LU3IN005

Statistiques & informatique

2024-2025

Fascicule de TDs

Nicolas Baskiotis
Pierre-Henri Wuillemin

semaine 1

Un certain nombre d'exercices sont tirés du cours de A. Guyader `http://www.lsta.upmc.fr/guyader/`

Exercice 1 – Dénombrement 1

Q 1.1 On lance trois fois de suite un dé.

Q 1.1.1 Quel est l'espace de probabilité lié à cette expérience ?

Q 1.1.2 On s'intéresse au nombre de valeurs distinctes obtenues. Quelle est la distribution de probabilité ?

Q 1.2 Nicolas et Pierre-Henri ont l'habitude de se retrouver chaque semaine autour d'un café et de décider à pile ou face qui règle l'addition. Nicolas se lamente d'avoir payé les quatre dernières additions et Pierre-Henri, pour faire plaisir à son collègue, propose de modifier exceptionnellement la règle : “Nicolas, tu vas lancer la pièce cinq fois et tu ne paieras que si on observe une suite d'au moins trois piles consécutifs ou d'au moins trois faces consécutives". Nicolas se félicite d'avoir un collègue si conciliant. A tort ou à raison ?

Q 1.3 Supposons qu'une file de personnes est composée de h hommes et f femmes.

Q 1.3.1 Quelle est la probabilité que la personne au i-ème rang soit un homme si la file est constituée au hasard ?

Q 1.3.2 Combien de files sont possibles si on considère chaque individu comme unique ?

Q 1.3.3 Et si on s'intéresse uniquement à l'ordre homme/femme dans la file ?

Q 1.3.4 On introduit c chiens et on s'intéresse à la suite h, f, c. Combien de files ?

Q 1.4 9 personnes doivent s'aligner, parmi eux 4 français, 3 italiens, 2 allemands. Combien de manières de faire si les français doivent être tous à gauche et les italiens à droite ? Si les nationalités doivent rester grouper ?

Q 1.5 Une urne contient 3 boules numérotées $1, 2, 3$. On effectue 5 tirages avec remise, on note le nombre de fois où chaque boule est apparue.

Q 1.5.1 Nombre de résultats possibles ?

Q 1.5.2 Nombre de résultats sachant que la boule 2 n'est pas apparue.

Q 1.5.3 Nombre de résultats sachant que chaque boule est apparue au moins une fois.

Q 1.6 On lance une pièce 5 fois et on note dans l'ordre l'apparition de pile ou face.

Q 1.6.1 Combien de suites différentes ? combien de suites avec 2 pile ?

Q 1.6.2 Combien de suites avec au moins deux pile ?

Q 1.6.3 Combien de suites avec au moins un pile et un face ?

Q 1.7 Combien d'anagrammes des mots : *Flute*, *Proposer*, *Taratata* ?

Q 1.8 Le père-noël dispose de 10 pochettes cadeaux et 14 jouets tous différents. Il veut préparer 6 cadeaux. Combien de résultats possibles ?

Q 1.9 8 écrans identiques sont à partager entre 4 bureaux différents, combien de possibilités ? Et si chaque bureau doit recevoir au moins 1 écran ?

Exercice 2 – Traduction

Soit Ω un univers et trois événements A, B, C sur cette univers. Traduire grâce aux symboles d'union, intersection et complémentaire les phrases suivantes :

- Seul A se réalise
- A et C se réalisent, pas B
- Au moins un des 3 événements se réalise
- Les 3 événements se réalisent
- Aucun ne se réalise
- Au plus l'un des 3 se réalise
- Exactement deux des trois se réalisent
- Au plus trois se réalisent.

Exercice 3 – Dates d'anniversaire

Dans cet exercice, on suppose que la date d'anniversaire d'une personne (jour et mois) est distribuée uniformément sur l'ensemble des dates possibles, en ignorant les années bisextiles. Ainsi, une personne tirée aléatoirement a une probabilité $\frac{1}{365}$ d'être née un jour particulier de l'année.
n personnes se retrouvent à une soirée ($n \geq 2$).

Q 3.1 À partir de quelle valeur de n est-on sûr qu'il y a deux personnes nées le même jour ?

Q 3.2 Soit n, un entier compris entre 2 et 365. Quelle est la probabilité que les n personnes soient toutes nées des jours différents ?

Q 3.3 Donnez une formule de récurrence simple permettant de calculer p_n à partir de p_{n-1}.

Q 3.4 À partir de quelle valeur de n a-t'on une probabilité supérieure à $\frac{1}{2}$ que au moins deux personnes soient nées le même jour ?

Exercice 4 – Dénombrements avancés

Q 4.1 Soit $\mathcal{A} = \{1, 2, 3\}$ et $\mathcal{B} = \{1, 2, 3, 4, 5\}$. Combien d'applications de $\mathcal{A}$ dans $\mathcal{B}$? d'applications injectives ? d'applications injectives telles qu 4 n'admet qu'un antécédent ?

Q 4.2 Soit une application d'un ensemble à k éléments vers un ensemble à n éléments. Quelle est la probabilité qu'elle soit injective ?

Q 4.3 Combien de suites strictement croissantes de 3 chiffres pris dans $\{1, 2, 3, 4, 5\}$? décroissantes ?

Q 4.4 Sur $\mathbb{N} \times \mathbb{N}$, on note $\mathcal{E}$ l'ensemble des chemins de longueur 12 qui partent de $(0, 0)$ et terminent en $(7, 5)$. Combien de chemins ? Et de chemins qui passent par $(4, 3)$? (un chemin est une suite de segments connectés de longueur 1)

Q 4.5 On considère dans cet exercice un polygone convexe à n sommets. On s'intéresse à l'expérience aléatoire consistant à tirer arbitrairement deux segments reliant des sommets tous distincts.

Q 4.5.1 Définir l'univers des possibles et déterminer son cardinal.

Q 4.5.2 Déterminer le cardinal de l'ensemble des cas où les deux segments s'intersectent. En déduire la probabilité de l'événement aléatoire "les deux segments s'intersectent".

Q 4.6 Un ascenseur commence au RDC avec 8 personnes et va jusqu'au 6ème étage en s'arrêtant entre temps. On note x_i le nombre de personnes qui sont descendus au i-ème étage, et on suppose qu'il n'y a plus personne dans l'ascenceur à la fin. Combien de n-uplet $(x_1, \ldots, x_6)$ sont possibles ? Et s'il y avait 5 femmes et 3 hommes et qu'on s'intéresse aux variables f_i, h_i le nombre de femmes et d'hommes descendus à chaque étage ?

Exercice 5 – Espace probabilisé infini

Q 5.1 On lance une pièce équiprobable un nombre indéterminé de fois.

Q 5.1.1 Probabilité d'obtenir le premier pile après le 4ème lancer (préciser l'univers) ?

Q 5.1.2 D'attendre au moins 4 lancers avant le premier pile ?

Q 5.1.3 D'obtenir le deuxième face avant le 5ème lancer ?

Q 5.1.4 D'obtenir le premier pile à un rang impair ? d'obtenir le deuxième face à un rang impair ? (on admettra $\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$ si $0 < x < 1$).

semaine 2

Exercice 6 – Histoires de portables

Q 6.1 Le gérant d'un magasin d'informatique a reçu un lot de téléphone chacun dans sa boîte. 5% des boîtes sont abîmées. Le gérant estime que :

- 60% des boîtes abîmées contiennent un téléphone défectueux.
- 98% des boîtes non abîmées ne contiennent aucun téléphone défectueux.

Un client achète une boite. On désigne par A l'événement : "la boîte est abîmée" et par D l'événement "la boîte achetée contient un téléphone défectueux".

Q 6.1.1 Donner les probabilités de $P(A)$, $P(\overline{A})$, $P(D|A)$, $P(D|\overline{A})$, $P(\overline{D}|A)$, $P(\overline{D}|\overline{A})$. En déduire $P(D)$.

Q 6.1.2 Le client constate qu'un téléphone acheté est défectueux. Quelle est la probabilité pour qu'il ait acheté une boîte abîmée ?

Q 6.2 En sortant de l'université, vous vous rendez compte que vous n'avez plus votre nouveau téléphone portable : vous l'avez oublié en amphi avec une probabilité de 20% et dans ce cas vous avez 90% de chance de le retrouver ; il est possible également que vous l'ayez perdu en chemin, mais dans ce cas vous n'avez qu'une chance sur quatre de le retrouver

Q 6.2.1 Calculer la probabilité de trouver votre portable dans le cas où vous le cherchez en amphi. Et dans le cas où vous le cherchez sur le chemin ? Ou vaut-il mieux l'avoir perdu ?

Q 6.2.2 Vous avez cherché le portable sur le chemin, vous ne l'avez pas trouvé. Quelle est à présent la probabilité de le trouver en amphi ?

Q 6.2.3 Vous trouvez un portable en amphi. Quelle est la probabilité que ce soit le votre ?

Exercice 7 – Loi Totale et autres

Q 7.1 Soit x la proportion de prestidigitateurs dans la population. Nicolas joue à pile ou face avec Pierre-Henri. Ce dernier parie sur pile, lance la pièce, et obtient pile. Quelle est la probabilité pour qu'il soit un prestidigitateur ?

Q 7.2 Une population est composée de familles de 0, 1, 2 ou 3 enfants. Il y a une famille sans enfant pour 3 de 1 enfant, 4 de 2 enfants et 2 de 3 enfants. On suppose que les deux sexes sont équiprobables et qu'ils sont indépendants pour deux enfants différents.

Q 7.2.1 Donner les probabilités de nombres d'enfants par famille p_0, p_1, p_2, p_3.

Q 7.2.2 On choisit une famille au hasard : quelle est la probabilité qu'il n'y ait aucun garçon ?

Q 7.2.3 Toujours pour une famille choisie au hasard, quelle est la probabilité qu'elle ait 2 enfants sachant qu'elle n'a aucun garçon ?

Q 7.3 Vous voyagez en avion de Los Angeles à Paris avec deux escales, à New York puis Londres. La probabilité p que votre bagage ne soit pas mis en soute est la même à chaque étape. Arrivé à Paris, vous n'avez pas votre valise. Calculer les probabilités que celle-ci soit restée à Los Angeles, New York et Londres respectivement.

Q 7.4 Un questionnaire à choix multiples propose m réponses pour chaque question. Soit p la probabilité qu'un étudiant connaisse la bonne réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quelle est pour le correcteur la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il l'a donnée ?

Exercice 8 – Indépendance

Q 8.1 Une boîte contient des boules numérotées de 1 à 10. Une est tirée au hasard, soit X_1 ce nombre. On remet la boule dans la boîte et un deuxième tirage est effectué, noté X_2. Quelle est la distribution de X_1, X_2 ? Les deux sont-ils indépendants ? Et dans le cas où il n'y a pas remise ?

Q 8.2 On tire trois fois à pile ou face. Soit les événements :

- A : Pile au premier lancé
- B : Face au second
- C : Pile au troisième
- D : les trois lancés sont tous pile ou tous face
- E : exactement un pile dans les trois lancés.

Q 8.2.1 Quelles paires d'événements sont indépendants parmi (A,B), (A,D), (A,E), (D,E) ?

Q 8.2.2 Quels triplets d'événements sont indépendants parmi (A,B,C), (A,B,D), (C,D,E) ?

Q 8.3 Montrer que pour G et H deux événements indépendants, alors G et $\overline{H}$ sont indépendants ainsi que $\overline{G}$ et $\overline{H}$.

Exercice 9 – Risque conditionnel

Une compagnie d'assurance répartit ses clients en deux classes R_1 et R_2 : ceux qui sont enclins aux accidents et ceux à faible risque. Ses statistiques montrent qu'un individu enclin aux accidents a une probabilité de 0.4 d'en avoir un au cours d'une année, alors que cette probabilité tombe à 0.2 pour les gens à faible risque. On suppose que 30% de la population appartient à la classe à haut risque.

