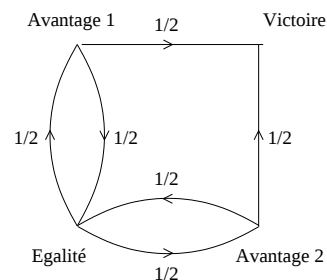
$$f(x) = \sum_{n=1}^{\infty} x^n.$$

En dérivant f , calculer $\mathbb{E}[N]$.

- 3) Calculer de même la variance de N .

Exercice 2.21. *Prof sympa ?* Un professeur note ses élèves en lançant trois dés icosaédriques et en affectant comme note le nombre le plus élevé qui est apparu.

- 1) Quel est le pourcentage de 20 prévisible (autrement dit la probabilité que la note donnée soit 20) ?
- 2) Calculer la fonction de répartition de la variable aléatoire « note ».
- 3) Quelle est la médiane de la classe prévisible (autrement dit la médiane de la variable aléatoire « note ») ?
- 4) Quel est la moyenne de la classe prévisible (autrement dit l'espérance de la variable aléatoire « note ») ?



3 Lois discrètes usuelles

Exercice 3.1. *Variable constante...* Soit X une variable aléatoire constante.

- 1) Donner la densité de probabilité de X .
- 2) Calculer l'espérance et la variance de X .
- 3) Représenter la fonction de répartition de X .

Exercice 3.2. *Anniversaires.* On choisit 500 personnes au hasard.

- 1) Quelle est la probabilité qu'exactly 3 d'entre elles aient leur anniversaire le premier mars ?
- 2) Quelle est la probabilité qu'au moins une d'entre elles soit née un 29 février ?
- 3) Combien faut-il choisir de personnes pour que la probabilité qu'au moins une d'entre elles fête son anniversaire le même jour que vous soit supérieure à $1/2$?
- 4) Combien faut-il choisir de personnes pour la probabilité qu'au moins deux d'entre elles fêtent leur anniversaire le même jour soit supérieure à $1/2$?

Exercice 3.3. *Tireur perfectionniste.* Un tireur touche une cible une fois sur deux.

- 1) Combien de fois doit-il tirer, au minimum, pour être sûr, à 99 % au moins, d'atteindre la cible ?
- 2) Le même tireur vise la cible 20 fois de suite. Quelle est la probabilité qu'il l'atteigne 20 fois ?
- 3) 720 000 tireurs du même acabit réalisent en même temps l'expérience précédente. Quel est l'événement le plus probable parmi : « Aucun n'atteint 20 fois la cible » et « L'un au moins atteint 20 fois la cible » ?

Épilogue : Si après tout cet entraînement les tireurs ratent encore la cible une fois sur deux, leur cas est sans doute désespéré !

Exercice 3.4. *Chuck-a-luck.* Un jeu d'argent se déroule ainsi : on parie et on mise sur un nombre entre 1 et 6. On jette ensuite trois dés. Si le nombre choisi apparaît $k \geq 1$ fois, on gagne k fois sa mise ; sinon, on perd sa mise. A-t-on intérêt à jouer ?

Variantes : Ce jeu, appelé *Chuck-a-luck* sur les foires américaines, est basé sur le même principe que le *Crown and Anchor* des Britanniques, l'*Anker en Zon* des Flamands, notre *Ancre*, *Pique*, et *Soleil*, le *Hoo Hey How* des Chinois, le *Bầu cua cá cốp* des Vietnamiens, et encore la *Big Six wheel* des Américains.