Q 9.1 Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant la première année de son contrat ?

Q 9.2 Quelle est la probabilité qu'il fasse partie de la classe à haut risque s'il est victime d'un accident dans l'année qui suit la conclusion de son contrat ?

Q 9.3 Quelle est la probabilité conditionnelle pour un nouveau client d'avoir un accident durant la deuxième année de son contrat s'il a eu un accident durant la première année ?

Exercice 10 – Paradoxes

Q 10.1 Grincheux, Simplet et Joyeux partent en picnic et Blanche-Neige leur prépare leur panier. Elle ne dispose cependant que de 2 pommes (et pas de couteau), elle décide donc de mettre les pommes dans deux paniers tirés aléatoirement. Elle leur demande de ne pas ouvrir leur panier avant l'heure du déjeuner pour ne pas leur gâcher la journée . . . Mais voilà que Prof - qui a vu les préparatifs - propose à Simplet de lui dire qui de Grincheux ou Joyeux a une pomme ; si les deux ont la pomme, il lui dira au hasard un des deux noms. Simplet décline l'offre en disant que cela réduira ses chances d'avoir la pomme à une sur deux, soit lui soit l'autre nain aura la pomme restante. Quelle erreur fait Simplet ? Quelle est la probabilité qu'il ait la pomme sachant que Grincheux à la pomme ?

Q 10.2 Simplet et Prof jouent à pile ou face.

Q 10.2.1 Simplet parie que la séquence *pile, pile, face* arrivera avant *face, face, pile*. Prof refuse le pari, en prétextant que cela revient à tirer une fois à pile ou face. A-t-il raison ?

Q 10.2.2 Cette fois Simplet parie que la séquence *pile, face, face* arrivera avant *pile, pile, face*. Prof acceptera-t-il le pari ?

Q 10.3 Dans un parc il y a trois bancs à 2 places. Blanche-Neige puis Simplet vont s'asseoir au hasard. Simplet pense ainsi qu'il a une chance sur 3 de se trouver sur le même banc. Blanche-Neige lui dit que c'est plutôt une chance sur 5. Qui a raison et pourquoi ?

Exercice 11 – Conditionnel premier

On suppose qu'on a un espace probabilisé tel que l'univers Ω est un ensemble fini de cardinal un nombre premier p, et que le modèle choisi soit celui de l'équiprobabilité. Prouver que deux événements A et B non triviaux ne peuvent pas être indépendants.

Exercice 12 – Loi de succession de Laplace

On dispose de $(N+1)$ urnes, numérotées de 0 à N. L'urne k contient k boules rouges et $(N-k)$ boules blanches. On choisit une urne au hasard. Sans connaître son numéro, on en tire n fois de suite une boule, avec remise après chaque tirage.

Q 12.0.1 Quelle est la probabilité que le tirage suivant donne encore une boule rouge sachant que, au cours des n premiers tirages, seules des boules rouges ont été tirées ?

Q 12.0.2 Calculer la limite de cette probabilité lorsque N tend vers l'infini (rappel : si f est continue sur $[0,1]$, alors $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(k/n) = \int_0^1 f(x)dx$).

semaine 3

Exercice 13 – Calendrier aléatoire

On lance un dé à 12 faces où chaque face correspond à un mois de l'année. Soit M le mois obtenu et X le nombre de jours du mois obtenu (année non bissextile).

Q 13.1 Quelle est le nom mathématique de M et X ? Sa définition et son utilité ?

Q 13.2 Donner la loi de X en fonction de la loi jointe $P(X, M)$ et représenter sa fonction de répartition.

Q 13.3 Calculer l'espérance de X et son écart-type.

Q 13.4 Supposons qu'on tire maintenant au hasard un jour de l'année et qu'on appelle Y le nombre de jours du mois correspondant à ce tirage. Quelle est sa loi ? Son espérance ?

Exercice 14 (3.5 points) – Loi jointe

Soit X et Y deux variables aléatoires qui suivent des lois de Bernoulli, chacune de paramètre 0.5. De plus, on suppose que $P(X = 1, Y = 1) = \frac{1}{12}$.

Q 14.1 Donner la loi jointe de (X, Y).

Q 14.2 Quelle est la loi de $Z = XY$? Quelle est son paramètre ?

Q 14.3 Calculer la covariance $Cov(X, Y)$. X et Y sont elles indépendantes ?

Exercice 15 – Loi binomiale

A et B sont deux avions ayant respectivement 4 et 2 moteurs. Les moteurs sont supposés indépendants les uns des autres, et ils ont une probabilité p de tomber en panne. Chaque avion arrive à destination si moins de la moitié (incluse) de ses moteurs tombe en panne.

Q 15.1 Soient $X_4 \in \{0, 1, 2, 3, 4\}$ et $X_2 \in \{0, 1, 2\}$ les v.a. correspondant respectivement au nombre de moteurs en panne des avions A et B. Quelles sont les lois de X_4 et X_2 ?

Q 15.2 Donnez les probabilités que chaque avion arrive à destination.

Q 15.3 Pour quelles valeurs de p faut-il choisir de voyager avec l'avion B ?

Exercice 16 – Queue de poisson

Q 16.1 Dans un bureau de poste, il y a deux guichets. Chacune des personnes arrivant à la poste choisit le premier guichet avec une probabilité p, ou le deuxième guichet avec une probabilité $1 - p$. Les personnes effectuent leur choix de façon indépendante. En une heure, le nombre X de personnes arrivées à la poste suit une loi de Poisson de paramètre λ. On désigne par Y le nombre de personnes ayant choisi le premier guichet.

Q 16.1.1 Exprimez la probabilité conditionnelle de $Y = k$ sachant que $X = n$. En déduire la loi conjointe du couple (X, Y).

Q 16.1.2 Montrez que Y suit une loi de Poisson de paramètre λp.

Exercice 17 – Deux dés

On lance deux dés équilibrés. On note X_1 et X_2 les variables aléatoires correspondant aux résultats obtenus.

Q 17.1 Rappeler la loi de X_1, son espérance et sa variance.

Q 17.2 Soit $X = X_1 + X_2$ et $Y = X_1 - X_2$. Que valent $\mathbb{E}[X]$, $\mathbb{E}[Y]$? Montrer que $\mathbb{E}[XY] = 0$. X et Y sont-elles corrélées ? indépendantes ?

Q 17.3 Soit $U = min(X_1, X_2)$ et $V = max(X_1, X_2)$.

Q 17.3.1 Quelle est la loi de U ? Son espérance ?

Q 17.3.2 Exprimer $U + V$ en fonction de X_1 et X_2. En déduire l'espérance de V.

Q 17.3.3 Exprimer UV en fonction de X_1 et X_2. En déduire $E[UV]$ et $Cov(U, V)$.

Exercice 18 – Loi géométrique

On considère une épreuve de Bernoulli de paramètre p, que l'on répète un nombre indéterminé de fois de façon indépendante. On note S_i la variable aléatoire correspondant au i-ième essai.
La loi géométrique est la loi suivie par la variable aléatoire $X \in \{1, ...\}$, qui compte le nombre d'essais jusqu'au premier succès.

Q 18.1 Déterminez la loi de X.

Q 18.2 Calculez l'espérance de X.
On pourra utiliser les formules suivantes :

$$\forall x\ 0 < x \neq 1, \sum_{k=0}^{n} x^k = \frac{x^{n+1} - 1}{x - 1} \quad et \quad \forall k \in \mathbb{N}, \left(-(1-y)^k\right)' = k(1-y)^{k-1}$$

Exercice 19 – Exemples d'algorithmes stochastiques : RandomSort et RandomPerm

Q 19.1 `RandomPerm`
On suppose que l'on a accès à une fonction `RandomPerm`(n), qui renvoie une permutation de l'ensemble $\{1, ..., n\}$, prise selon la distribution uniforme.

Q 19.1.1 Quelle est la probabilité que `RandomPerm`(n) tire une permutation particulière ?

Q 19.1.2 Soit $[x_0, ..., x_n]$ un tableau contenant n valeurs distinctes. On considère l'algorithme de tri suivant :
`RandomSort(k)` :

1. tri $\leftarrow$ `faux`
2. $i \leftarrow 1$
3. tant que `non`(trie) et $i \leq k$ faire :
4. $\quad$ $p \leftarrow$ `RandomPerm`(n)
5. $\quad$ si $\forall j \in \{0, ..., n-1\}, x_{p(j)} \leq x_{p(j+1)}$:
6. $\quad\quad$ tri $\leftarrow$ `vrai`
7. $\quad$ fin si

8. $i \leftarrow i+1$
9. fin tant que
10. retourner $[x_{p(0)}, ..., x_{p(n)}]$

Q 19.1.3 En fonction de la valeur de k, quelle est la probabilité que `RandomSort` renvoie le tableau trié ?

Q 19.1.4 Pour quelle valeur de k a-t-on une probabilité supérieure à 0.5 que la fonction renvoie le tableau trié ? Cet algorithme est-il efficace en pratique ?

Q 19.2 `RandomPerm`
On suppose que l'on a accès à un générateur de nombres aléatoires `rand`(k), qui renvoie un entier entre 0 et k. Considérons l'algorithme suivant :

1. $tab = [0, 1, ..., n]$
2. pour k allant de n à 1 faire :
3. $p \leftarrow$ `rand`(k)
4. échanger les valeurs de $tab[k]$ et $tab[p]$
5. fin pour
6. retourner tab

On note X la variable aléatoire qui correspond au tableau tab en sortie de l'algorithme. Montrez que :

Q 19.2.1 pour tout $i \in \{0, ..., n\}$, $P(X[n] = i) = \frac{1}{n+1}$;

Q 19.2.2 pour tout $(i, j) \in \{0, ..., n\}^2$, si $i \neq j$ alors $P\big(X[n-1] = j \mid X[n] = i\big) = \frac{1}{n}$;

Q 19.2.3 Soit $(i_0, ..., i_n) \in \{0, ..., n\}^{n+1}$ un $(n+1)$-uplet contenant $n+1$ valeurs distinctes. Que vaut $P(X = [i_0, ..., i_n])$?

semaine 4

Exercice 20 – Loi binomiale - le retour

Q 20.1 On dispose de deux dés qu'on lance simultanément 12 fois de suite et on appelle X le nombre de double six obtenus sur les 12 lancers.

Q 20.1.1 Quelle est la loi de X ? Moyenne et variance ?

Q 20.1.2 Calculer $P(X \leq 2)$

Q 20.1.3 Estimer cette quantité en approximant par une loi de Poisson.

Exercice 21 – Casque variable

Q 21.1 Une entreprise fabrique des casques audios dont 6% sont défectueux. Chaque casque est soumis à un contrôle qui n'est pas parfait : le contrôle rejette 98% des casques défectueux et 5% des corrects. Soit D l'événement *Casque défectueux* et R *Casque rejeté lors du contrôle*

Q 21.1.1 Quelle est la probabilité $P(E)$ d'une erreur lors du contrôle ?

Q 21.1.2 Quelle est la probabilité qu'un casque soit rejeté ?
Quatre contrôles indépendant sont réalisés. Le casque est commercialisé avec le logo de l'entreprise si les quatres contrôles sont passés avec succès, détruit s'il rate au moins deux contrôles, commercialisé sans logo sinon. Le coût de fabrication d'un casque est de 50$. Son prix de vente est de 120$ avec logo et de 60$ sans logo. Soit G la variable aléatoire du gain pour la vente d'un casque.

Q 21.1.3 Donner la loi de probabilité de G, son espérance et sa variance.

Exercice 22 – Marketing

Un magasin vend deux produits : A au prix de 8 euros, B au prix de 12 euros. Soit X la variable aléatoire du nombre de produits A acheté par un client et Y pour la variable aléatoire pour le produit B. On suppose que X suit une loi de poisson de paramètre 2 et Y de paramètre 1, chacune indépendante de l'autre [1]

Q 22.1 Calculer la probabilité pour qu'une personne (a) n'achète rien ; (b) n'achète que des produits de type A ; (c) achète des produits de type A et B.

Q 22.2 Que vaut la variable aléatoire Z dénotant le montant d'achat d'un client ? Calculer son espérance et sa variance.

Q 22.3 On a remarqué que 350 personnes passent au magasin chaque jour. Soit W le chiffre d'affaires du magasin sur une journée.

Q 22.3.1 Déterminer l'espérance et la variance de W

Q 22.3.2 En utilisant Tchebychev [2], trouver un minorant de $P(9000 < W < 10600)$.

1. Loi de poisson : $P(X = k) = \lambda^k/k!e^{-\lambda}$, espérance et variance : λ.
2. Inégalité de Tchebychev : $P(|X - E(X)| > a) \leq Var(X)^2/a^2$.

Exercice 23 – Communication à travers un canal bruité

Deux personnes, Alice et Bob, souhaitent s'envoyer des messages à travers un réseau. Pour cela, le message (initialement du texte) est codé par une suite de 0 et de 1, transmis par l'intermédiaire du réseau. La séquence de 0 et 1 est décodée de l'autre côté du réseau pour retrouver le message original :

$$\begin{array}{ccc} \text{Alice} & & \text{Bob} \\ \text{bonjour !} & & \\ \downarrow & & \\ 011...110 & \text{canal de transmission} & 011...110 \\ & & \downarrow \\ & & \text{bonjour !} \end{array}$$

En pratique, pour des raisons physiques, les canaux de transmission sont légèrement bruités : avec une probabilité $p < 1/2$, un bit (0 ou 1) peut être inversé. Autrement dit, il existe une petite probabilité pour qu'un 1 initialement transmis par le canal soit reçu comme 0, et inversement. La conséquence directe est que le message peut être reçu incorrectement :

$$\begin{array}{ccc} \text{Alice} & & \text{Bob} \\ \text{bonjour !} & & \\ \downarrow & & \\ 011...110 & \text{canal de transmission bruité} & 010...110 \\ & & \downarrow \\ & & \text{bmnjour !} \end{array}$$

Dans le suite de l'exercice, on suppose que l'inversion d'un bit se fait avec une probabilité constante, et indépendamment du reste du message transmis.

Q 23.1 Supposons que le message qu'envoie Alice à Bob soit d'une longueur de k bits. Quelle est la probabilité qu'il y ait au moins une erreur dans la transmission ?

Q 23.2 Une méthode simple pour comtrôler la probabilité d'une erreur de transmission sans changer le canal de communication est d'utiliser un *code correcteur d'erreur*. Le plus simple d'entre consiste à répéter chaque bits n fois (n impair). Á la réception d'un message, on inspecte les séquences successives de longueur n, et on considère que le bit initial est le plus fréquent parmi les n reçus. Voici un exemple avec $n = 3$:

$$\begin{array}{llll} \text{message original :} & 1010 & & \\ \text{message recodé :} & 111000111000 & \text{canal de transmission bruité} & 110000111000 \\ & & & \downarrow \text{décodage} \\ & & & 1010 \end{array}$$

On note X_n le nombre d'erreurs du canal de transmission sur une séquence de n bits. Quelles sont l'espérance et la variance de X_n ?

Q 23.3 On note p_n la probabilité d'une erreur de décodage avec le codage qui consiste à répéter chaque bit n fois. Quelle est la relation entre p_n et X_n ?

Q 23.4 Afin d'avoir rapidement une borne supérieure sur p_n, on décide d'utiliser l'inégalité de Chebychev. Pour cela, on procède en deux étapes :

1. exprimez p_n en terme d'une probabilité d'écart entre X_n et son espérance,
2. appliquez l'inégalité de Chebychev pour avoir une borne supérieure sur cette probabilité d'écart.

Q 23.5 On suppose que le canal a une probabilité $p = 0.001$ d'inverser un bit. Donnez une valeur de n telle que $p_n \leq p/10$.

Q 23.6 Bien que très générale, l'inégalité de Chebychev peut être extrêmement pessimiste dans certains cas particulier. Notamment dans le cas des variables binomiales, où l'on peut démontrer *l'inégalité de Hoeffding* (on admet ce résultat) :

$$P(X_n - \mathbb{E}(X_n) > nt) \leq e^{-2nt^2}$$

A l'aide de cette nouvelle inégalité, donnez une valeur de n telle que $p_n \leq p/1000$.

Exercice 24 – Salaires

Le salaire moyen d'un travailleur français est (environ) de 2000 euros brut par mois. Considérant uniquement cette donnée, nous allons essayer de répondre à la question suivante : si l'on échantillonne aléatoirement et uniformément n travailleurs français, quelle est la probabilité que le *moins bien payé* de ces n travailleurs ait un revenu inférieur à 2200 euros brut ?
On note X_1, ..., X_n, n variables aléatoires indépendantes et identiquement distribuées, où chaque X_i est le revenu de la i-ième personne échantillonnée. On a $\forall i \ \mathbb{E}(X_i) = 2000$.

Q 24.1 Fixons $i \in \{1, ..., n\}$. En utilisant l'inégalité de Markov, donner une borne supérieure sur la proba que X_i soit supérieur à 2200.

Q 24.2 En déduire une borne supérieure sur la proba que $\forall i \ X_i > 2200$.

Q 24.3 En déduire la plus petite valeur de n pour laquelle avec avec une probabilité supérieure à $0{,}95$, le plus petit des n revenus observés est inférieur à 2200 euros.

Q 24.4 Dans cette dernière question, nous essayons de répondre à la question suivante : supposons que l'on échantillonne 50 personnes aléatoirement (uniformément et indépendemment) parmi les travailleurs français. Quel est le plus petit revenu possible r pour lequel nous pouvons garantir qu'avec une probabilité supérieur à $0{,}95$, le plus petit des n revenus observés est inférieur à r euros ?
Répondez à cette question en reprenant les étapes des questions précédentes. La seule différence à prendre en compte est que maintenant n est connu ($n = 50$), mais la valeur du revenu (qui était fixée à 2200) est maintenant l'inconnue que vous devez déterminer.

semaine 5

Exercice 25 – Tir à l'arc sur une cible de rayon R

Considérons deux archers visant une cible de rayon R. On suppose que les archers sont assez adroits pour ne jamais rater la cible. La performance de chaque archer est mesurée par la distance au centre de la cible : si un archer envoie un flèche à une distance t du centre, il gagne t points. L'objectif est de posséder le moins de points possibles. On note X_1 (respectivement X_2), la variable aléatoire comptant le nombre de points du premier (resp. deuxième) archer. Le niveau de chaque archer est connu :
Pour $t \in [0, R]$, $P(X_1 \leq t) \propto t$, et $P(X_2 \leq t) \propto \sqrt{t}$.

Q 25.1 Calculez et tracez les fonctions de répartitions de X_1 et X_2.

Q 25.2 Soit $t \in [0, R]$. Que vaut $P(X_1 = t)$?

Q 25.3 Déterminez les densités et les espérances de X_1 et X_2.

Q 25.4 Qui est le meilleur archer ?

Exercice 26 – Loi exponentielle

Une variable aléatoire à valeurs réelles X suit une loi exponentielle de paramètre λ si elle admet comme fonction de densité :

$$p_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & sinon \end{cases} \tag{1}$$

Q 26.1 Calculez la fonction de répartition de X ;

Q 26.2 Soient $X_1, X_2, ..., X_n$, n variables aléatoires indépendantes suivant des lois exponentielles de paramètres respectifs $\lambda_1, ..., \lambda_n$. Montrez que $Y = \min(X_1, ..., X_n)$ suit une loi exponentielle de paramètre $\lambda_1 + ... + \lambda_n$.

Exercice 27 – Diverses Densités

Q 27.1 Parabolique Soit X une variable aléatoire de densité $f(x) = c(1 - x^2)\mathbf{1}_{-1<x<1}$

Q 27.1.1 Déterminer c pour que f soit bien une densité de probabilité.

Q 27.1.2 Quelle est la fonction de répartition de X ?

Q 27.1.3 Calculer $\mathbb{E}[X]$ et la variance.

Q 27.2 Loi de Pareto Soit la variable aléatoire T représentant la durée de vie en heures d'un composant électronique, qui a pour densité $f(t) = \frac{K}{t^2}\mathbf{1}_{t>10}$.

Q 27.2.1 Que vaut K ?

Q 27.2.2 Calculer $P(T > 20)$.

Q 27.2.3 Représenter la fonction de répartition.

Q 27.2.4 Quelle est la probabilité que parmi 6 composants indépendants, au moins 3 d'entre eux fonctionnent durant au moins 15h.

Q 27.3 Soit X une variable aléatoire qui suit une loi uniforme sur le segment $[0,1]$ et soit $Y = 1 - X$. Donner la fonction de répartition de Y. En déduire la densité de Y. Est-ce que la variable aléatoire $Z = X + Y$ admet une densité ?

Q 27.4 On construit une variable aléatoire X en commençant par lancer une pièce équilibrée : si on obtient Pile, alors $X = 1$; si on obtient face, X est le résultat d'un tirage uniforme dans $[0,1]$. Donner la fonction de répartition de X.

Exercice 28 – Histoire de segments

Soit X un point au hasard sur le segment $[0,1]$.

Q 28.1 Quelle est la probabilité que $X > \frac{3}{4}$?

Q 28.2 Quelle est la probabilité que X soit supérieur à $\frac{3}{4}$ sachant qu'il est supérieur à $\frac{1}{3}$?

Q 28.3 Le point X définit les deux segments $[0,X]$, $[X,1]$. Quelle est la probabilité pour que le rapport entre le plus grand et le plus petit des deux soit supérieur à 4 ?

Q 28.4 On tire deux points X et Y au hasard sur le segment $[0,1]$, indépendamment. Quelle est la probabilité que le plus petit des deux nombres soit supérieur à $\frac{1}{3}$? Que le plus grand des deux nombres soit supérieur à $\frac{3}{4}$ sachant que le plus petit des deux est supérieur à $\frac{1}{3}$?

Exercice 29 – Loi normale

Q 29.1 On suppose que la masse X d'une pomme de terre suit une loi normale de moyenne $m = 200g$ et d'écart type $\sigma = 70g$. Quelle est la probabilité qu'une pomme de terre :

Q 29.1.1 pèse plus de 250 grammes ? pèse moins de 180 grammes ?

Q 29.1.2 ait une masse comprise entre 190 et 210 grammes ?

Q 29.2 On admet que la longueur du pied d'un homme adulte suit une loi normale de moyenne $24cm$ et d'écart-type $3cm$. Un fabriquant de chaussettes étudie cette loi pour programmer sa production de chaussettes en taille et en quantité. Il décide de répartir sa production selon 5 tailles numérotées de 1 à 5 de la façon suivante : il prend un intervalle symétrique autour de la moyenne, de probabilité $p = 0.9$; il divise cet intervalle en 3 intervalles égaux correspondant aux tailles 2, 3 et 4. Il obtient donc ainsi son total de 5 tailles.

Q 29.2.1 Déterminer les longueurs de pied qui délimitent ces 5 intervalles.

Q 29.2.2 Quelle est la part, en pourcentage, de la production totale à affecter respectivement à chacune des 5 tailles ?
(Données : $\Phi(4.95/3) = 0.95$ et $\Phi(4.95/9) = 0.71$ où Φ est la fonction de répartition de la variable centré réduite.)

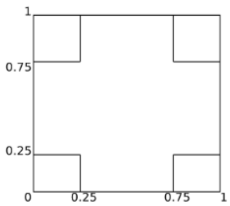
Exercice 30 – Simulation

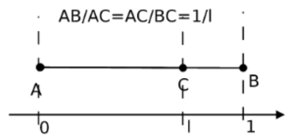
On dispose d'un générateur aléatoire $G(t)$ qui permet d'engendrer des nombres aléatoires réels entre 0 et 1 de manière uniforme pour chaque $t > 0$ entier.