Exercice 3.5. *Élection surprenante...* *Tant que ça ? (premier épisode)* Pour mémoire, voici les résultats du trio de tête du premier tour de l'élection présidentielle de 2002 :

Jacques Chirac	19,88 %
Lionel Jospin	16,18 %
Jean-Marie Le Pen	16,86 %

- 1) Modéliser l'expérience : on choisit au hasard un électeur avant l'élection et on lui demande son intention de vote.
- 2) Modéliser l'expérience : on choisit au hasard 1 000 électeurs avant l'élection et on compte parmi eux le nombre de ceux qui ont l'intention de voter pour Lionel Jospin. On notera J la variable aléatoire associée.
- 3) Exprimer, en fonction de J , la proportion d'électeurs (parmi les 1 000) ayant l'intention de voter pour Lionel Jospin.
- 4) Pour tout $\varepsilon > 0$, on définit :

$$f(\varepsilon) = \mathbb{P}\left(0,1618 - \varepsilon \leq \frac{J}{1000} \leq 0,1618 + \varepsilon\right).$$

Montrer qu'il existe $\varepsilon_0 \in [0, 1]$ tel que

$$f(\varepsilon) \geq 95\% \iff \varepsilon \geq \varepsilon_0.$$

[On pourra montrer que f est croissante.]

On trouve que $\varepsilon_0 \approx 2,28\%$. L'institut B.V.A. a effectué un sondage sur 1 000 électeurs deux jours avant l'élection et a obtenu les résultats suivants :

Jacques Chirac	19 %
Lionel Jospin	18 %
Jean-Marie Le Pen	14 %

5) *A posteriori*, le résultat du sondage B.V.A. est-il surprenant ?

Exercice 3.6. *Machine-outil.* Une machine-outil produit des vis. En marche normale, la probabilité qu'une vis soit défectueuse est 0,1%.

- 1) Quelle est la probabilité qu'on prenne 999 vis en bon état, puis une millièème qui soit défectueuse ?
- 2) Quelle est la probabilité qu'un lot de 1 000 vis comprenne au plus une vis défectueuse ?
- 3) Comparer le résultat obtenu à celui approché donné par une loi de Poisson.
- 4) On pioche des vis au hasard. Combien va-t-on prendre de vis, en moyenne, avant de tomber sur une vis défectueuse ?

Exercice 3.7. *Un livre sans coquilles : une denrée rare.* La première édition du prix Goncourt (2011), *L'Art français de la guerre*, d'Alexis Jenni, parue chez la prestigieuse maison Gallimard, n'est pas épargnée par la présence d'une dizaine de coquilles sur le million de caractères qui composent le livre.

- 1) À combien peut-on estimer la probabilité qu'un caractère pris au hasard soit erroné ?
- 2) Quelle est la probabilité qu'il y ait au moins une coquille dans les 64 premières pages (le livre contient 640 pages).

Exercice 3.8. *Radioactivité.* On veut mesurer le nombre de particules α émises en une seconde par un gramme de matière radioactive. Des expériences ont montré qu'en moyenne ce nombre est 3,2. Quelle est la probabilité qu'au plus deux particules α seront enregistrées ?

Exercice 3.9. On gagne à un jeu avec probabilité $p \in [0, 1]$. On joue cinq fois, mais si on gagne la cinquième fois, on joue jusqu'à ce qu'on perde.

- 1) Quelle est l'espérance du nombre de fois que l'on joue ?
- 2) Quelle est l'espérance du nombre de fois où l'on perd ?

Exercice 3.10. On jette une pièce de monnaie jusqu'à obtenir pile pour la deuxième fois. Donner l'espérance et la variance du nombre de face obtenus.

Exercice 3.11. La probabilité de naissance d'un garçon dans une famille est $p \in [0, 1]$. Dans chacun des cas suivants, calculer la taille moyenne de la fratrie :

- 1) on continue jusqu'à avoir un garçon ;
- 2) on continue jusqu'à ce que les deux sexes soient représentés ;
- 3) on continue jusqu'à ce que le sexe de l'aîné réapparaisse ;
- 4) on continue jusqu'à ce que les deux sexes soient représentés, mais en s'arrêtant à quatre enfants.

Exercice 3.12. *Trousseau de clefs.* Un ami vous a prêté les clefs de son appartement pour que vous puissiez y déposer vos affaires. Arrivé devant la porte, vous ne vous souvenez plus de la clef qui ouvre la porte. Le trousseau compte n clefs.