Q 30.1 Quelles sont les fonctions f de densité et F de répartition de cette loi ?

Q 30.2 Donner un algorithme permettant de simuler un tirage de manière uniforme dans le rectangle $[0,1] \times [0,1]$ à l'aide du générateur aléatoire G. Jusitifer.

Q 30.3 Donner un algorithme permettant de simuler un tirage de manière uniforme dans la croix de la figure ci-dessous. Quelle est la complexité de votre algorithme pour tirer un point ? Peut-on faire mieux ?





Q 30.4 Le nombre d'or ϕ est la racine du polynôme $f(x) = x^2 + x - 1$. Pour un segment AB de longueur 1, le point C de ce segment donnant le nombre d'or est tel que $\phi = AB/AC = AC/CB$. En notant $l = AC$ la longueur du segment AC, l est l'inverse de $\phi = 1/l = l/(1-l)$. De quel polynôme l est la solution ? Comment se comporte ce polynôme sur l'intervalle $[0, 1]$? Comment savoir simplement si un point de cet intervalle est inférieur ou supérieur à l ? Donner un algorithme de simulation permettant de vous donner une approximation de l. Comment appelle-t-on ce type de simulation ?

Interlude - Révisions

Exercice 31 – Petites questions indépendantes de révision

Q 31.1 Quel est le nombre de groupes de 8 personnes que l'on peut former avec 5 garçons et 7 filles si on veut qu'ils contiennent :

Q 31.1.1 3 garçons donnés (et les autres peuvent être les autres garçons ou des filles)

Q 31.1.2 exactement 3 garçons et donc 5 filles.

Q 31.1.3 au moins 3 garçons.

Q 31.2 On répartit sept jetons numérotés de 1 à 7 dans trois urnes.

Q 31.2.1 Combien re répartitions possibles ? Quelle est la probabilité que l'urne 1 soit vide ?

Q 31.2.2 Que l'urne 1 soit la seule vide ?

Q 31.3 Un échiquier comporte 5 lignes, 5 colonnes, soit 25 cases. De combien de façons peut on placer 5 pions de telle sorte qu'il y en ait un seul par ligne et un seul par colonne ?

Q 31.4 Quelle est la probabilité qu'un 6 apparaisse 1 fois dans un lancer de dé à 6 faces (équilibré) ? Combien de fois faut-il le lancer pour que la probabilité qu'un 6 apparaisse soit de $\frac{1}{2}$?

Q 31.5 On lance deux dés équilibrés. Soit X la variable aléatoire qui dénote le plus grand des numéros obtenus et Y le plus petit. Quelle est la loi jointe $P(X, Y)$? Quelles lois suivent X et Y ? Quelle est l'espérance de X et Y ?

Q 31.6 Vous arrivez à un arrêt de bus à 22h30.

Q 31.6.1 Quelle est la probabilité que vous attendiez plus de 10 minutes le bus sachant qu'il arrivera à un certain instant distribué uniformément entre 22h30 et 23h ?

Q 31.6.2 Il est 10h40, le bus n'est toujours pas passé. Quelle est la probabilité que vous attendiez au moins 10 minutes supplémentaires ?

Q 31.7 Soit X et Y deux variables aléatoires indépendantes suivant toutes les deux la loi uniforme dans $[0, 1]$.

Q 31.7.1 Donner une représentation graphique de la probabilité $P\left(Y \leq \frac{X}{2}, Y \geq X - \frac{X}{2}\right)$ et de $P\left(Y \leq \frac{X}{2} | Y \geq X - \frac{X}{2}\right)$.

Q 31.7.2 Calculer $P\left(Y \leq \frac{X}{2} | Y \geq X - \frac{X}{2}\right)$.

Q 31.7.3 Comment feriez-vous pour vérifier expérimentalement par une simulation votre résultat ?

Q 31.8 On réalise des expériences indépendantes chacune avec une probabilité de succès de $p = 0.1$. On répète chaque expérience jusqu'à obtenir 100 succès. On note X le nombre d'expériences totales nécessaires.

Q 31.8.1 Proposer une modélisation du problème pour calculer l'espérance de X et sa variance.

Q 31.8.2 En utilisant une approximation par la loi normale, quelle est la probabilité que le nombre d'expériences ne dépasse pas 1050 ?

Q 31.9 On observe à une borne vélib que en moyenne il y a deux emprunts par 5 min. Les emprunts sont indépendants.

Q 31.9.1 L'hypothèse d'indépendance vous semble-t-elle juste ?

Q 31.9.2 Un emprunt est observé à t_0. Quelle est la probabilité qu'il y ait au moins un autre emprunt dans les 5 prochaines minutes ?

Q 31.9.3 On suppose que la borne contient 4 vélibs (et qu'il n'y a pas de retour de vélibs). Chaque emprunt rapporte 1.5 euro, le coût d'entretien de la borne est de 1 euro pendant 5 min. Quel est l'espérance de gain ?

Q 31.10 Dans une station service, la demande hebdomadaire en essence, en milliers de litres, suit une variable aléatoire X de densité $f(x) = c(1-x)^2$ si $x \in [0, 1]$, 0 sinon.

Q 31.10.1 Déterminer c

Q 31.10.2 Calculer $\mathbb{E}(X)$.

Q 31.10.3 La station est réapprovisionnée le lundi matin une fois par semaine. Quelle doit être la capacité minimale du réservoir d'essence pour que la probabilité d'épuiser ce réservoir soit inférieure à 10^{-3} ?

Q 31.11 On suppose que la distance entre deux stations-service consécutives le long d'une autoroute est de 50km. On appelle X la distance parcourue par une automobile qui tombe en panne entre deux stations-service depuis le passage devant la première. On fait l'hypothèse qu'elle suit une loi uniforme sur $[0, 50]$.

Q 31.11.1 Déterminer la probabilité d'être à plus de 10km d'une station de service.

Q 31.11.2 La distance entre le lien de la panne et la station-service la plus proche est une fonction de X, notée $g(X)$. Donner l'expression de $g(x)$ en fonction de la valeur x prise par la variable X.

Q 31.11.3 Donner l'espérance de $g(X)$ sans calcul si possible.

Q 31.12 Insecte et loi de Poisson
Un insecte pond des oeufs, en nombre aléatoire X de loi de Poisson de paramètre a. Chaque oeuf éclôt ou n'éclôt pas, indépendamment de tous les autres, et la probabilité p d'éclosion est la même pour chaque oeuf.
Calculer la loi du nombre Y d'oeufs éclôts.

Q 31.13 On estime que 1% des ordinateurs du parc informatique mondial fonctionnent sous Linux. Parmi les ordinateurs tournant sous Linux, 50% utilisent le navigateur Firefox, contre 5% pour les ordinateurs fonctionnant sous un autre système d'exploitation. Déterminer la probabilité qu'un ordinateur fonctionne sous Linux dès lors qu'on a détecté que le navigateur Firefox est employé.

Q 31.14 Deux joueurs lancent une pièce de monnaie parfaitement équilibrée, n fois chacun. On note X, Y le nombre de "`pile`" obtenus respectivement par A, B.

Q 31.14.1 Pour tout k dans $[[0, n]]$, calculer la probabilité de l'événement : $(X = k)$, $(Y = k)$, et $(X = k)$ et $(Y = k)$.

Q 31.14.2 En déduire la probabilité que A et B obtiennent le même nombre de fois "pile'.

Q 31.15 Dans un jeu, il y a n lots et p joueurs. Chaque joueur choisit un lot au hasard et on appelle X le nombre de joueurs choisissant le lot 1. Donner la loi de X.

Q 31.16 Soit une fonction `unknowRandom()` qui renvoie au hasard 0 ou 1 de manière possiblement biaisé. En utilisant deux tirages simultanées mais possiblement plusieurs fois, comment simuler un tirage non biaisé ?

Q 31.17 Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et suivant toutes la loi uniforme sur $[0, 1]$, et $M = \max(X_1, \dots, X_n)$. Quelle est la fonction de répartition de M ? même question avec le minimum.

Q 31.18 Une page contient en moyenne 1.5 fautes, quelle est la probabilité que ce document contient au moins 3 fautes ?

Q 31.19 La durée de vie d'une ampoule suit une loi exponentielle [3], avec une durée de vie moyenne de 5 mois. Quelle est la probabilité que l'ampoule dure moins de 3 mois ? Un luminaire utilise 5 ampoules, durée de vie indépendante, quelle est la probabilité qu'il faille en remplacer exactement 2 avant 3 mois ?

Q 31.20 Soit X une variable aléatoire. On suppose que la loi de X a pour densité la fonction f définie par : $\forall x \in \mathbb{R}, f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{b}{3} & \text{si } 0 \leq x \leq 1 \\ \frac{c}{x^4} & \text{si } x \geq 1 \end{cases}$ où c est une constante.

Q 31.20.1 En supposant la fonction de densité continue en $x > 0$, quelle est le lien entre b et c ? Quelles sont les autres contraintes pour ces deux constantes ? Quelles sont les valeurs possibles ? Représenter graphiquement la fonction f.

Q 31.20.2 Calculer la fonction de répartition F_X de X.

Q 31.20.3 Calculer : $P(X = 2)$, $P(0.5 < X < 2)$, et $P(X > 3|X > 1)$.

Q 31.20.4 Calculer l'espérance de X.

Q 31.20.5 Soit $Y = X^3$. Calculer $P(Y \leq 1)$. Donner les fonctions de répartition et de densité de Y.

Exercice 32 – De passage...

Deux personnes A et B partent en vacances de façon indépendante dans un pays E. Leur séjour dans ce pays peut s'étaler sur n journées (n > 3) numérotées de 1 à n. Pour éventuellement s'y rencontrer, elles ont projeté d'y séjourner trois jours consécutifs (et trois jours seulement) dans un hôtel H, choisi par elles. On suppose que les jours d'arrivée possibles sont uniformes et indépendantes. Les arrivées ont lieu le matin et les départs le soir deux jours plus tard.

Q 32.1 Quelle est la probabilité que A et B arrivent le même jour ?

Q 32.2 Quelle est la probabilité qu'elles arrivent avec un jour d'écart ?

Q 32.3 Quelle est la probabilité qu'elles puissent se rencontrer dans l'hôtel ?

Q 32.4 Sachant que A et B se sont rencontrées, quelle est la probabilité qu'elles ne puissent passer qu'une journée ensemble ?

Exercice 33 – Détection du paludisme

Il fut un temps où, dans l'armée française, pour rechercher les hommes atteints d'une maladie (le paludisme), on effectuait sur chacun un prélévement sanguin. On regroupait dans un même récipient les prélévements de k hommes et on cherchait le parasite dans le mélange

- si le parasite était absent, on décrétait que aucun des k hommes n'était atteint

3. $p(t) = \lambda e^{-\lambda t}$, $E(x) = 1/\lambda$, $V(x) = 1/\lambda^2$

- si le parasite était présent, on prélevait à nouveau les k hommes concernés et on examinait un à un les prélévements.

On suppose que chaque homme a une probabilité p d'être atteint et que la maladie touche les hommes indépendament les uns des autres.

Q 33.1 Soit $X^{(k)}$ le nombre d'analyses de sang à effectuer pour un groupe de k hommes. Quelle est la loi de $X^{(k)}$? Calculer $E(X^{(k)})$.

Q 33.2 En déduire que, si l'on cherche le parasite chez N hommes (N pourra être supposé multiple de k), l'espérance du nombre de tests à faire vaut $N/k + N(1-(1-p)^k)$.

Q 33.3 Si on prend $k = 10$, pour quelle valeur de p cette procédure permet de faire moins d'analyses en moyenne que la procédure simple consistant à tester dès le début chacun des hommes séparément ?