- 1) Première stratégie : vous essayez une clef, puis une autre, jusqu'à ouverture de la porte, et ce complètement au hasard, sans vous souvenir des clefs déjà essayées. On note N_1 le nombre de clefs essayées. Donner la loi et l'espérance de N_1 .
- 2) Seconde stratégie : vous essayez une clef après l'autre, jusqu'à ouverture de la porte, et ce en éliminant après chaque essai la clef qui n'a pas convenu. On note N_2 le nombre de clefs essayées. Donner la loi et l'espérance de N_2 .

Exercice 3.13. Un échantillon de trois objets est choisi au hasard dans une boîte en contenant vingt, dont quatre sont défectueux. Trouver l'espérance du nombre d'objets défectueux dans l'échantillon.

4 Conditionnement et indépendance

Exercice 4.1. On jette deux dés.

- 1) Si vous savez que le résultat fait apparaître deux nombres distincts, quelle est la probabilité que la somme soit paire ?
- 2) Si vous savez que la somme est paire, quelle est la probabilité que les deux nombres soient distincts ?

Exercice 4.2. Une grive se pose sur un fil téléphonique. Un chasseur, en visant la grive, peut l'atteindre avec une probabilité égale à 0,5. La probabilité de couper le fil est égale à 0,2. La probabilité de réussir les deux est égale à 0,1. Un groupe de trois chasseurs tire sur la grive (sans atteindre un autre chasseur...).

- 1) Calculer la probabilité de pouvoir encore téléphoner.
- 2) Calculer la probabilité de voir la grive s'envoler.
- 3) Calculer la probabilité d'avoir le fil indemne et la grive en bonne santé.
- 4) Justifier l'indépendance des événements « le fil est coupé » et « la grive est touchée ».

Exercice 4.3. Le 14 juillet, à Saint-Tropez, il fait beau huit fois sur dix. Le comité des fêtes dispose de deux sources de prévisions météorologiques indépendantes :

- la météo nationale, qui se trompe deux fois sur cent ;
- une grenouille verte, qui se trompe une fois sur vingt.

La météo annonce de la pluie, alors que le comportement de la grenouille laisse prévoir du beau temps. Est-il plus probable qu'il fasse beau ou pas ?

Exercice 4.4. *Assurance.* Une compagnie d'assurance estime que les gens peuvent être répartis en trois classes : personnes à haut risque, à risque modéré, à bas risque. Ses statistiques montrent qu'un individu à haut risque a une probabilité de 0,3 d'avoir un accident dans l'espace d'un an ; cette probabilité tombe à 0,15 pour un individu à risque modéré et à 0,05 pour quelqu'un à bas risque. De plus, elle estime que 30% de la population appartient à la classe à haut risque et 50% à la classe à risque modéré.

- 1) Quelle est la probabilité pour qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat ?
- 2) Un nouveau signataire a un accident dans l'année qui suit la signature de son contrat. Quelle est la probabilité qu'il fasse partie de la classe à haut risque ?
- 3) Quelle est la probabilité qu'une personne n'ait pas d'accident durant la deuxième année, sachant qu'elle n'en a pas eu lors de la première ?
- 4) Vérifier que cette probabilité est plus grande que celle de ne pas avoir d'accident la première année. Donner une explication intuitive de ce résultat.

Exercice 4.5. Soit A , B et C trois événements indépendants. Montrer que :

- 1) A et B^c sont indépendants ;
- 2) A et $B \cup C$ sont indépendants.

Exercice 4.6. Grand-père a trois bérets et une casquette. Quand il descend acheter sa baguette, il saisit un couvre-chef au hasard. Sachant qu'il prend une fois sur trois une baguette moulée et que deux fois sur cinq il oublie de chausser ses souliers, calculer la probabilité de le voir remonter en charentaises, un béret sur la tête et une baguette non moulée sous le bras.

Exercice 4.7. La mafia subtilise 15 % des colis expédiés de New York par avion. Dominique veut envoyer deux cadeaux à sa sœur. Elle peut faire soit deux paquets séparés, soit un paquet groupé. Calculer, dans les deux cas, les probabilités des événements suivants : « un cadeau au moins est bien arrivé » et « les deux cadeaux sont bien arrivés ».