Q 33.4 Connaissant p, quelle est la valeur de k qui minimise l'espérance du nombre de tests à faire.

semaine 6

Exercice 34 – roues et intervalle de confiance

Une entreprise fabrique, en grande quantité, des roues métalliques pour l'industrie. On s'intéresse au diamètre des roues, exprimé en mètres. On prélève au hasard et avec remise un échantillon de 50 roues dans la production d'un journée.
Soit D la variable aléatoire qui, à tout échantillon de 50 roues prélevées au hasard et avec remise dans la production d'un journée, associe la moyenne des diamètres des roues de cet échantillon.
On suppose que D suit une loi normale de moyenne inconnue μ et d'écart-type $\frac{\sigma}{\sqrt{50}}$ avec $\sigma = 0.19$. Pour l'échantillon prélevé, la moyenne obtenue est $x = 9.99$.

Q 34.1 À partir des informations portant sur cet échantillon, donner une estimation ponctuelle de la moyenne μ des roues produites dans cette journée.

Q 34.2 Déterminer un intervalle de confiance centré sur de la moyenne μ des diamètres des roues produites pendant la journée considérée, avec le coefficient de confiance de 95%.

Q 34.3 On considère l'affirmation suivante : la moyenne μ est obligatoirement dans l'intervalle de confiance obtenue à la question précédente. Est-elle vraie ? Pourquoi ?

Exercice 35 – Sondages et intervalles de confiance

Lors d'un sondage électoral, on interroge 1000 électeurs : 52% des ces électeurs interrogés affirment qu'ils voteront pour le candidat A. On note p_A la proportion des électeurs qui votent pour le candidat A dans la population (on suppose la population totale de l'ordre de plusieurs millions).
Q 35.1 Donnez un intervalle de confiance de la proportion p_A. Peut-on en déduire la victoire du candidat A avec un niveau de confiance à 95% ?

Q 35.2 Quel effectif serait suffisant pour annoncer la victoire du candidat recueillant 52% des voix au seuil 5% ?

Q 35.3 On veut maintenant une borne sur la précision d'un sondage sans estimation de vote préalable. Calculez la valeur de p_A pour lequel la variance $\sigma^2(X)$ est maximale. Quel est le nombre de sondés n_α nécessaire pour obtenir un niveau de confiance d'au plus α% ?

Exercice 36 – Intervalle de confiance

Un développeur vient de terminer un filtre spam. Pour évaluer la performance de son programme, il sélectionne n = 1000 e-mails au hasard, tirés indépendamment et selon la même distribution que les e-mails de ses futurs utilisateurs. Le développeur étiquette lui-même les 1000 e-mails (chaque mail est donc étiqueté selon que c'est du spam ou non), puis applique son filtre sur les 1000 e-mails. Parmi ces e-mails, il mesure que son filtre fait 50 erreurs.

Q 36.1 Soit X la variable aléatoire qui compte le nombre d'erreurs du filtre sur un échantillon de 1000 e-mails. Quelle est la loi de X ?

Q 36.2 On admet que, dans notre cas, la probabilité p que le filtre fasse une erreur sur un e-mail choisi aléatoirement est suffisante pour que le théorème de la limite centrale s'applique. Quelle est l'approximation de $P(X \leq t)$ donnée par le théorème de la limite centrale dans notre cas ?

Q 36.3 On admet que $\sqrt{p(1-p)}$ peut être approximé par $\sqrt{\hat{p}_n(1-p_n)}$ avec $\hat{p}_n = X/n$ (rappel : $n = 1000$ est le nombre d'e-mails). Donnez un intervalle de confiance à 90% pour p.

Exercice 37 – Intervalles de confiance

Un chercheur en informatique mène des recherches sur l'efficacité d'un algorithme d'apprentissage. Il étudie en particulier son temps de calcul. Sur un échantillon de 15 base de données générées aléatoirement, il trouve les résultats suivants (en ms) :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 900	1 500	3 100	1 800	3 900	1 850	5 400	2 700	1 150	3 000	1 700	3 500	1 800	4 000	2 200

L'informaticien suppose que les temps de calcul sont répartis suivant une loi normale d'espérance μ et de variance σ^2.

Q 37.1 Donner la meilleure estimation possible de μ et de σ^2 à partir de l'échantillon.

Q 37.2 En supposant que $\sigma = 1\,158.35$, donner pour μ un intervalle de confiance de niveau de confiance 95% (au seuil de 5%). Quelle est la taille de cet intervalle de confiance (son amplitude) ?

Q 37.3 L'informaticien voudrait obtenir un intervalle de confiance d'une amplitude de 500 millisecondes (toujours au seuil de 5%). Sur quel paramètre de son expérience peut-il jouer pour modifier l'amplitude de l'intervalle de confiance ? Calculer la valeur que devrait prendre ce paramètre pour obtenir cette amplitude de 500ms ?

semaine 7

Exercice 38 – Test de conformité d'une moyenne

On souhaite s'assurer que la teneur en matière grasse d'un jaune d'œuf est conforme à celle annoncée, à savoir $308\,g/kg$. Pour cela des mesures par extraction éthérochlorhydrique ont été réalisées sur 12 œufs différents. A partir de ces données, on a obtenu une estimation de la moyenne et de la variance de la teneur en matière grasse : $\bar{x} = 305.33\,g/kg$ et $s'^2 = 10.06\,(g/kg)^2$

Q 38.1 En première approximation, quelles hypothèses sont nécessaires si l'on veut construire un intervalle de confiance à α% ou des tests sur la teneur moyenne en matière grasse ?
Q 38.2 On suppose vérifié l'hypothèse évoquée à la question précédente. Peut-on admettre

que la teneur moyenne en matière grasse de l'œuf est de $308\,g/kg$ (valeur fréquemment trouvée dans la littérature) ?

Exercice 39 – Dépistage de maladies

Un examen systématique de dépistage est institué pour détecter une maladie M dans une population dite à risque et comportant 10% de malades.
L'examen consiste, après une série d'analyses préliminaires, à relever le nombre de globules blancs. Ce nombre est dit normal s'il est compris entre 4280 et 4420, insuffisant s'il est compris entre 4260 et 4280 et très insuffisant en-dessous de 4260.
On adopte la règle de décision suivante :

- Si le nombre de globules blancs est très insuffisant, on accepte l'hypothèse H_1 : "l'individu est malade"
- s'il est normal ou insuffisant, on accepte l'hypothèse H_0 : "l'individu n'est pas malade".

Une étude a montré que, dans cette population à risque, 20% des individus malades ont un nombre normal de globules blancs, et 15% un nombre insuffisant. Ces chiffres sont respectivement de 90% et 5% pour les individus bien portants.

Q 39.1 Quelles sont les erreurs de 1ère et de 2ème espèce du test décrit ci-dessus ?
Q 39.2 Que deviennent ces erreurs si on accepte H_1 lorsque le nombre de globules blancs

est insuffisant ou très insuffisant et H_0 lorsqu'il est normal ?

Exercice 40 (5 points) – Nouvelle espèce

Dans un laboratoire de biologie, on étudie des modifications génétiques afin de rallonger la vie d'une espèce de mouche particulière. On suppose que la durée de vie de cette mouche suit une loi normale de moyenne 10000 minutes et d'écart-type 1200 minutes. Sur un échantillon de 100 mouches génétiquement modifiées, on observe une durée de vie de 10200 minutes (on suppose que l'écart-type reste le même).
La question posée est la suivante : la modification génétique a-t-elle vraiment modifiée la durée de vie de cette espèce de mouches ?

Q 40.1 Poser le problème sous la forme d'un test d'hypothèse. On supposera que l'on veut prouver que la modification génétique n'a pas modifié la durée de vie de la mouche. Quel type de test utilise-t-on ?

Q 40.2 Donner la région critique de ce test pour un seuil de 0.05. Qu'en concluez-vous ?

Q 40.3 La question est maintenant légèrement modifiée : il s'agit de savoir si la modification génétique a **augmenté** la durée de vie de cette espèce de mouches. Quelle modification cela implique-t-il sur le test ? Quelle est la nouvelle région critique ? Quelle est la conclusion ?

Exercice 41 – Ver de terre

Le nématode (*Caenorhabditis elegans*) vit dans le sol. Le pH de son milieu naturel se situe entre 5 et 9. Dans ces conditions, le taux de survie moyen après 14 jours est de $p = 0,5$.
Un biologiste désire tester l'influence du pH sur la survie des nématodes. Pour cela, il place 20 nématodes dans un milieu acide ($pH = 4$). Au bout de 14 jours, il constate que 8 nématodes sont encore vivants.
Que peut-il conclure ?

Exercice 42 – rats enragés

Un laboratoire pharmaceutique veut évaluer l'efficacité d'une nouvelle molécule sur la rage. Sur l'ensemble de la population des rats, la probabilité de souffrir de la pathologie étudiéé est $p = 0,5$. Le laboratoire a traité 20 rats avec la nouvelle molécule, et a observé une proportion $\hat{p} = 0,25$ de rats malades.
Que concluez-vous quant à l'efficacité du traitement ?
Le tableau suivant donne $P(X = k)$ lorsque X suit une loi binomiale $\mathcal{B}(20; 0,5)$.

	$k \leq 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
$P(X = k)$	0%	0,1%	0,5%	1,5%	3,7%	7,4%	12,0%	16,0%	17,6%
$P(X \leq k)$	0%	0,1%	0,6%	2,1%	5,8%	13,2%	25,2%	41,2%	58,8%

	$k = 11$	$k = 12$	$k = 13$	$k = 14$	$k = 15$	$k = 16$	$k = 17$	$k \geq 18$	
$P(X = k)$	16,0%	12,0%	7,4%	3,7%	1,5%	0,5%	0,1%	0%	
$P(X \leq k)$	74,8%	86,8%	94,2%	97,9%	99,4%	99,9%	100%	100%	

Exercice 43 – Couples de jumeaux

On considère 10 couples de jumeaux et on se demande s'il y a une différence de capacité respiratoire entre l'ainé et le cadet.

Q 43.1 Quel test statistique peut-on mettre en oeuvre ? Préciser les hypothèses H_0 et H_1.
Q 43.2 Si H_0 est vraie, quelle est la loi du nombre X de couples de jumeaux où l'ainé à une capacité respiratoire supérieure à celle de son cadet.
Q 43.3 Déterminer avec précision la règle de décision au niveau 10%, 5%, et 1%.

Le tableau suivant donne $P(X = k)$ et $P(X \leq k)$ lorsque X suit une loi binomiale $\mathcal{B}(10; 0,5)$.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
$P(X = k)$	0,1%	1,0%	4,4%	11,7%	20,5%	24,6%	20,5%	11,7%	4,4%	1,0%	0,1%
$P(X \leq k)$	0,1%	1,1%	5,5%	17,2%	37,7%	62,3%	82,8%	94,5%	98,9%	99,9%	100%

Q 43.4 On observe $X = 8$. Pour chaque niveau, que concluez-vous ? Quel niveau préférez-vous prendre ?

Exercice 44 – Marqueurs chromosomiques

On est aujourd'hui capable de déterminer à bon marché ce que porte un individu en certains locus de ses chromosomes (notion de marqueurs). Sous l'hypothèse H_0 qu'une dystrophie est indépendante de ce que porte l'individu au marqueur M, un calcul simple montre que dans certaines conditions, la probabilité que deux frères atteint de dystrophie soient identiques au locus marqueur est $p^* = 1/4$. Si H_0 est faux, cette probabilité est $> 1/4$.

Q 44.1 On recueille un échantillon de $n = 20$ couples de frères. Si H_0 est vraie, quelle est la loi du nombre X de couples identiques au locus marqueur? Donner la formule générale de $P(X = k)$ et calculer les valeurs numériques de $P(X = k)$ et $P(X \leq k)$ pour $k = 0$ à 5. On admettra que la suite du calcul que vous venez de faire donne le résultat suivant :

	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	...
$P(X=k)$	16,9%	11,2%	6,1%	2,7%	1,0%	...
$P(X \leq k)$	78,6%	89,8%	95,9%	98,6%	99,6%	...

Q 44.2 On ne veut rejeter l'hypothèse H_0 que si l'on courre un risque $\leq 5\%$. Construire le test d'hypothèse $H_0 : p^* = 1/4$ contre $H_1 : p^* > 1/4$.

Q 44.3 Que conclure si $X = 7$? Donner le degré de significativité.

Q 44.4 Un généticien suédois prétend que p^* vaut non pas 1/4 mais 1/2. S'il a raison, quelle est la loi de X?

On admettra que le tableau ci dessus devient alors :

	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	...
$P(X=k)$	3,7%	7,4%	12,0%	16,0%	17,6%	...
$P(X \leq k)$	5,8%	13,2%	25,2%	41,2%	58,8%	...

Calculer alors la puissance du test.

Exercice 45 – Optimisation de graines

Une entreprise commercialise des graines de fleur qui, de fait se répartissent en :

- graines fertiles (c'est à dire donnant vraiment une plante - et des fleurs - dans les conditions voulues),
- graines non fertiles (disons mortes).

On notera q le pourcentage de graines non fertiles. L'entreprise souhaite que q ne dépasse pas 10%. Pour s'assurer de cette "qualité", elle procède de temps en temps à un controle en prélevant n graines et en les mettant en culture; l'observation donne alors le nombre X de graines non fertiles dans cet échantillon. Cette fois ci, l'entreprise prélève $n = 25$ graines.

Q 45.1 Par quelle loi peut on modéliser la variable aléatoire X?

Q 45.2 Tester l'hypothèse $H_0 : q = 10\%$ contre $H_1 : q > 10\%$.

Q 45.3 Que doit on conclure si $X = 6$. Donner le degré de significativité.

Q 45.4 Quelle est la puissance de ce test si en réalité q vaut 12%?

Q 45.5 Tracer sommairement la courbe puissance. Le tableau suivant donne les lois $P(X = k)$ lorsque X suit une loi binomiale $\mathcal{B}(25; 0,1)$ ou $\mathcal{B}(25; 0,12)$.

$P(X=k)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$\mathcal{B}(25;0,1)$	7,2%	19,9%	26,6%	22,6%	13,8%	6,4%	2,4%	0,7%
$\mathcal{B}(25;0,12)$	4,1%	13,9%	22,8%	23,9%	17,9%	10,2%	4,6%	1,7%

Exercice 46 – Pièce truquée ?

En jetant une pièce de monnaie quatre fois de suite, on veut tester deux hypothèses alternatives H_0 et H_1. On teste l'hypothèse nulle :

$$H_0 : \text{"P(obtenir pile)=1/2"}$$

contre l'hypothèse

$$H_1 :\text{"P(obtenir pile)=3/4"}$$

On convient de rejeter l'hypothèse H_0 si l'on obtient au moins trois fois pile.

Q 46.1 Calculer le risque de première espèce encouru en prenant cette décision. Calculer $P_{H_1}(X < 3)$. Que mesure cette quantité ?
On lance maintenant la pièce 25 fois de suite et on convient de rejeter l'hypothèse H_0 si le nombre de pile obtenu est supérieur ou égal à un certain seuil n_0 qu'on se propose de déterminer.
Q 46.2 Justifier la forme de la région de rejet de l'hypothèse H_0.

Q 46.3 Déterminer n_0 pour que le risque de première espèce α soit inférieur à 5%. On admettra que $P(X \leq 16) = 0,946$ et que $P(X \leq 17) = 0,978$ lorsque X suit une loi $\mathcal{B}(25; 0,5)$.
Q 46.4 On suppose qu'on obtient 17 fois pile lors de ces 25 lancers. Peut-on dire que la pièce est bien équilibrée ? Quel est le niveau réel du test ?

semaine 8

Exercice 47 – Test d'hypothèses

Le responsable qualité d'une piscine surveille la température du bassin principal pour éviter qu'elle ne dérive trop. On appelle T_0 la température de l'eau à respecter (la consigne). On note T la variable aléatoire représentant la température obtenue par le module de chauffage. D'après les connaissances sur ce module, on sait que l'écart-type respecté par ce module de chauffage sera $\sigma_T = 1$ (en degré Celsius) quelque soit le réglage de la consigne.
Le responsable qualité prélève 40 fois la température du bassin pour évaluer si il y a déréglage ou non de la température.

Q 47.1 Quelle est la procédure de test statistique qu'il va pouvoir utiliser (hypothèses H_0 et H_1, règle de décision au seuil de 5%) ?

Q 47.2 Quelle loi suit T si il y a un déréglage de 0.2 degrés de la température du bassin ?

Q 47.2.1 Représenter graphiquement (schématiquement) la distribution de T sous H_0, les frontières de la règle de décision de la question précédente, la distribution de T sous l'hypothèse de déréglage.

Q 47.2.2 Représenter sur le même graphique la probabilité de détecter ce déréglage avec le règle de décision de la question précédente. Calculer cette probabilité.

Q 47.3 Le responsable veut détecter au moins 9 fois sur 10 un tel déréglage. Combien de prélèvements doit-il faire ?

Exercice 48 – Test d'ajustement

On a observé 500 valeurs d'une distribution continue dans $[-3, 3[$ que l'on a réparties en 10 classes de la façon suivante :

Classes	$[-3,-2[$	$[-2,-1.5[$	$[-1.5,-1[$	$[-1,-0.5[$	$[-0.5,0[$	$[0,0.5[$	$[0.5,1[$	$[1,1.5[$	$[1.5,2[$	$[2,3[$	
Effectifs	17	17	44	80	92	98	72	43	24	13	500

On voudrait tester si la distribution est symétrique ou non par rapport à 0.

Q 48.1 Quelle est l'hypothèse nulle ? Quelle est l'hypothèse alternative ?

Q 48.2 En remarquant que dans une distribution symétrique, il devrait y avoir le même effectif dans la classe $[x, y[$ que dans la classe $[-y, -x[$, proposer un tableau des effectifs théoriques sous H_0 puis le tableau des contributions au χ^2_{obs}.

Q 48.3 Quel est le nombre de degrés de liberté pour ce test ? Attention, il y a un piège : il faut prendre en compte la remarque de la question précédente.

Q 48.4 Construire le test du χ^2 et conclure.

Exercice 49 – Naissances sur l'année

On a relevé les dates de naissance de 1708 personnes puis on a réparti ces naissances par trimestre.

Trimestre	Nombre de naissances
Janvier – mars	435
Avril – juin	483
Juillet – septembre	410
Octobre – décembre	380
total	1708

Peut-on considérer que les naissances sont uniformément réparties sur les quatre trimestres de l'année ?

Exercice 50 – Les bretons et le beurre salé

On veut vérifier si la consommation de beurre salé est liée au fait d'être breton. Une enquête réalisée auprès de 30 individus a donné les résultats suivants :

	Breton	Non breton	Total
Consomme du beurre salé	10	3	13
Consomme du beurre doux	2	15	17
Total	12	18	30

Q 50.1 Quelles sont les variables et leurs types ?

Q 50.2 Le type de beurre consommé est-il indépendant de la région du consommateur ?

Exercice 51 – Préférences entre 5 chocolats noirs

On a demandé à 30 personnes d'indiquer, parmi 5 chocolats noirs du marché leur chocolat préféré. Le tableau ci dessous donne le nombre de personnes ayant préféré chacun des chocolats.

Chocolat	Excellence	Amère	Amazonie	Pâtisser	Supérieur	total
# préféré	7	5	5	4	9	30

La question générale que l'on se pose est : peut-on considérer que chaque chocolat est préféré par la même proportion de personnes ?
Q 51.1 Quelle est l'hypothèse que l'on veut tester ? Quelle est l'hypothèse alternative ?

Q 51.2 Pour résoudre ce problème, on peut utiliser un test d'ajustement à une distribution.

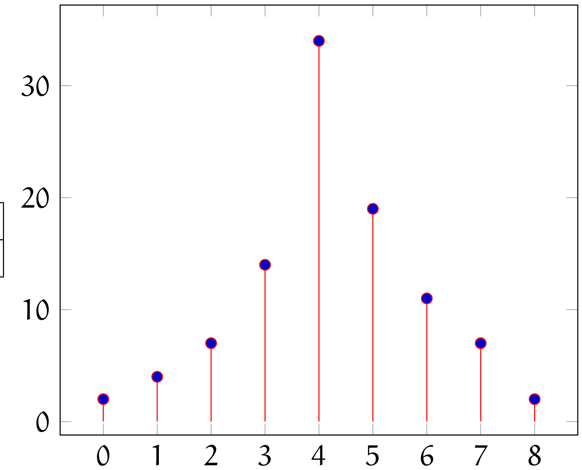
Quelle est la loi de cette distribution ?
Q 51.3 Calculer la statistique de test puis conclure.

Q 51.4 Refaire l'exercice en multipliant les effectif observés par 10. Conclure.

Exercice 52 – Ajustements

Soit la série suivante :

0	1	2	3	4	5	6	7	8	total
2	4	7	14	34	19	11	7	2	100



L'analyste n'a aucune idée du phénomène qui a donné cette série d'effectifs (ou tableau de contingence). Il n'a donc pas beaucoup d'hypothèse sur la loi théorique.

Q 52.1 Premier essai : loi triangulaire
L'analyste propose une loi triangulaire pour modéliser ce comportement.

Q 52.1.1 En supposant que les effectifs théoriques en 0 et 8 sont 1, construire le tableau des effectifs théoriques si la loi est triangulaire, centrée sur 4 (il s'agit essentiellement de trouver a l'accroissement ou décroissement entre 2 effectifs théoriques consécutifs.)

Q 52.1.2 Tester l'ajustement des données à cette hypothèse au seuil de 5%

Q 52.2 Second essai : loi normale L'analyste essaye ensuite d'ajuster sur une loi normale.

Q 52.2.1 Trouver la meilleure estimation possible de la moyenne μ et de la variance σ^2 de cette série.

Q 52.2.2 Dans l'hypothèse d'une loi normale $\mathcal{N}(\mu, \sigma^2)$, utiliser la table de la loi normale centrée réduite pour calculer les effectifs théoriques. **Note** : En notant T_i l'effectif théorique, on discrétise de la façon suivant $T_4 = p(4.5 \leq T \leq 5.5)$.

Q 52.2.3 Tester cet ajustement au seuil de 5%.

Exercice 53 – Liaison et causalité

Lors de l'analyse des résultats des élections présidentielles en 2002, un journaliste confronte les résultats du candidat L... à un grand nombre de variables géographiques. Parmi celles-ci, se trouve le taux de contamination (en Becquerel) des retombées radioactives de l'accident nucléaire de Tchernobyl (1984).
Les résultats sont les suivants :

O_{ij}	taux≤ 750	$\in [750, 1500]$	$\in [1500, 3000]$	$\in [3000, 6000]$	Total
L... (=1er)	4	4	7	20	35
L... (=2ème)	4	8	6	3	21
L... ($\geq$3ème)	23	12	4	1	40
Total	31	24	17	24	96

Q 53.1 Quels sont les variables et leurs types de ces variables ?

Q 53.2 Construire un test permettant d'évaluer le lien entre ces variables ?

Q 53.3 Commenter : que peut-on en conclure ?

Exercice 54 – Test de symétrie

On souhaite évaluer 5 bières à l'aide de comparaison par paires : on goûte les bières 2 à 2 et on indique notre préférence. L'analyste suspecte qu'un biais existe par l'impact du rang de dégustation : certaines bières seraient favorisées quand elles sont goûtées en premier, et certaines autres le seraient quand elles sont goûtées en second. On fait donc goûter à 71 consommateurs les bières 2 à 2, dans les 2 ordres possibles . ce qui nous donne la matrice des O_{ij} qui indique combien ont préféré i à j quand la bière i était goûtée avant la bière j.

	j=1	j=2	j=3	j=4	j=5	Total
i=1	0	22	18	26	21	87
i=2	54	0	23	40	33	150
i=3	48	34	0	49	28	159
i=4	44	17	11	0	21	93
i=5	52	24	20	39	0	135
total	198	97	72	154	103	624

Q 54.1 Calculer le nombre de fois où la biere i a été préféré à la bière j.