Exercice 4.8. Une rue compte vingt lampadaires. On modélise la durée de vie en mois de chacune des lampes par une variable aléatoire de loi $\mathcal{G}(p)$ avec $p \in]0, 1[$. Montrer que la durée pendant laquelle tous les lampadaires fonctionnent suit une loi géométrique dont on précisera le paramètre.

Exercice 4.9. Le capitaine Flint, croyant son équipage endormi, enterre son coffre de doublons sur la falaise de l'île déserte, puis repère son emplacement en y traînant une grosse pierre en forme de tortue. Malheureusement, le autres l'ont suivi et ont observé son manège. Dans la nuit, chacun des flibustiers revient, à l'insu des autres, et déplace la pierre de sept pieds, soit vers le nord, soit vers le sud, avec la même probabilité. On note X la variable aléatoire égale au nombre de pieds dont la pierre se trouve déplacée après le passage de n pirates (positif vers le nord, négatif vers le sud).

- 1) Calculer $\mathbb{E}[X]$ et $\text{Var}(X)$.
- 2) En discutant selon la parité de l'entier n , calculer la probabilité pour que la pierre ne se retrouver pas écartée de plus de sept pieds de l'emplacement du coffre.

Exercice 4.10. Chaque samedi, Georges fait son marché. Selon l'envie, il décide d'acheter 1, 2 ou 3 concombres avec les probabilités respectives $1/6$, $1/2$ et $1/3$. Il découvre ensuite le prix du concombre, qui varie selon les semaines, et, en été, il vaut 0,90 €, 1,00 € ou 1,10 € avec les probabilités respectives $1/4$, $1/2$ et $1/4$. En moyenne, combien Georges dépense-t-il en concombre chaque semaine ?

Exercice 4.11. Soit $(X_1, X_2, \dots, X_n)$ des variables indépendantes et admettant un second moment.

- 1) Démontrer que $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$.
- 2) Démontrer que

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n).$$

Exercice 4.12. Le nombre de blessés à l'œil par un bouchon de champagne arrivant aux urgences ophtalmologiques la nuit de la Saint-Sylvestre suit une loi de Poisson de moyenne 6. Sachant que les victimes potentielles sont réparties équitablement entre les deux sexes, on note X le nombre de femmes qui figurent parmi ces éborgnés du Jour de l'An et Y le nombre d'hommes.

- 1) Etablir les lois de probabilité de X et de Y .
- 2) Calculer leurs espérances et leurs écarts-types.
- 3) Montrer que X et Y sont indépendantes.

[On suppose que l'interne de garde ne fait pas partie des victimes...]

Exercice 4.13. *Le premier arrivé a gagné!* « Je vais lancer deux dés jusqu'à ce que la somme des deux dés fasse 7 ou 8. Si 8 sort en premier, vous gagnez ; sinon, si c'est 7, c'est moi qui gagne. Quelle est la probabilité que vous gagniez ? » Essayons de répondre à la question en quelques étapes.

- 1) Modéliser un lancer de deux dés.
- 2) Quelle est la probabilité que la somme obtenue fasse 7 ? Et 8 ?

On s'intéresse maintenant au jeu proposé. On imagine qu'on lance les dés une infinité de fois. On modélise la liste des sommes observées par une variable aléatoire $X = (S_n)_{n \geq 1}$.

- 3) Préciser l'univers de la variable aléatoire X et sa loi.

On note $G \subset \Omega_X$ l'événement « Vous gagnez », de sorte qu'on cherche $\mathbb{P}(X \in G)$.

- 4) Justifier que

$$\mathbb{P}(X \in G) = \mathbb{P}(X \in G | S_1 = 7) \mathbb{P}(S_1 = 7) + \mathbb{P}(X \in G | S_1 = 8) \mathbb{P}(S_1 = 8) + \mathbb{P}(X \in G | S_1 \notin \{7, 8\}) \mathbb{P}(S_1 \notin \{7, 8\})$$

- 5) En déduire $\mathbb{P}(X \in G)$.
- 6) Quelle est la probabilité que je gagne ?
- 7) Quelle est la probabilité qu'aucun de nous ne gagne ?
- 8) Retrouver ce dernier résultat en remarquant que l'événement « Aucun de nous ne gagne » est égal à l'événement « 7 et 8 n'apparaissent jamais ».

Pour tout $n \geq 1$, on note E_n l'événement « Durant les $n - 1$ premiers lancers, ni 7 ni 8 ne sortent et, au n^{e} lancer, 8 sort ».

- 9) Pour tout $n \geq 1$, calculer $\mathbb{P}(X \in E_n)$.
- 10) En déduire $\mathbb{P}(X \in G)$ et comparer au résultat déjà obtenu.

5 Estimation

Exercice 5.1. Le taux de personnes à soigner pour un problème de cholestérol élevé est de 7,5%.

- 1) Etant données 10 000 personnes, donner un intervalle dans lequel on est sûr à 95% de trouver le nombre de personnes qu'il y aura à soigner.
- 2) L'intervalle précédent est-il un intervalle de confiance ?

Exercice 5.2. Le pôle médical d'une grande entreprise mène une enquête sur le taux de cholestérol de ses employés ; les observations sur 100 employés tirés au sort sont les suivantes :

taux de cholestérol (en cg)	120	160	200	240	280	320
nombre d'employés	9	22	25	21	16	7

- 1) Proposer un modèle statistique pour cette expérience.
- 2) Calculer la moyenne et l'écart-type du taux de cholestérol sur l'échantillon.
- 3) Estimer la moyenne et l'écart-type du taux de cholestérol dans toute l'entreprise.
- 4) Déterminer un intervalle de confiance asymptotique à 95% pour le taux de cholestérol moyen dans l'entreprise.
- 5) Déterminer la taille minimum d'un échantillon pour que l'amplitude de l'intervalle de confiance asymptotique à 95% soit inférieure à 10.

Exercice 5.3. *Estimation de viscosité.* Une entreprise chimique commercialise un polymère servant à la fabrication de microprocesseurs. Ce polymère est stocké dans une cuve où l'on contrôle sa viscosité η ; celle-ci doit être comprise entre 75 et 95 pour que le polymère puisse être mis sur le marché. Quatre extractions ont été réalisées dans des zones différentes de la cuve et ont conduit aux valeurs de la viscosité suivantes : $x_1 = 78$, $x_2 = 85$, $x_3 = 91$, $x_4 = 76$. Aidez l'entreprise à fournir aux clients un intervalle de confiance asymptotique à 95% pour η .

Exercice 5.4. *Taux de germination.* Un semencier a récolté cinq tonnes de graines de tournesol. Il a besoin de connaître le taux de germination p de ces graines avant de les mettre en vente. Il extrait un échantillon de 40 graines, les dépose sur un buvard humide et compte le nombre de graines ayant évolué favorablement : il en trouve 36.

- 1) Proposer un modèle statistique pour cette expérience.
- 2) Quel est l'estimateur du maximum de vraisemblance pour p ?
- 3) Cet estimateur est-il un estimateur sans biais de p ?
- 4) Dans le modèle d'échantillonnage associé, la suite des estimateurs du maximum de vraisemblance pour p est-elle une suite d'estimateurs consistante de p ?
- 5) Application numérique : donner une estimation de p .
- 6) Proposer un intervalle de confiance à 80% pour p .
- 7) Application numérique : donner l'intervalle de confiance estimé.
- 8) Proposer un intervalle de confiance asymptotique à 95% pour p .
- 9) Application numérique : donner l'intervalle de confiance asymptotique estimé.

Remarque : On remarque que ce contrôle est de type destructif : l'échantillon ayant servi au contrôle ne peut plus être commercialisé. Même avec une population d'effectif restreint, un contrôle destructif impose de faire confiance à un échantillon restreint et la valeur exacte de p ne peut être calculée.

Exercice 5.5. *Taux d'albinisme.* Sur 12 000 individus d'une espèce, on a dénombré 13 albinos.

- 1) Proposer un modèle statistique et un modèle d'échantillonnage pour cette expérience.
- 2) Estimer la proportion d'albinos dans l'espèce en donnant un intervalle de confiance (exact) à 95%.
- 3) Estimer la proportion d'albinos dans l'espèce en donnant un intervalle de confiance asymptotique à 95%.