Q 54.2 Pour vérifier l'idée de l'analyste, quelle hypothèse H_0 doit-on tester ? Quelle est son alternative ?

Q 54.3 Donner le tableau des effectifs théoriques sous l'hypothèse H_0.

Q 54.4 Donner le tableau des contributions au χ^2, donner le nombre de de degrés de liberté.

Q 54.5 Construire le test et conclure.

semaine 9

Exercice 55 – Simulation sociale

Une enquête sociologique (très rapide) propose de ranger les individus d'une société dans 3 classes sociales différentes :

- B : bourgeoisie
- M : classe moyenne
- O : classe ouvrière

Dans ce modèle, on s'intéresse principalement à la "migration" sociale qui rend compte des transferts générationnels entre classes dans la société. Le principe du modèle simpliste est que pour un enfant et hormis ses qualités intrinsèques considérées comme distribuées aléatoirement, la classe qu'il rejoindra dépend uniquement de la classe de sa famille d'origine.

Le recueil de données statistiques basé sur cet hypothèse permet d'évaluer ces différentes probabilités :

- L'enfant rejoindra la classe B si sa famille était de classe B (0.5), de classe M (0.2) ou de classe O (0.1) ;
- L'enfant rejoindra la classe M si sa famille était de classe M (0.7), de classe B (0.5) ou de classe O (0.3) ;
- L'enfant rejoindra la classe O si sa famille était de classe O (0.6) ou de classe M (0.1).

Q 55.1 Comment formaliser ces probas ? Pourquoi ne somment-elles pas à 1 par ligne ?

Q 55.2 Montrer que ce comportement de la société peut être modélisé par une chaîne de Markov à temps discret. Donner le graphe de transition et la matrice de transition.

Q 55.3 Cette chaîne est-elle irréductible, apériodique, ? Quelle est la nature des états de la chaîne ? Que peut-on en conclure pour ce comportement ?

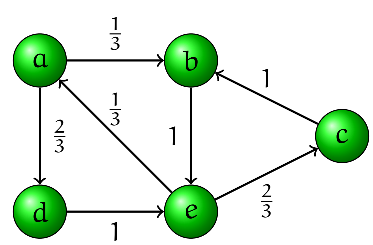
Q 55.4 Déterminer la probabilité stationnaire π de cet chaîne de Markov.

Q 55.5 Dans le régime permanent, quelle est la probabilité d'être d'avoir changé de classe en 2 générations ?

Exercice 56 – Chaîne de Markov

Soit $\mathcal{M}$ la chaîne de Markov ci-contre.

- $\mathcal{M}$ est-elle irréductible ? Écrire la matrice de transition T de $\mathcal{M}$.
- Quelle est la période du nœud a ? $\mathcal{M}$ est-elle apériodique ? Si non, quelle est sa période ?
- Donner une distribution stationnaire pour $\mathcal{M}$.
- Peut-on calculer $\lim_{t\to\infty} T^t$? Pourquoi ?



Exercice 57 – Cursus d'un étudiant

On considère le cursus d'un étudiant en L. Pour chaque année $a \in \{1,2,3\}$, on note p_a la probabilité de réussir son année (et donc de passer à l'année suivante), q_a la probabilité de redoubler et r_a la probabilité d'abandon. On appelera M un succès en L3 (et donc un passage en M) et A un abandon (quelle que soit l'année).

Q 57.1 Pourquoi ce problème peut-il être représenter par une chaîne de Markov ? Quels sont les états de la chaîne de Markov ?

Q 57.2 Dessiner le graphe de la chaîne et la matrice de transition ?

Q 57.3 Caractériser la chaîne et ses états (chaîne irréductible, acyclique, etc. ? états transitiores, absorbants, etc.)

Q 57.4 Soit $\alpha_a^t(n)$ la probabilité d'arriver dans l'état $t \in \{A, M\}$ en n étapes à partir de l'état a. Calculer $\alpha_a^t(n+1)$ en fonction de $\alpha_a^t(n)$.

Q 57.5 Calculer α_a^t la probabilité d'arriver dans l'état $t \in \{A, M\}$ sachant qu'il est parti de l'état a à l'instant initial.

Q 57.6 Avec les probabilités suivantes : $(p_1, p_2, p_3) = (0.4, 0.5, 0.7)$ et $(q_1, q_2, q_3) = (0.3, 0.1, 0.1)$, calculer la probabilité de réussir son cursus L, la probabilité de le réussir en 3 ans, la probabilité de doubler au moins une fois.

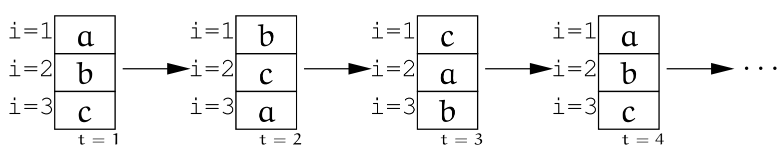
Exercice 58 – File aléatoire

Si, dans un algorithme, on veut sélectionner régulièrement un élément parmi trois (qu'on nommera a, b et c), on peut utiliser une liste et effecuter un algorithme simple de rotation :

ChoixSimple

1. Retirer l'élément à la position 1 de la liste
2. Ranger l'élément sélectionné en fin de liste
3. Retourner l'élément sélectionné

Ce qui donne comme représentation graphique des états de la liste produits par application de **ChoixSimple**, en supposant un état initial de la liste $(1{:}a, 2{:}b, 3{:}c)$:



Afin de sélectionner de manière aléatoire tout en privilégiant une rotation régulière des 3 éléments, on propose d'associer à chaque position i de la liste une probabilité p_i d'être sélectionnée ; les (p_i) vérifiant $p_1 \geq p_2 \geq p_3 \geq 0$ et $1 = \sum_{i=1}^{3} p_i$. L'algorithme de sélection devient donc :

ChoixAléatoire

1. Retirer de la liste l'élément à la position i avec la probabilité p_i.
2. Ranger l'élément sélectionné en fin de liste
3. Retourner l'élément sélectionné

Q 58.1 Peut-on considérer **ChoixSimple** comme un cas particulier de **ChoixAléatoire** ?
Q 58.2 En considérant la suite (x_t) des états de la liste produits par l'algorithme **ChoixA-**

léatoire aux temps t, ce processus est-il une chaîne de Markov (à temps discret) ? Pourquoi ? Quels en seraient les états ? La chaîne de Markov des états produits par **Choix-Simple** est-elle périodique et si oui, de quelle période ?

Q 58.3 Désormais, on suppose $p_1 \geq p_2 \geq p_3 > 0$.
Proposer une représentation graphique et la matrice de transition de cette chaîne de Markov.

Q 58.4 Cette chaîne est-elle irréductible ? Apériodique ? Quelle est la nature des différents états de la chaîne de Markov ?

Q 58.5 En utilisant les symétries de la chaîne, déterminer la probabilité stationnaire de chaque état.

Q 58.6 Donner la probabilité que l'état de la liste soit la même après de 3 itérations.

semaine 10

Exercice 59 – Mono-Bestiole aléatoire

On modélise le comportement d'un animal abstrait qui se déplace linéairement comme suit :

- l'animal se déplace dans un couloir qui ne possède que 4 cases numérotées de 0 à 3 de gauche à droite,
- l'animal se trouve forcément sur une des 4 case,
- à chaque itération, l'animal saute soit vers la gauche soit vers la droite. Un saut lui fait parcourir 2 cases. Toutefois, si le saut le fait se heurter au mur, l'animal se retrouve sur la case la plus proche (0 ou 3).
  Ainsi si l'animal saute depuis la case 1 : si c'est un saut à droite, il se retrouve sur la case 3; si c'est un saut à gauche, il se retrouve sur la case 0 après s'être heurté au mur gauche.
- la probabilité de sauter à gauche est notée L, la probabilité de sauter à droite est notée R.

Q 59.1 Donner les raisons pour lesquelles on peut modéliser ce problème sous la forme d'une chaîne de Markov à temps discret.

Q 59.2 Quelle est la seule information pertinente pour décrire l'état de l'animal ? Combien d'états possèdent donc la chaîne ? Proposer un graphe de transition ainsi que la matrice de transition de la chaîne.

Q 59.3 Cette chaîne de Markov converge-t-elle ? Calculer la distribution stationnaire.

Exercice 60 – Sociologie des BLOGs.

Un sociologue décide d'étudier comment sont maintenus les BLOGs d'une certaine communauté. Son centre d'intérêt est principalement la fréquence de mise à jour de ces BLOGs (comme indicateur de l'investissement du blogueur). Pour cela, il propose de les caractériser par un critère de vitalité : chaque BLOG peut être considéré comme Inactif, en Sommeil ou Actif en fonction de sa fréquence de mise à jour. Le sociologue s'intéresse alors aux évolutions des BLOGs au court du temps (unité de temps = 1 semaine).
Il s'aperçoit que sur 10 BLOGs Inactifs, 1 passe en Sommeil et 4 deviennent Actifs. De même,sur 10 BLOGs en Sommeil, 2 deviennent Inactifs et 5 deviennent Actifs. Enfin, sur 10 BLOGs Actifs, 2 passent en Sommeil et 2 deviennent Inactifs.

Q 60.1 Les hypothèses d'une modélisation en chaîne de Markov sont-elles respectées ? Tracer le graphe et la matrice de transition de cette chaîne de Markov (garder l'ordre I, S, A).

Q 60.2 Pour un BLOG donné, quelle séquence est la plus probable : $ASIS$ ou $ASSI$?

Q 60.3 Quelle est la probabilité pour qu'un BLOG soit Actif 2 semaines après avoir été en Sommeil ?

Q 60.4 Soit $\pi_0 = \frac{1}{21}[6, 4, 11]$, que vérifie cette distribution de probabilité ? Qu'en concluez-vous pour la chaîne de Markov ? pour l'état des BLOGs de la communauté étudiée ?

Exercice 61 – Commutateur

On considère un nœud de commutation et on s'intéresse à la file de sortie des paquets vers une destination donnée. Celle-ci est constituée d'un « *buffer* » stockant les paquets en attente d'émission vers la destination considérée et d'un « serveur » transmettant un paquet bit par bit sur la liaison de sortie. La capacité totale de la file (*buffer* + serveur) est limitée à K paquets. Un paquet se présentant à l'entrée de la file alors que celle-ci est pleine, est rejeté.
On s'intéresse à l'état de la file à des instants particuliers tk = k.T, k = 0, 1, 2, ..., où T est un intervalle de temps constant (*time slot*). On s'intéresse au processus stochastique Xk k = 0, 1, 2, ... , où Xk est le nombre de clients dans la file à l'instant tk.
On suppose que les arrivées de paquets et les fins de service sont distribués selon une loi de Bernoulli. Durant chaque *time slot* de durée T, il y a donc une probabilité p pour qu'il arrive un paquet (et 1–p pour n'en arrive aucun), et, conditionné par le fait que la file contient au moins un paquet au début du *time slot*, une probabilité q pour que la transmission du paquet en service se termine, libérant alors une place dans le *buffer* à la fin du *time slot* (et 1–q pour qu'elle ne se termine pas).

Q 61.1 Montrer que $(X_k)_{k=0,1,\cdots,n}$ est une chaîne de Markov à temps discret (CMTD). Donner la matrice et le graphe de transition associés.