Exercice 5.6. *Élection surprenante... Tant que ça ? (deuxième épisode)* L'institut B.V.A. a effectué un sondage sur 1000 électeurs deux jours avant l'élection présidentielle de 2002 et a obtenu les résultats suivants :

Jacques Chirac	19 %
Lionel Jospin	18 %
Jean-Marie Le Pen	14 %

Donner les « fourchettes » correspondantes, autrement dit les intervalles de confiance asymptotiques à 95% pour chacune des proportions.

Remarque : Pour mémoire, voici les résultats du trio de tête du premier tour de l'élection présidentielle de 2002 :

Jacques Chirac	19,88 %
Lionel Jospin	16,18 %
Jean-Marie Le Pen	16,86 %

A posteriori, le résultat du sondage B.V.A. est-il surprenant ?

Exercice 5.7. Service client. Afin de mieux satisfaire ses clients, une grande société cherche à estimer le nombre λ d'appels reçus par le son service client par minute ; elle pourra ainsi évaluer le temps d'attente pour le client et le nombre d'employés à mettre au standard. Les résultats de l'enquête portent sur 200 séquences d'une minute, durant lesquelles le nombre d'appels moyen a été de 3 par minute.

- 1) Proposer un modèle statistique pour cette expérience.
- 2) Quel est l'estimateur du maximum de vraisemblance pour λ ?
- 3) Cet estimateur est-il un estimateur sans biais de λ ?
- 4) Dans le modèle d'échantillonnage associé, la suite des estimateurs du maximum de vraisemblance pour λ est-elle une suite d'estimateurs consistante de λ ?
- 5) Application numérique : donner une estimation de λ .
- 6) Proposer un intervalle de confiance asymptotique à 95% pour λ .
- 7) Application numérique : donner l'intervalle de confiance asymptotique estimé.

Exercice 5.8. Problème des chars allemands Pendant la Deuxième guerre mondiale, les Alliés se sont posés la question cruciale de l'estimation du nombre de chars allemands en circulation. Les mathématiciens de l'*Intelligence Service* ont alors eu recours à une simple méthode probabiliste qui a surpassé toutes les autres approches.

Imaginez que vous soyez sur le champ de bataille, et que vous parveniez à détruire trois chars d'une division ennemie, chaque char étant numéroté. En vous approchant de leurs carcasses, vous repérez les trois numéros de série : $x_1 = 17$, $x_2 = 66$ et $x_3 = 42$. Combien y avait-il de chars au total dans cette division ?

- 1) Proposez un modèle statistique et un modèle d'échantillonnage pour cette expérience aléatoire qu'est votre relevé.
- 2) Proposez des estimateurs du nombre r de chars (méthode des moments, méthode du maximum de vraisemblance...).
- 3) Ces estimateurs sont-ils sans biais ? Les suites associées dans le modèle d'échantillonnage sont-elles consistantes ?
- 4) Application numérique : donner une estimation de r .
- 5) Proposez un intervalle de confiance à $1 - \alpha$ pour r de la forme $[M, \beta M]$ avec

$$M = \max(X_1, X_2, X_3)$$

- 6) Application numérique : donner l'intervalle de confiance à 90% estimé.

Remarque : Les mathématiciens de l'*Intelligence Service* estimèrent à 270 le nombre de chars Panther qui avaient été produits au mois de février 1944. C'était bien plus que ce que les Alliés avaient supposé jusque là. Les archives allemandes retrouvées après-guerre montrèrent que le chiffre réel était de 276 chars Panther fabriqués ce mois-là. L'approche statistique s'était montré bien plus précise que les méthodes d'investigation classiques des services de renseignement, et l'expression « le problème des chars allemands » se répandit rapidement comme le symbole de ce genre d'analyste statistique.