Q 61.2 Sous quelle condition cette chaîne admet-elle un régime stationnaire ? Quelle est la nature des différents états de la chaîne ?
Par la suite, on prendra les valeurs numériques suivantes : $K = 3$, $p = \frac{1}{3}$ et $q = \frac{1}{2}$.

Q 61.3 Existence et valeur de la distribution de probabilité stationnaire ?

Q 61.4 En déduire le nombre moyen de paquets dans le commutateur (à la fin d'un *time slot*).

Exercice 62 – Urnes aléatoires

On répartit aléatoirement N boules noires et N boules blanches dans 2 urnes de N boules chacune. Puis, à chaque instant, on choisit une boule au hasard dans chaque urne et on les échange. On désigne par X_n le nombre de boules noires dans la première urne après n échanges.

Q 62.1 Pourquoi ce processus peut-il être modélisé comme une chaîne de Markov à temps discret ? Quel est l'espace d'état S de cette chaîne de Markov ?

Q 62.2 Soit l'état $i \in S$, quels sont les état atteignables à partir de i ? Calculer les différentes probabilités de transition. Représenter son graphe de transition et écrire sa matrice de transition.

Q 62.3 Cette chaîne est-elle irréductible ? Existe-t-il un n_0 telle que $\forall i, j \in S, P_{n_0}(i, j) > 0$ où $P_{n_0}(i, j)$ représente la probabilité de passer de l'état i à l'état j en n_0 étapes.

Q 62.4 On rappelle que $C_n^k = \frac{n!}{k! \cdot (n-k)!}$. Soit la distribution de probabilité π vérifiant

$$\forall i \in S, \pi(i) = K \cdot \left(C_N^i\right)^2$$

où K est une constante.

Q 62.4.1 En notant que $\pi(i)$ correspondant à la probabilité de tirer i boules noires (parmi N boules noires) et $N - i$ boules blanches (parmi N boules blanches) sachant qu'on tire en tout N boules parmi 2N boules, et en se rappelant que $C_n^k = C_n^{(n-k)}$, montrer que $K = \dfrac{1}{C_{2N}^N}$.

Q 62.4.2 Montrer que π est une distribution stationnaire pour la chaîne de Markov. Que peut-on en conclure en terme de convergence ?

Q 62.4.3 Quel est le temps moyen de retour en chaque état i ?

semaine 11 : annales

Exercice 63 – 2011 Commerce et hypothèses

Pour une étude de marché au sujet d'un produit P, on constitue deux échantillons de consommateurs de tailles 300 et 200 respectivement dans deux villes A et B. Cette étude met en évidence des préférences pour P pour $f_A = 56\%$ de l'échantillon de A et $f_B = 48\%$ de l'échantillon de B.

Q 63.1 On rappelle que la fréquence f est la meilleure estimation d'une probabilité p sur un échantillon; et que la variable F des fréquences sur les échantillons de même taille n vérifie :

$$F \sim \mathcal{N}\left(\mu, \sigma^2\right) \text{ avec } \mu = p \text{ et } \sigma^2 = \frac{f(1-f)}{n}$$

En notant p_A et p_B les probabilités de préférence du produit P dans les villes A et B, quelle loi suit la variable F_A des fréquences f_A parmi les échantillons de taille 300 de la ville A ? Même question pour la variable F_B des fréquences f_B parmi les échantillons de taille 200 de la ville B ? Quelle loi suit la variable $D = F_A - F_B$?

Q 63.2 Tester, au seuil de 5%, l'hypothèse : "Le produit P est préféré dans la ville A".

Q 63.3 Tester, au seuil de 5%, l'hypothèse : "Il y a une différence significative entre les 2 villes (deux comportements d'achat différents)."

Exercice 64 – Blés & hypothèses

Deux variétés de blé ont été cultivées sur $n = 8$ parcelles chacune. Les rendements observés sur les diverses parcelles, exprimés en quintaux par hectare, ont été les suivants (on les a rangés par ordre croissant) :

Variété 1	78.6	78.9	79.3	79.8	80.5	80.6	81	81.3
Variété 2	80	80.3	80.9	81.5	82	82.2	82.5	82.6

On se demande si l'on peut considérer que les deux variétés ont même rendement ou pas.

Q 64.1 Calculer la moyenne $\bar{x}_i$, la variance s_i^2 et l'écart-type s_i des observations des rendements de chacune des variétés $(i = 1, 2)$. Dans les résultats obtenus, qu'est-ce qui suggère que les différences observées s'expliquent par des fluctuations aléatoires ?

Q 64.2 On fait l'hypothèse suivante : les observations précédentes sont deux échantillons indépendants tirés de deux v.a. X_1 et X_2 de lois normales $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$.

Q 64.2.1 On estime σ_i^2 par $\frac{n}{n-1} s_i^2$ $(i = 1, 2)$; pourquoi ne prend-on pas s_i^2 ? Calculer ces estimations.

Q 64.2.2 Dorénavant on admet que les lois de X_1 et X_2 sont de variances connues, de valeurs données par les estimations précédentes. On suppose en outre que les deux échantillons sont indépendants l'un de l'autre. On sait que $\bar{X}_1$, $\bar{X}_2$ et $D = \bar{X}_2 - \bar{X}_1$ suivent aussi des lois normales; montrer que leurs espérances valent respectivement m_1, m_2 et $m_2 - m_1$ et que leurs variances valent respectivement $\frac{1}{n}\sigma_1^2$, $\frac{1}{n}\sigma_2^2$ et $\sigma_D^2 = \frac{1}{n}\sigma_1^2 + \frac{1}{n}\sigma_2^2$.

Q 64.3 Quelle est la loi de D sous l'hypothèse H_0 : "X_1 et X_2 ont même espérance" ? Quelle est alors la loi de $\frac{D}{\sigma_D}$?

Tester l'hypothèse H_0 contre l'hypothèse H_1 :"$m_1 \neq m_2$" (test bi-latéral) avec probabilité d'erreur de première espèce $\alpha = 5\%$. Doit-on considérer une variété comme meilleure que l'autre ?

Exercice 65 – 2009 - Ajustement et SNCF

La SNCF a rendu public le nombre moyen de pannes par mois sur certaines de ces lignes :

Numéro de ligne	Nombre d'accidents
ligne U	84
ligne V	99
ligne W	81
ligne X	66

Une association d'usagers vous demande de vérifier que le nombre de pannes ne dépend pas de la ligne.

Q 65.1 Quels sont les hypothèses à formuler ?

Q 65.2 Quels seraient les effectifs théoriques dans le cadre de l'hypothèse $H0$?

Q 65.3 Pour un seuil de signification $\alpha = 0.05$, que pouvez-vous répondre à l'association ?

Exercice 66 – 2009 - Chaînes de Markov

Soit une chaîne de Markov dont la matrice de transition est la suivante (avec p, q et $p + q \in [0, 1]$) :

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ (1 - p - q) & p & q \\ \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

Q 66.1 Donner le graphe de cette chaîne de Markov.

Q 66.2 Discuter, en fonction de p et q, de l'irréductibilité de cette chaîne de Markov. Quelle condition supplémentaire faut-il pour que cette chaîne possède un état stationnaire ?

Q 66.3 Avec $q = 0$ et l'état 1 comme état initial du processus, calculer la probabilité de revenir dans l'état 1 si $p = 1$ puis si $p \neq 1$. L'état 1 est il transient ? récurrent ?

Q 66.4 On suppose $q = 0$ et $p = 1$, calculer le nombre moyen de passages par l'état 3 (toujours avec comme état initial 1).

Exercice 67 – 2009 - Ajustement

Un étalon gris hétérozygote accouplé à des juments non grises produira des poulains gris avec une probabilité de 0,25 d'après les lois de Mendel.
50 juments non grises accouplées à cet étalon et ayant produit chacune 5 poulains (1 poulain par portée) ont donné les résultats suivants :

Nombre de poulains gris sur les 5	0	1	2	3 et plus
Nombre de juments	10	18	16	6

Q 67.1 Soit X la variable aléatoire prenant pour valeur le nombre de poulains gris par jument. Montrer que, d'après les lois de Mendel, X suit une loi binomiale $\mathcal{B}(5, 0.25)$.

Q 67.2 À l'aide d'un test de χ^2 (au seuil 5%), vérifier si les résultats observés permettent d'accepter l'hypothèse de Mendel.

Exercice 68 – Urnes aléatoires

On répartit aléatoirement N boules noires et N boules blanches dans 2 urnes de N boules chacune. Puis, à chaque instant, on choisit une boule au hasard dans chaque urne et on les échange. On désigne par X_n le nombre de boules noires dans la première urne après n échanges.

Q 68.1 Pourquoi ce processus peut-il être modélisé comme une chaîne de Markov à temps discret ? Quel est l'espace d'état S de cette chaîne de Markov ?

Q 68.2 Soit l'état $i \in S$, quels sont les état atteignables à partir de i ? Calculer les différentes probabilités de transition. Représenter son graphe de transition et écrire sa matrice de transition.

Q 68.3 Cette chaîne est-elle irréductible ? Existe-t-il un n_0 telle que $\forall i, j \in S, P_{n_0}(i, j) > 0$ où $P_{n_0}(i, j)$ représente la probabilité de passer de l'état i à l'état j en n_0 étapes.

Q 68.4 On rappelle que $C_n^k = \frac{n!}{k! \cdot (n-k)!}$. Soit la distribution de probabilité π vérifiant

$$\forall i \in S, \pi(i) = K \cdot \left(C_N^i\right)^2$$

où K est une constante.

Q 68.4.1 En notant que $\pi(i)$ correspondant à la probabilité de tirer i boules noires (parmi N boules noires) et $N - i$ boules blanches (parmi N boules blanches) sachant qu'on tire en tout N boules parmi 2N boules, et en se rappelant que $C_n^k = C_n^{(n-k)}$, montrer que $K = \dfrac{1}{C_{2N}^N}$.

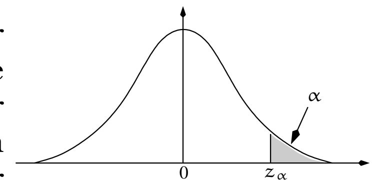
Q 68.4.2 Montrer que π est une distribution stationnaire pour la chaîne de Markov. Que peut-on en conclure en terme de convergence ?

Q 68.4.3 Quel est le temps moyen de retour en chaque état i ?

Tables

Table de la loi Normale

La table ci dessous donne pour chaque valeurs α la valeur du seuil z_α telle que $P(X > z_\alpha) = \alpha$ avec $X \sim \mathcal{N}(0,1)$ une variable aléatoire gaussienne centrée réduite. Attention sur cette table les dixièmes sont en ligne et les centièmes en colonnes (par exemple $P(X > 0,31) = 0,3783$ est la valeur sur la 4ème ligne et seconde colonne).

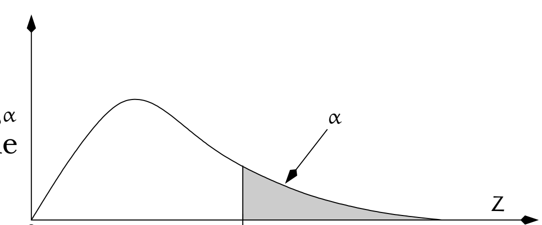


valeurs dans le tableau ci-contre : les α tels que $¶(Z > z_\alpha) = \alpha$

z_α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0859	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0466	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Table du χ^2

La table ci dessous donne la valeur de seuil $c_{r,\alpha}$ telle que $P(Z \geq c_{r,\alpha}) = \alpha$ avec $Z \sim \chi^2_{(r)}$ une variable aléatoire suivant un χ^2 à r degrés de libertés.



valeurs dans le tableau ci-dessous : les $c_{n;\alpha}$ tels que $P(Z > c_{n;\alpha}) = \alpha$

$n \setminus \alpha$	0,995	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,005
1	0,0000393	0,000157	0,000982	0,00393	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	22,3	25,0	27,5	30,6	32,8