6 Tests

Exercice 6.1. *Doué en calcul mental ?* Il paraît que Maximilien est un petit prodige du calcul mental (aux dires de son père). Il saurait effectuer n'importe quelle division d'un nombre de neuf chiffres par un autre, en donnant du tac au tac les dix premiers chiffres significatifs du résultat. Pour savoir ce qu'il en est, je vais faire l'expérience suivante : demander le résultat de

$$\frac{123456789}{987654321}$$

à Maximilien et noter ce qu'il me répond du tac au tac.

- 1) En supposant que le père de Maximilien dit vrai (hypothèse notée H_0), modéliser le résultat de l'expérience ci-dessus.
- 2) On construit un test de H_0 en prenant comme zone de rejet $R_0 = \mathbb{R} \setminus \{0.1249999988\}$. Quel est le niveau d'erreur d'un tel test ?

Remarque : Si on pouvait toujours se ramener à de tels tests, on n'aurait plus besoin de probabilités pour décider...

Exercice 6.2. *Galles du hêtre (deuxième épisode).* La cécidomyie du hêtre provoque des galles (petites boules) sur les feuilles de cet arbre. On observe 375 feuilles de hêtre et on note, pour chacune d'elles, le nombre de galles qu'elle porte. On obtient les résultats suivants :

nombre de galles	0	1	2	3	4	5	6	7	8	9	10
nombre de feuilles	182	98	46	28	12	5	2	1	0	1	0

On souhaite modéliser le nombre de galles sur une feuille de hêtre. Plus précisément, on veut modéliser le résultat de l'expérience suivante : on choisit au hasard une feuille sur un hêtre et on observe le nombre de galles sur cette feuille. On peut modéliser le résultat de cette expérience (le nombre de galles observées sur la feuille) par une variable aléatoire $X : \Omega \rightarrow \Omega_X = \mathbb{N}$. On propose comme loi de X la loi suivante :

x_i	0	1	2	...	9	10
$\mathbb{P}(X = x_i)$	1/2	1/4	1/8	...	1/2 ¹⁰	1/2 ¹⁰

Construire un test d'adéquation du χ^2 pour déterminer, avec un niveau d'erreur de 5%, si la loi proposée est raisonnable.

Exercice 6.3. *Théorie de Mendel.* Le croisement d'une variété américaine et d'une variété éthiopienne de sorgho donne des graines toutes rouges. Ces graines rouges, croisées entre elles, donnent : 195 rouges, 73 jaunes et 100 blanches. La théorie de Mendel prévoit les proportions 9/16, 3/16 et 4/16. Construire un test d'adéquation du χ^2 pour déterminer, avec un niveau d'erreur de 10%, si la théorie de Mendel est confirmée. Quelle est la conclusion du test ? [Indication : $\frac{144}{207} + \frac{16}{69} + \frac{64}{92} \approx 1,62$.]

Exercice 6.4. *Contrôle qualité* Un commerçant d'articles de chasse reçoit de son fournisseur un lot important de cartouches. Le fournisseur affirme que celui-ci contient une proportion p de cartouches défectueuses inférieure à 10%. Le commerçant souhaite vérifier les dires du fournisseur et la seule façon de contrôler ces cartouches est de les tirer. On comprend que le commerçant ne peut contrôler tout le lot (et même quand le contrôle n'est pas, comme ici, destructif, il est souvent cependant onéreux, ce qui explique qu'un contrôle exhaustif de la qualité est généralement exclu). Ainsi, le commerçant décide de prélever un échantillon aléatoire de 100 cartouches et de les tirer : il trouve 14 cartouches défectueuses.

1) Doit-il renvoyer tout le lot à son fournisseur ? [On pourra proposer un modèle statistique pour cette expérience, puis construire un intervalle de confiance à 80% pour p à l'aide de l'inégalité de Bienaymé-Tchebychev.]

On se propose d'améliorer la conclusion.

2) En supposant $p = 10\%$, quelle est la loi du nombre D de cartouches défectueuses dans un échantillon de 100 cartouches ? Quelle est la probabilité que $D \geq 14$?

3) Construire une zone de rejet de l'hypothèse $p = 10\%$ au niveau d'erreur 20% et de la forme $[[n_0, 100]]$, où n_0 est à déterminer.

4) Conclure.

Exercice 6.5. *Taste-tastevin.* Les tastevins utilisent souvent ce test triangulaire : deux verres contiennent le même vin, le troisième un autre. Le tastevin doit isoler celui qui n'est que dans un verre. Un béotien a la probabilité $1/3$ de désigner le bon verre.

1) Modéliser le résultat de l'expérience : une personne fait 10 tentatives et on compte le nombre de bonnes réponses qu'elle donne.

Un soi-disant connaisseur se présente. On souhaite tester ses performances en lui faisant faire l'expérience décrite ci-dessus.

2) Donner une zone de rejet de l'hypothèse : « Le soi-disant connaisseur est un béotien » de la forme $R_0 = \llbracket k, 10 \rrbracket$ et au niveau d'erreur de 5%.

3) Notre homme désigne 7 fois le bon verre. Béotien ou connaisseur ?

Pour être sûr, on recommence l'expérience (des 10 essais) et notre homme désigne cette fois seulement 4 fois le bon verre.

4) En tenant compte seulement de cette seconde expérience, le déclarez-vous béotien ou connaisseur ?

5) En tenant compte des 20 essais, dites-vous qu'il est béotien ou connaisseur ?

6) Avez-vous une préférence entre les conclusions 4 et 5 ?

Remarque : Le pauvre ne doit plus avoir les idées très claires après toute cette séance de dégustation ! C'est d'ailleurs une contrainte pour la réalisation de ce test : on doit pouvoir s'assurer de l'expertise du tastevin en lui faisant boire le minimum de verres possibles.

Exercice 6.6. *Pêcheur doué ou chanceux ?* Trois pêcheurs ont attrapé six poissons. Parmi eux, Saturnin en a pêché quatre : il parle de don, ses amis de chance.

1) On suppose que Saturnin a eu de la chance, autrement dit que les trois pêcheurs ont le même niveau. Modéliser l'expérience : « les trois amis attrapent six poissons ». [Indication : on pourra faire un arbre représentant le fait que le premier poisson a autant de chance d'être attrapé par chacun des trois amis, et de même pour le deuxième, etc.]

2) Quelle est alors la probabilité que Saturnin attrape au moins quatre poissons ?

3) Quelle est la valeur minimale de k pour laquelle la probabilité que Saturnin attrape au moins k poissons est inférieure ou égale à 20% ?

4) Saturnin a-t-il eu de la chance (avec un niveau d'erreur de 20%) ?

5) Quelle est la probabilité que l'un des trois amis attrape au moins quatre poissons ?

6) Quelle est la valeur minimale de k pour laquelle la probabilité que l'un des trois amis attrape au moins k poissons est inférieure ou égale à 20% ?

7) Cela remet-il en cause la conclusion de la question 4 ?

Indications : $233/729 \approx 0,32$, $73/729 \approx 0,10$ et $13/243 \approx 0,05$.

Exercice 6.7. *Truites dans un étang.* On souhaite estimer le nombre de truites dans un étang. Des écologistes semblent dire qu'il ne resterait plus que 8400 truites. Pour tester cette hypothèse, étant donné qu'on n'a pas les moyens de vider l'étang et de recenser exactement toutes les truites qui s'y trouvent (les écologistes ne seraient d'ailleurs sans doute pas d'accord avec cette méthode !), on décide de prodéder ainsi : on prélève 1000 truites, on les marque et on les relâche ; une semaine après, on prélève à nouveau 1000 truites dans l'étang et on compte le nombre de truites marquées.

1) Si l'on note x ce dernier nombre, quel est approximativement, et intuitivement, le nombre de truites dans l'étang (en fonction de x) ?

2) Proposer un modèle statistique pour l'expérience : une semaine après le marquage des truites, on prélève 1000 truites dans l'étang et on compte le nombre X de truites marquées.

3) On suppose que les écologistes ont raison. Trouver le nombre x_0 minimal tel que

$$\mathbb{P}(X \leq x_0) \geq 5\%.$$

4) On fait l'expérience et on relève 100 truites marquées. Les écologistes ont-ils raison ?

5) Que donne l'estimation de la question 1 ?